The Seven Sacred Cows of Academia

Paul E. Meehl

Improving the cost efficiency of higher education was a topic Paul Meehl thought about over many years, dictating material on various aspects of the problem whenever another "it's-a-crisis" article or book came to his attention—"We could fix that" being his motivating response. Teaching, research, and other writing projects usually claimed higher priority, but he had a rough book draft by 1997. Here is that manuscript, edited to incorporate additional inserts in various stages of completion.

As he says in the Afterword, it was 'written partly for catharsis.' The result is less formality in style and more variety in topics. Text would have become tighter, more polished, and contain many more cites if Paul had done the editing himself. Sometimes notes and annotations have been left on references; if they were helpful to Paul, they may help readers. Cost estimates were already out of date in this draft; but actual dollars saved would always depend upon contemporary economics. The 1997 draft heading was kept to remind readers of the date of this manuscript and that the arguments and conclusions do not depend on specific dollar amounts, precise percentages of time or cost changes, or newer technologies available for education. The power of his message is in *how challenging Sacred Cows and using social science research findings can reduce the cost of higher education*.

I am grateful to Val Escher, who transcribed Paul's dictation for many years, including this book. Her patience, intelligence, and library research skills helped Paul Meehl stay productive throughout his happy retirement.

—Leslie J. Yonce-Meehl, December 2018

Table of Contents

Introduction

There is a way that a state university can lower student fees, reduce its tax support, increase faculty salaries, and improve the quality of instruction. This combination of desirable outcomes, prima facie impossible because they seem to conflict with one another, is easily achievable by applying the clear results of social science research to modify the way we go about our educational task. The economic changes possible are not small. I show in concrete detail, relying on hundreds of quantitative research studies, how a sufficiently clear-headed, determined program of efficiency could reduce the faculty salary component of instructional cost *to one-fourth of its present value*. This astonishing figure should surely induce academics, legislators, parents, and taxpayers to take my (admittedly radical) proposals seriously.

It is common knowledge that the economics of higher education is a mess, and matters are expected to get worse before they get better. College professors complain because their annual salary increases run below cost of living changes, having fallen farther in real wages over the last two decades than for any major employed group except librarians. College students find it difficult to pay the fees, and families with college-age children may spend a third of their disposable income on higher education. Meanwhile, the taxpayer revolt means that the dollars—already scarce to run colleges and universities—are likely to get scarcer, while inflation makes these scarce dollars buy less. The situation is widely documented and discussed in newspapers and alumni publications (see, e.g., C. Benda, 1994; Casse & Manno, 1998).[1]

Awareness of this economic problem and its grave implications has led informed persons in both private and public sectors to consider ways of paying the high cost of college education that will be less painful, risky, and inequitable to students, taxpayers, and professors. Congressional committees have scrutinized the situation and heard expert testimony about possible remedies. State legislatures have considered educational tax breaks of various kinds. Savings and investment schemes have been devised, anticipating the newborn's college costs two decades in the future. The problem of student loan defaults is still with us. Political and ethical issues arise when we ask "who should pay what fraction for a student's getting a BA, MD, PhD, or JD?"

If we're having trouble paying for something, we may ask, "How can we arrange to pay for this, when it costs so much?" However, there are two sides to any economic coin. *We should also ask*, "*Does it have to cost this much*? *Can we buy it cheaper*?" I have the impression that almost no attention is being paid to this other side of the coin. Discussions seem to take for granted that college costs can only stay fixed or increase,

[1] [*An appendix with collected examples of headlines was planned, but they are ephemeral and new ones appear routinely.—LJY*]

and that this parameter must be taken as given. I believe this presupposition is a serious mistake, and that is what this book is about.

TABLE 1. The Seven Sacred Cows of Academia

I.	**Teaching and research complement each other, they go hand in hand, so that an institution, group, or individual cannot do a good job of one without concurrently doing the other.**
II.	**Formal classroom lecturing is an efficient and indispensable method of basic instruction at all levels and for all subject matters.**
III.	**When classroom lecture is the method of instruction, small classes are much preferable to big ones over the whole range of sizes and subject matters.**
IV.	**Most research published by college professors is worthwhile.**
V.	**No one can assign priorities among fields because we each believe our own field is valuable.**
VI.	**Going to college is good for almost everybody, hence the more persons who get a college education the better; and, ideally, almost everybody should go to college.**
VII.	**Providing persons with a college education is socially valuable, a collective good; for that reason, the taxpayer should be the main source of support**.

While I am neither an economist nor an academic administrator, I am a professor emeritus who has served as chairman of a large department at a state university and on the executive committees of three other departments. With those perhaps marginal qualifications, I will show here how college costs could be dramatically reduced. Listed in Table 1 are seven assumptions about college education that seem almost universally taken for granted by professors and academic administrators, assumptions that we do not know to be true, and which jointly cost us needless expense if they are false. Because of their usually unquestioned status I call them Sacred Cows, meaning that they are devoutly believed by the denizens of academia and that it is considered wicked, disloyal, or cranky

to raise any doubts about them. By challenging some of them, we can painlessly reduce teaching costs by as much as 75%. Were we sufficiently informed, rational, and courageous to challenge all seven assumptions concurrently, we could almost double faculty real wages *and* reduce both student fees and tax support materially. I realize such numbers will seem preposterous, but the text justifies a savings equation that proves it. I further show, relying on a large body of empirical research, that the efficacy of instruction will not suffer but, on the contrary, will be improved.

The term 'Sacred Cow' does *not* refer to whether a certain belief is objectively true or false, nor whether it might be rationally defended by theoretical arguments or empirical evidence. Neither its truth status nor its rational defensibility is what makes it a Sacred Cow. The crucial point about a Sacred Cow is the culture's attitude toward it. There are Sacred Cows that are true, Sacred Cows that are false, Sacred Cows that are true but hard to defend at this time, Sacred Cows that are false in the eyes of Omniscient Jones but seem well supported by present data, and so on. I do not attempt to prove conclusively that these seven unquestioned postulates are all false. I do not even assert that they are all false, although I am inclined to believe they are. My purpose is to list them and to make four points about them: (a) they are widely believed, (b) they have not been proved to be true, (c) some are contradicted by sizeable bodies of evidence, and (d) they waste money.

Miscounting by Pretending Not to Count[2]

Several of my arguments involve judgments that hinge upon some sort of quantification (counting, measuring, rating) of something that exists in varying degrees or frequencies. For example, an administrative decision must be made as to which departments ought to have research units attached (imposing the conventional publication requirement upon faculty to be tenured in such departments) and which departments are to have no research unit attached, hence avoid the publication requirement to be tenured. Although the administrative decision is in the nature of things a dichotomy—like a legal judgment of bankruptcy, or a criminal judgment of delinquency, or the admission of someone to dental school—the factors that enter into the decision maker's judgment are usually not dichotomous, not "all or none," but are matters of degree. I find that proposing a list of matters to be looked into, quantifying "criteria" (using that term in the sense of a set of rough guidelines rather than as something definitory of the variable of interest) of research productivity and scholarly impact meets a good deal of resistance, even from social scientists, who themselves regularly engage in such counting and measuring in their social science research.

---

[2] [*This separately written insert* (*initially to be an appendix*) *is a broader message, generally applicable to issues throughout the book. It bears understanding before, keeping in mind during, and remembering after reading the book. I elected to simply put it at the beginning.—LJY*]

Introduction

We decide to attach a research unit and retain the publish-or-perish rule for Department A because its faculty, at least on the average or collectively, turn out many books and articles. They bring in a good deal of outside grant money for their research projects, generating (inflated) overhead to the University. Their work is frequently cited in the research literature by scholars around the world. Many of the faculty are recipients of prestigious prizes and awards for their scholarly achievements. They receive numerous invitations to participate in convention panel discussions and to give colloquia at other schools. Their graduate program succeeds in attracting large numbers of doctoral candidates from elsewhere, and these candidates tend to have high scores on the Graduate Record Examination and were graduated with honors from high-quality colleges. When one suggests that some sort of composite index of these plausible criteria of scholarly merit should be used by the institutional decision-maker in sorting departments into the pure teaching and the mixed research and teaching categories, I find many persons, including psychologists (not, I am happy to report, all of them), express objections on the grounds that these things cannot reliably be quantified.

I do not address the difficult question how reliably and validly they can be quantified, but content myself with pointing out an element of contradiction in the objector's attitude, which, if I were judgmental about it, I would characterize as a not-very-subtle form of hypocrisy. The incoherence is simply this: When those who object to tallying research papers, citations, prizes, etc., are not discussing Meehl's Reformed University, but are in a faculty meeting discussing a concrete case, they themselves proceed to invoke the above criteria; and when pressed to *prove* that a certain decision is warranted, they use *words of degree*. I have participated in faculty meetings for a half-century and served for six years as department chair; I cannot recall a single instance of promotion to tenure, promotion to full professor, raise in pay, or meeting a competing offer in which these criteria were not invoked, pro or con, and where adverbs and adjectives of *quantity* were avoided. The reason for that, of course, is that they cannot be avoided. There is no rational basis for arguing that having produced *one* publication is "better than *zero*" as indicative of scholarly merit,[3] but ten publications is "no better than" one. I cannot imagine what kind of idealized utility function someone would have in mind who would postulate such a step-function discontinuity, once he allows himself to get into the business of appraising scholarly merit by referring to books and papers in the first place. Furthermore, the empirical evidence makes it clear that the common allegation of an inverse relationship between quantity and quality of scholarly publications is factually false, although one knows of individuals who manage to grind out lots of mediocre papers and find a publisher for silly books. I take it for granted, as I have throughout, that adopting an institutional decision policy that presupposes infallibility of our judgments

[3] *I* do not identify scholarly merit with grinding out books and papers, but the people I am talking about regularly do, and to make my point, I am here accepting their erroneous definition of "scholarship."

and perfect validity of indicators, whether of teaching or research or public service, is ruled out at the start. No social science, or, for that matter, no biological science, can get off the ground if the requirement of perfect validity of counts or measures for the variables that interest us is a prerequisite.

But I need not argue that case, because the empirical fact is that the same people who, in an abstract discussion of my proposed reform, object to counting books and articles, nevertheless proceed to tell me in a faculty meeting that "Jones is very productive," and, when pressed, they say it is because he writes "lots of books and papers"! Or, if somebody challenges with, "Well, he writes a lot, but I don't think his stuff is very impressive," what do they do? They say that his stuff is "very well received by the profession on other campuses than ours." And then, if you ask what makes them think so, they either tell you some anecdotes (e.g., they heard a favorable remark about candidate Glotz at a cocktail party in Milwaukee), or they tell you that Glotz is "often cited in the literature." I won't belabor the point, which is so obvious it's a little embarrassing that one has to explain it.

The basic mistake here is not realizing that when you're talking about things that exist in degree or frequency (e.g., how often, or how strong, or how intense), *you are engaging in a tallying or measuring operation whether you use words or numbers*. The main difference is that the words are less precise and less objective than the numbers. If you are going to invoke how many articles somebody writes in ten years, the burden is on you to show that she does write a lot. The way you do that is look at her publication list. If you are going to claim further that her publications, whatever their quantity, are of a high qualitative merit because people in her specialty cite her in the research literature, how do you prove that? Do you prove it by an anecdote, or do you prove it by counting citations?

It is not necessary to assume that the *Science Citation Index* or the *Humanities Citation Index* are infallible measures; nobody maintains that, their defects are well known. For example, you don't get cited if you're fourth author. Sometimes your name is misspelled or your initials are wrong. (I, for example, am often cited for some reason as "P. A. Meehl," which a careful counter would presumably spot, and, if necessary, can check by which specific work is being cited.) But if I am cited 200 times per year in the *Social Science Citation Index*, what difference does a few misspellings make? In implementing proposed institutional arrangements we are not comparing me with a colleague who is cited, say, 175 times a year. What we are comparing is somebody who is cited this many times per year with someone who is cited less than a half-a-dozen times in *five years*. And, of course, we aren't comparing a single member of a strong department with a single member of a weak department. Rather, we are considering *groups* of faculty, and what we find when we consult the *Social Science Citation Index* that way is discussed in the Sacred Cow IV chapter.

The basic confusion here is widespread, even among social scientists who have no excuse for it. What many people do not understand (I see it as so obvious that I am at a loss to explain why they don't) is that when we use words like 'frequently,' 'usually,' 'a lot,' 'seldom,' 'typical,' 'many,' 'stronger,' 'weaker,' *these are all quantity words*. They all make an empirical claim, either (a) about how often some kind of event takes place or (b) how large or intense a variable quantity is. The first kind of question is settled by *tallying* (counting instances). The second kind is settled by *measuring*, if possible; otherwise, by using some kind of judgmental *rating* scale. Suppose the opponents of tallying, measuring, and rating were consistent. How would it go? When somebody is up for tenure or promotion or a raise in salary to meet a competing offer, they would scrupulously avoid using any of these adverbs and adjectives of degree or frequency. But what would they then be talking about with respect to the criteria? I find it hard to conceive what it would be. The sad thing about this incoherent position is that it substitutes the speaker's impressionistic, subjective, anecdotal idea of a frequency or degree for a more accurate (although, of course, still fallible) indicator of the criterion invoked. The net result is that by pretending not to count, and deceiving themselves as to whether they are counting in concrete cases being adjudicated, what they propose is substituting a less reliable count for a more reliable count. If such opposition to a systematic way of counting carries the day, *they are achieving a miscounting as a result of a pretext that they don't count*. I know of no way to cure this cognitive disorder. In the field of clinical psychology, I have noticed its recalcitrance to cognitive therapy for a half-century; so, I am not optimistic about the chances of curing it in the present institutional context.

There is a rational concern that a clear-headed thinker could have, and it may be that the inconsistent position analyzed above is partly motivated by this valid consideration, but not thought through clearly enough to be properly expressed. Whenever a property or state of affairs that can exist in *degrees*, or a kind of event or property of a subclass of events that occurs with a certain *frequency*, is subjected to numerical representation, it is psychologically tempting to be so impressed with a *number* that one becomes subject to dangerous mistakes in appraisal of whatever one is attempting to judge. Two mistakes, distinguishable but related, can occur on this basis.

The first mistake is to exaggerate the accuracy of the number as a representation of the inferred intensity or frequency we are attempting to estimate by our fallible instrument. This is the other side of the coin from "anti-numerical" types, who think that they can avoid the difficulties and implications of quantification by eschewing explicit numeration or mensuration in favor of adverbs and adjectives of degree. "Numerical types" (e.g., engineers, industrial psychologists, statisticians, as well as clinical psychologists of the tough-minded skeptical variety) are favorably disposed to the numerical expression of somewhat fuzzy properties and variables, just as the anti-quantitative and many "humanities" types of academics are unfavorably disposed to numbers and prefer words, even when these words are inherently quantitative in their import.

Secondly, most decisions we make in academia—e.g., concerning whether somebody should be admitted to graduate school, whether an assistant professor should be granted tenure, or whether a particular department's research publications are of sufficient merit and scholarly impact to warrant classifying it as a research department—these terminal administrative judgments are almost never based upon a single property or relation, but upon some composite of considerations. A decision maker faced with a set of evaluative criteria or predictor attributes is, in reality, exhibiting a multiple regression equation "in his head," which is revealed by a mathematical analysis of the judgments made, despite his not having explicit weights consciously in mind in reaching the judgments. Even if an index were a perfectly reliable and valid measure of one of the set of appraisal criteria of individuals or departments, a systematic bias could still creep in if a subset of the criteria were subject to quantitative expression and others not. Suppose we had a perfect measure of scholarly impact (anticipating the verdict of history after we are all dead, as would be available to Omniscient Jones); a decision maker who is favorably disposed to numerical indices and prefers to think quantitatively, might err by giving this perfect measure that happens to be available in numerical form a greater weight than he would if it were embedded in a collection of criteria all of which were in numerical form.

As a clinical practitioner, I am fully aware of these two dangers, so I do not present them as a grudging concession to "the opposition" in this matter. To ascertain the seriousness of these two kinds of tempting mistakes for quantitatively oriented decision makers would be a major research project in its own right. Although the great social importance of the issue involved would warrant a sizeable expenditure of money and brains to find a replicable answer, it has not been done. Absent such research, let me try to soften the blow of these two *qualitatively* sound criticisms, not by saying they are not real, because they definitely are, but by making a plausibility argument that they are quantitatively not as big sources of error as might be feared.

What is a reasonable benchmark for an acceptable error in decision making attributable to the joint influence of these two mistakes? It is not infallible judgments, which are never attainable in complex human affairs. The proper question to ask is, "What is the expected error in *not* employing some sort of numerification, however crude, for criteria that are inherently quantitative in their intrinsic nature?" In order for exaggerated ideas of validity and spuriously high β-weights to render using counts, measures, or quantitative format ratings undesirable, the exaggeration must be bad enough to be worse than the pseudo quantification represented by anecdotal and impressionistic "evidence." It is a psychometric theorem that the upper limit of validity coefficients is the square root of the reliability. No social scientist familiar with the research on human ratings and judgments, especially judgments about another person's attributes, will maintain that such judgments are highly reliable. For example, the interjudge reliability of teachers' ratings of schoolchildren's academic ability is such that in order to match the reliability of an intelligence test, it is necessary to pool the ratings of seven or more teachers.

Secondly, aside from the unreliability factor in subjective judgments, a faculty member arguing pro or con as to promotion, tenure, salary increase, or an administrative committee evaluating the scholarly eminence of a department, if they are at all consistent in their anti-index attitude, will be confined to basing judgments upon impressions and anecdotes, which involves a terrible problem of *sampling error* (a different source of error from rater unreliability). If one presses a colleague who doesn't offer any quantitative support for a claim that somebody's publications are having an impact on the field, what does one find? Typically one or two anecdotes, stories of what some scholar alleged about the candidate. One hasn't the faintest idea how representative that minuscule sample is of the population of scholars capable of making an informed, reasoned judgment about the candidate's work. Anybody familiar with the research literature on human ratings and judgments in relation to more objective measures—even in very fuzzy fields like clinicians' appraisals of anxiety or of hostility in mental patients, or the leadership ability of business executives—will accept my claim that this body of research, which is vast and varied in context, does not warrant the belief that the errors of human impressionistic ratings and of anecdotal impressions based on happenstance samples of two or three persons that you happen to have run into in a cocktail party are likely to be smaller sources of error than a reasonable decision maker's being overly trusting of a numerical index and thereby giving it a greater weight in the equation.

Thirdly, those faculty members or administrators who are suspicious of numerical indexes (whether for good or bad reasons) will not remain silent in group discussions about decisions. I can myself testify that when one points to a long list of publications, or mentions the number of citations in the *Social Science Citation Index* of a candidate for a job or promotion, those opposed—whether they are generic anti-quantifiers or not—are not slow or shy about bringing up the usual objections. These objections, whether rationally or irrationally engendered in the minds of the objectors, since they do have an intrinsic qualitative merit, and should rationally be taken seriously, operate as a buffer against "deification of numbers" offered in evidence.

Finally, let me make a simple, obvious point that's often ignored in such discussions and that I find extremely powerful. However the group discussion has proceeded, with or without numerification of intrinsically quantitative attributes of individuals, and regardless of how the faculty is composed with regard to their pro- or anti-numerical biases, what do we finally do to reach a terminal judgment? *We take a vote*. What does it mean to take a vote in a group? It means to tally, to count, to numerify a bunch of human judgments. If you don't think that tallying frequencies of opinions has any validity with regard to what is being judged, why do you take a vote? Why not flip a coin, or rely on the chairperson's intuitions, or a small randomly chosen sample of the full professors? It is well known by political scientists that the classic defenses of majority rule (by Locke and followers) are surprisingly difficult to defend by any rigorous kind of argument, whether utilitarian or on some other "moral" basis. I am not here trying to argue the merits of majority rule. I only point out that we don't say, after discussion pro and con,

and some faculty seem to be vacillating, "Well, it looks to me like most of us favor promoting Assistant Professor X." We take a vote, we *tally* and *count*. Note that the fact of the importance of actually tallying yeas and nays is even greater in doubtful cases than in cases in which there is a near-Quaker consensus in one direction or the other.

Putting these four considerations together, I believe this is an adequate reply to those who argue that the admitted dangers of quantification are so important and incurable that we should rely upon a subjective, impressionistic pseudo-quantifying, instead of counting and measuring.

Additional References for Introduction:

Rescher, N. (1978). *Scientific progress: A philosophical essay on the economics of research in natrual science*. Pittsburgh, PA: University of Pittsburgh Press.

Rescher, N. (1980). *Unpopular essays on technological progress*. Pittsburgh, PA: University of Pittsburgh Press.

Roche, G. (1994). *The fall of the ivory tower: Government funding, corruption, and the bankrupting of American higher education*. Washington, DC: Regnery.

Smith, P. (1990). *Killing the spirit: Higher education in America*. New York: Viking Penguin.

Sykes, C. J. (1988). *ProfScam: Professors and the demise of higher education*. New York: St. Martin's.

Sacred Cow I

Teaching and research complement each other, they go hand in hand, so that an institution, group, or individual cannot do a good job of one without concurrently doing the other.

As Bertrand Russell observes, one occurrence of an event suffices to establish its possibility; so we know that this Sacred Cow is false when stated thus strongly. There are numerous examples of research facilities in private industry and government (and, increasingly, in universities) that have no teaching functions and turn out excellent research. On the other side, there are many liberal arts colleges that educate undergraduates superbly, but whose faculty do not publish, or are under no pressure to do so. Some instruction-oriented schools discourage it except in the form of textbooks (a pedagogical contribution), a kind of publication that in research oriented institutions gives one little or no credit toward raises, tenure, or promotion to full professor. It is odd that so many college professors claim to believe this Sacred Cow, when the first thing they do on obtaining a research grant is take themselves off teaching time! Single quarter and single semester research leaves, not to mention the traditional sabbatical year, are rather hard to explain if one believes that scholarly research and writing are facilitated by teaching responsibilities. Authors of scholarly books today frequently express thanks to the dean or other administrator, or to their department, or to a fund-granting agency for "relief from teaching and administrative duties enabling me to bring this work to completion." A major inducement offered to job-seekers, whether hiring a new PhD or attracting a tenured professor from another institution, is a "light teaching load." Eminent scholars and researchers (e.g., Nobel laureates) are sometimes hired with the understanding that they will not be required to do any teaching except for their own PhD candidates (who also serve them as research assistants). Is there anyone who is unaware of these facts? Writing in 1993, Thomas Sowell described changes in time spent on teaching:

> When Jacques Barzun [1945] wrote his classic *Teacher in America* back in the 1940s, he referred to a typical college professor spending 15 hours a week in the classroom. Today, even half of that time would be considered an excessive teaching load at many institutions. Indeed, 35 percent of today's faculty teach undergraduates only 4 hours a week or less. At research universities, 51 percent of the faculty teach undergraduates only 4 hours or less, and fewer than 10 percent spend as much as 11 hours a week teaching undergraduates [Boyer, 1987]. However, more than half of research university faculty spend 11 or more hours per week on research [Boyer, 1987]. (Sowell, 1993, p. 13)

Unless otherwise stated, in what follows I shall employ the term 'teaching' to denote classroom lecturing in undergraduate courses, as well as graduate instruction—lecture or seminar format—for students not one's own research assistants or a few other advanced

doctoral candidates interested in one's specialty or research domain. It goes without saying that seminar, laboratory, or clinical instruction of PhD candidates usually serves as a professor's research facilitator, and doctoral students cannot learn to be researchers without such instruction. As will be discussed below, a department offering the PhD absolutely requires *some* actively researching faculty, although not all instructors need be.

In challenging this Sacred Cow, I do not suggest that there is a negative correlation *over individuals* between teaching competence and scholarly production—that is, of course, a different question, referring not to whether the two jobs facilitate one another or would covary longitudinally for a single professor—but to whether there may be slight positive cross-sectional correlations due to task-relevant personal traits. That correlation, while not zero, is of negligible size (Feldman, 1987). That it is slightly positive (rather than negative, as disgruntled undergraduates sometimes complain) is theoretically predictable, because several personal traits (e.g., high intelligence, conscientiousness, achievement motive, verbal fluency, efficient work habits, wide but selective reading, energy, enthusiasm, and social potency) can be reasonably expected, *ceteris paribus*, to contribute to effectiveness as a classroom teacher *and* as a writer of scholarly works. Bresler (1968) showed that student ratings of teaching effectiveness were positively correlated with professors' number of publications and reception of government and local research support, a relation discernible in both junior and senior faculty and in sciences, social sciences, arts, and humanities faculty.

While some traits of individuals must, on the average, contribute both to teaching effectiveness and scholarly research productivity, lumping all "teaching" together is probably somewhat misleading unless carefully qualified. An overall positive correlation between the two behavioral sectors is statistically compatible with the existence of zero or negative relations in subcontexts. For example, the knack of exciting intellectual interest in a research seminar for advanced doctoral candidates may require quite different talents, values, and interpersonal skills than those needed to motivate sophomores taking a required course in introductory psychology. *Example*: The cliché "I teach *students*, not subject matter," I have never heard uttered by a first-class intellect, only by second-raters. The most inspiring teachers I had were passionately interested in the subject-matter, and that "cognitive passion" spread to me and students like me. Although we know little about this complex question, there is evidence that the traits conducive to high scientific achievement are not "altruistic" or "person-oriented" (e.g., Campbell, 1965, 1971, Table 3-31 on p. 111; Roe, 1953; Shaffer, 1953; Strong, 1943, Table 193, p. 716; Thorndike, 1955; Wispé, 1963).

There is, however, a distinction between a person whose abilities, interests and terms of employment are mainly those of a *knowledge transmitter* and one who is mainly a *knowledge producer*, most of us at high quality state universities being considered (and required) to be both. Given the dominant intellectualist values of the academy ("brains is the name of the game" and "ours is a research institution"), it is silly to pretend that we don't make some such distinction. We locate our colleagues (and ourselves) roughly on a

dimension of this kind, and any unbiased observer of the academic scene is aware of institutional arrangements, some tacit and others formalized, for making the distinction. I have already mentioned the most obvious institutional arrangement which does so, namely, that between a small but high quality liberal arts college devoted wholly to undergraduate instruction and a research oriented high prestige institution (whether state or private) that expects (and, increasingly, *requires*) "knowledge production" in the form of published papers and books as a condition for a faculty member to be tenured.

Within research oriented institutions the distinction is also made, and everybody knows that it is. The least invidious and most "automatic, invisible-hand" mode of recognizing the dimension is via research grants of such size as to permit a material reduction in teaching load. These are mostly available to promising younger faculty or highly visible older scholars who already have an established track record in the knowledge producing function. The selection mechanism is quiet and operates with different force among departments within an institution. When the grant is received the faculty member drops some portion of classroom teaching responsibilities. In recent years one knows of faculty (more in medical schools and institutes of technology than in liberal arts) who have taken themselves off teaching 100% for several consecutive years (thereby revealing their disbelief in Sacred Cow I). Money thus freed up for "unassigned instruction" is then available for a younger person, or an advanced doctoral candidate with the rank of instructor, or a non-tenure track assistant professor, or a professional in the community seeking part time teaching experience and an academic title, to cover these courses, assuming they are considered central to the department's offerings. When faculty go on sabbatical or single quarter leave or reduce their teaching time from grant money, one or more of the classes they would normally teach are sometimes simply not offered during the period of the grant.

Second, a less formalized procedure and more employed when I was a young man (one supposes because less outside money was available in those days), is a tacit understanding within a department, not challenged by higher administrative authority as long as instruction is being covered adequately, that Professor Glotz, being a "star" in our university's diadem, should be burdened with somewhat less stand up class room teaching time than Professor Sedlitz, who has a respectable but not exciting "yardage" and who has run out of gas as a knowledge producer. Perhaps Sedlitz is still turning out papers occasionally, but everyone knows that they are not very exciting and achieve little notice on the national scene.

Third, within the academy structure there may be officially designated *research units* whose faculty are relieved of some portion of the normally expected teaching load by virtue of their membership in such a unit. In my experience—and I have been involved either as a member or in the administrative monitoring of several such—these research units do have a disadvantage. They may become the object of envious, hostile attitudes by some who view them as a kind of illicit sinecure. This attitude may exist even when the unit is clearly the main source of whatever national prestige the associated department

possesses! Despite this disadvantage, I shall argue below that the most rational, fair, and *enforceable* way to save money through explicit recognition of the knowledge transmitter/knowledge producer distinction—jointly challenging Sacred Cows I and IV—is by formal attachment of research units to some departments and not others.

What about knowledge transmitters who are not knowledge producers? It is increasingly difficult to make an academic career being one of these, and at high-prestige state universities it is becoming well-nigh impossible. This trend costs the taxpayer money, because those who survive the publish-or-perish hurdle usually command high salaries. I believe there is a more than adequate supply of non-publishing knowledge-transmitters in the academic labor market who would be willing to teach more hours (even at lower wages, but on my system they would receive higher pay than currently). Any professor in a large research related school has had the experience of talking with advanced undergraduates who, in considering whether they should go on to graduate school, point out that they like the academic scene, that they are basically bookish, verbal, "idea"-oriented people who enjoy leading the life of the mind, who find "explaining things to others" rewarding, who would like job security from unpredictable economic fluctuations, and who dislike the time pressure and time-clock punching aspects of the business world, but who decide not to go into graduate work "to become a professor" because they are afraid of the "publish or perish" situation. These young people are aware that neither their abilities nor their passions are along lines that would make their publication track record adequate for tenure and promotion. To put the economics bluntly, they're bright scholarly and bookish people who *could* be had quite cheaply, especially as they move into the higher ranks on a "years-in-grade" promotion basis (rather than by presenting a retention case or achieving special international eminence by publications). If many such persons were hired, tenured, and promoted solely on the basis of their social function as knowledge transmitters, with no expectation on anybody's part *or their own* that they must also be knowledge producers, we would save money and we might improve the average quality of instruction of students at the undergraduate level. The money could be saved partly by their coming a bit cheaper, but, as we shall see, simultaneous challenge of all seven cows would save so much money that they could be paid considerably more than our (required) transmitters-*and*-producers are now paid. The chief source of cost reduction among the non-researching knowledge-transmitters comes via their carrying a markedly larger teaching load than we presently expect of our amphibious knowledge producer/transmitter faculty. Such further "institutional formalization" of the distinction between the two kinds of contribution requires explicitly challenging Sacred Cow IV along with Sacred Cow I. It should not be so hard to challenge these two, since current practice already reflects effective disbelief in both of them.

There may be a problem in formalizing the invidious distinction between knowledge producers and knowledge transmitters, as it is likely to be equated in the thinking of both parties as "bright workers" versus "super-bright drones." (It is, however, paranoid to imagine that the undergraduate student knows or cares about this.) The tough-minded

approach is to say that anybody at lower rank, a fresh-baked PhD at the Assistant Professor level, is counted as a bright worker-bee until he or she demonstrates unusual research talents. Perhaps a few exceptions could occur for a very small number of new PhDs who are initially hired as knowledge producers on the basis of early research track record, superb recommendations, and the kind of institution in which they were trained. It's a delicate matter how to formulate that, but it could be done once we drop our present hypocrisy and the Sacred Cows that necessitate it. A job candidate to be hired as a knowledge transmitter is told: "Your teaching load would be bigger than Glotz's, who has only been here a couple of years, and that's because he was hired mainly as a knowledge producer. We all know that, as a result of this agreement, he is under the sword of Damocles to produce—and brilliant stuff, too—which you are not. If that arrangement is even slightly unsatisfactory to you, then you had better not take this job. We are not going to require that you generate publication yardage in order to be promoted to tenure rank and to be given corresponding raises in your salary. When we say that teaching is just as important in this department and this university as research, we mean precisely that, and we can show you the data to prove that we have sincerely tried to implement that policy. If you later should begin to think otherwise about your career intentions, the chairperson and your colleagues would like to know about it. If, despite your somewhat heavier load of teaching and student advising and service responsibilities in this department, you succeed by extraordinary brains, good luck, and burning the midnight oil more than other people do, in generating a distinguished research output—truly excellent work, not merely competent yardage, *because we are here challenging the Sacred Cow that values merely competent yardage*—then your 'classification' as a worker-bee or drone will be up for re-examination. But meanwhile you must be clear about the terms of our agreement. Merely competent yardage will not suffice to shift you over into the other category."

Some say that a really bright, scholarly job-seeker would not like these conditions, as they seem to classify one as a second-rate person. I dare say a few would think so, and they would not take a job on such terms. But I deny that there would be such a scarcity of takers that the jobs would go unfilled, and I insist that nobody has shown any such consequence, especially when—on my proposals—the department chair can conclude his remarks by saying, "University X is offering to start you at $30,000 under the usual publish-or-perish condition. We impose no such condition, and your starting salary will be $50,000." It is absurd to hold that such terms will not be attractive to young persons of scholarly bent.

What happens to the poor fellow who had thought he wanted to be purely a knowledge transmitter, spending a lot of time with students, leading the life of the intellect in a leisurely manner, reading books and articles rather than writing them, out from under the publishing Sword of Damocles, but who now becomes inspired by a brilliant idea for a theory or experiment and wants to change vocational tracks? I have no hard data on how often this happens to people. I have myself known very few academics, I think not more

than 5% at the outside, to whom this happened. Still, even if it is relatively unusual, it must be admitted as a minor defect of the two-track approach. All I can say is that this kind of consideration is one of utilitarian weight, and must not be perceived as an injustice. The person made a vocational choice based on the best self-knowledge and information about the academy available at the time of job hunting. He entered into a contractual engagement with his present academic institution whose intentions and standards were, we presume, made crystal clear. Nobody has wronged him or mistreated him, or "not given him a fair chance to develop his talents," because *he* miscalculated.

Nevertheless, from a utilitarian standpoint, whether we speak in terms of individual dissatisfactions or the possible contribution someone could make to the advancement of knowledge and perhaps to some practical problem of society, there is a legitimate concern. All I can usefully say, absent hard data about the career lines of such rare individuals, is that the possibility that a person makes a mistake in judgment in planning a career is part of human life. This disadvantage to a small minority of professors who make such a miscalculation has to be put in the hopper along with other objections to the two-track proposal; but it cannot be considered dispositive, inasmuch as only disutilities rather than fundamental principles of justice are involved. I repeat—perhaps tiresomely, but I find that people have a hard time keeping their heads straight on this one—we began our inquiry by accepting an empirical statement that *the present state of affairs is not satisfactory and something drastic must be done to alleviate it*. Justice or unfairness not being an issue, the disutility of, say, 1 in 20 people miscalculating in their career plans has to be weighed in the balance along with all the other utilities and disutilities which are the subject matter of this book.

Furthermore, it would be erroneous to assume that none who change their minds at age 30 or 40 (I doubt that many of them will do it later, asking readers to consult their introspections and observations in this respect) would be able to make the change. A knowledge transmitter who decides to write a few scholarly articles may certainly do so without moving into a different kind of institution. It will take more burning of the midnight oil and weekend work than it would for those who are ensconced in institutions or departments that classify themselves as "knowledge producing." Professors at Columbia Law School in the late nineteenth century taught three hours a day for each of five days a week (Kaufman, 1998, p. 43). A teaching load of 10 or 12 hours weekly, while heavier than most of us in research productive institutions have been accustomed to since World War II (it is 2.5-to-3 times the "guideline minimum" in my college, not stringently enforced), is not excessive by the standards of the 1920s and 1930s. At American universities in the 1930s, "the average teaching load was twelve or fourteen hours a week" [and] "research was an optional activity" (Macrae, 1992, pp. 174-175).

Teaching a heavy load in a teaching oriented institution does not literally *prevent* a bright, creative and energetic individual from doing and publishing scholarly research. It simply puts a heavier burden upon that individual. My own view, as the reader will infer from other things in this book, is that this burden will tend (fallibly) to screen out persons

with marginal talent and motivation, so that what we will get from those who undergo such a change in career orientation will be of higher quality than what we are getting now in mediocre departments with the routine requirement of publication yardage. While admittedly people in their 30s or early 40s may have a somewhat harder time getting a new job, everybody familiar with the academy knows that it is not an impossibility.

That a few will suffer disappointment and frustration goes without saying. A large number of faculty are suffering disappointment and frustration now because they have what they consider to be inadequate salaries. As always in economic matters, there is a trade-off. If we want to get the advantages that I claim are associated with classifying institutions, departments, and individuals as knowledge transmitters versus mixed transmitter-generators, we will have to pay the price of some disutility in the careers of the minority who miscalculate about how they want to live their academic lives. I have no answer to those who assume, if I understand their reasoning, that society must somehow provide a safety net for everybody against all possible consequences of all possible mistakes. Radical changes in one's career goals are always troublesome. They involve some extra work and sacrifice whether they take place in business, government, the military, or wherever. I see no good reason why we college professors should feel entitled to be preserved from the inconvenience and anxiety attendant upon having made such a mistake.

One may be pardoned for an elitist speculation that such a definition of the role and possible change in it might be an automatic selective factor in that those most talented and most highly motivated would, statistically speaking, be the ones that moved over. But we must be clear that this is at best a statistical expectation, and we should not get hung up applying some inappropriate concept of inherent justice as "equal situation in all respects." As long as the contractual understanding is clear and honored by both parties, no injustice is done. In tough economic times, if the Sacred Cows about research have been effectively and sincerely challenged, one cannot operate a department or engage in decision making about personnel in the spirit of Gray's *Elegy*![4] At the risk of sounding like a cantankerous old codger, I permit myself to point out that, "in the old days," before there was so much outside research grant money, the typical young fresh-baked PhD assistant professor *both* carried a heavier teaching load than older and famous faculty colleagues *and* struggled to publish noteworthy research, and took the situation for granted. Finally, no industrial psychologist would require that a selection or placement procedure be infallible, and it is foolish for us to entertain such fantasies in running the academy.

Unless collective bargaining should become standard pretty much throughout the country, in both state and private institutions, and perhaps even some degree of "equalization" aimed at by the bargaining organization at the national level, there seems no feasible way to assure that the worker-bees will have the same average salary at a

[4] [*Thomas Gray's poem*, Elegy Written in a Country Churchyard, *published in 1751.—LJY*]

corresponding (later) age in grade as the minority of drones. The reason for this is obvious to anyone who knows how the academy works in regard to raises. The most important source of an individual's sizeable raises, even in good times, is neither a legislative mandate for cost of living increases or some average value "merit increase," but rather the necessity to meet a competing offer in a tough retention case. Professor Glotz at the University of Minnesota is known to Harvard and Berkeley for her publications, and they don't really care much about her teaching ability (or her "service") so long as these meet minimum standards. Whereas Professor Fisbee may be a brilliant and dedicated classroom teacher, who also gladly spends many hours a week counseling and encouraging students informally, facts which cannot easily come to the attention of the Harvard or Berkeley faculty, even if they did care deeply about these things. Hence, even if the Seven Sacred Cows were challenged at all institutions of higher learning simultaneously, it would still be the case that the visibility of one's high level contribution as knowledge transmitter (or knowledge applier in service to the community) cannot be in the running with the visibility produced by the publication of scholarly papers and books. A state university that sincerely and effectively challenges the Seven Sacred Cows, in firming up its contractual understanding with an assistant professor who is being hired as a knowledge transmitter, can honorably say and can subsequently deliver on the promise that tenure and ordinary merit raises will be forthcoming without publication yardage; but they cannot, unfortunately, make the promise that there will be no cumulative salary differential over the ensuing years between such an individual and a national "star" who can be retained only if the institution fends off high salary competing offers from other institutions. I believe that this problem, with anything like our present institutional arrangements, is incurable, so that a late stage salary differential will have to be accepted as part of the bargain. But again, no "injustice" is involved if one thinks rationally about the problem, so long as the understanding is clear, the contractual relations are explicit, and the realistic expectations thereby engendered are honored. Besides, *knowing that one's own starting and ultimate salaries are almost twice what other institutions* (*worshiping the Sacred Cows*) *offer*, only a paranoid or neurotic injustice collector would fret over the higher salaries of knowledge producers in *his* institution.

The institutional formalization and—of great psychological importance—the associated semantics of the knowledge transmitter/knowledge generator distinction need not be done in the same way at all universities that are prepared to challenge Sacred Cow I (along with Sacred Cow IV), or even within a single university. However, there must be a compromise between the undesirability of an inherently invidious labeling which, while it cannot be avoided, should be minimized, and the need to formalize implied by a sincere and courageous challenge of the Sacred Cows in question. As ways of recognizing the worker/drone distinction, one thinks of the following, the merits of which I shall not discuss in detail:

Faculty could attain tenure at the usual associate professor rank without publication, and as early in their career as knowledge generators who publish, but the rank of full pro-

fessor might be restricted to those who publish. Such a suggestion seems shocking to us now because we're not accustomed to it (although I believe it was once the case in some schools before Sacred Cow I became so widely worshipped). Needless to say, there's no point in challenging a Sacred Cow unless we think we have a decent chance of persuading people, given the imperative economics of our situation, to accustom themselves to a changed language that refers realistically to differences in expectations and status.

A variant on this method would be to increase the use of titles such as "Research Professor," or honorific ones like "University Professor," "Regents' Professor," or "Distinguished Service Professor," or professorial "chairs" named for private donors. This method might be preferable to the first because it doesn't prevent a non-publisher from achieving the standard top level rank of full professor. It merely *adds* something special, sort of like making a general a field marshal, or giving a few of those at high rank the Blue Max!

A third alternative is to increase the number of formally designated research units associated with various departments on the understanding that not every department (presumably *not the majority* of the departments at a particular university) will have such a unit attached. While the dollars in those units would be a mix of hard and soft money, there would be at least enough state money appropriated so that such a research unit would not depend for its very existence on outside grants, private or federal. I have worked in three such units during my career. This arrangement has the advantage of flexibility in the case of young faculty who would like to make a try at the knowledge producing business (even though they were hired mainly as knowledge transmitters), as short-term partial appointments in such an associated unit could be available to a carefully chosen subset of worker-bees. On balance, the attached research unit system appears best to me. Relying on the data below refuting Sacred Cow IV (that most research is worthwhile), I propose that not over one-fifth of teaching departments should have research units attached. Faculty in the other four-fifths of departments would be institutionally defined as knowledge transmitters, under no pressure to function as knowledge producers.

The least invidious approach is to rely simply on the differential availability of research grants, as is presently done. The disadvantage is precisely its lack of formalization. At least at my university, whether faculty are on such a grant or not, and even if they have teaching loads somewhat higher than in other departments more nationally known, they are still expected to be knowledge generators by producing publications. No matter how good one's record as teacher or in community service, in order to be tenured in the Arts College of my university it is an absolute requirement that one publish. The college dean's promotion and tenure committee (on which I have served) is not instructed that this is a "laudable goal," "policy," or "guideline." It is close to being a hard and fast *rule*, and the Committee is fully aware that deviations from it would be overruled by the dean or academic vice president. Now, you can't challenge Sacred Cows I and IV meaningfully, so as to make a major inroad on the university's budget problems, if you

continue to insist that persons of marginal scholarly talent and creativity, located in departments that have mediocre national status as knowledge generators, must nevertheless produce books and papers in refereed journals in order to be given tenure. It would be unfair to such individuals to require that they must teach 10–12 hours a week while they are trying to do this, usually without outside grant support for their research. The challenging of Sacred Cows I and IV is intended to make it fair and humanly possible for the pure knowledge-transmitters to teach twice as many hours a week as they now teach, *because they have this pressure to publish effectively removed from them*.

The problem with relying informally (without any official rank designation, special research unit, or similar institutional arrangements) upon the operation of the fund granting "invisible hand" to subsidize some faculty and not to subsidize others is that it cannot be used to yield a non-stressful and fair situation in which there can be a marked reduction in teaching time dollars without an equivalent reduction in classroom lecture courses being offered. The point of formalization is that the rational expectations of knowledge-transmitters should be that they don't have to publish articles, that they will be tenured, and meanwhile they will be teaching 10–12 hours a week. The small subset of knowledge generators teach fewer hours a week. Their knowledge-transmitting function consists in senior and graduate level classes and the advanced instruction of their own doctoral candidates who are functioning as research assistants in their research programs. They have these expectations in explicit form because they have the title 'Research Professor,' or because they're officially listed as professors in the Research Unit associated with their department, everybody knowing that on this particular campus only a minority of departments in the college have such associated research units. Reliance solely on a *non-labeled consequence* of invisible hand operation by fund granting agencies will not achieve this goal if those who do not engage in heavily subsidized kinds of research, or who are unsuccessful in obtaining external support, are nevertheless under the gun as nationally visible knowledge generators to generate books and articles (of whatever merit) but meanwhile also must teach about the same number of hours a week as other faculty.

I conclude that the optimal approach is creation of special research units having professorships listed in them, the institutional expectations being that this will be true only of a minority of departments in the college. It minimizes the invidiousness of titles, provides an obvious institutional funnel for outside grant money, and still makes sufficiently clear by the formal nature of an individual's appointment (designations *both* as a professor in the department *and* as a professor in the research unit) that the contractual understandings and expectations are different.

The problem of invidious comparisons impairing morale seems to trouble many persons when considering the idea of making a clear administrative distinction between departments to which research units are attached and those lacking such units. I have had personal experience in three settings where such negative feelings were discernible in a few—not many!—people in the main department but not in the attached unit. This

within-same-discipline situation is the most conducive to envy, as I shall explain shortly, so I view these experiences as providing a worst-case sample of the problem, and the three (very different) experiences are illuminating, despite their anecdotal character. Lacking statistical studies of a representative sample of such research unit situations, anecdotal evidence is the best we have.

Being neither a frenzied egalitarian nor a utopian about institutions, I do not set up the absurd aim of creating arrangements which will assure that everybody, including litigious, paranoid injustice collectors and envious mediocrities, will be "happy." But I shall try to reduce this concern by four considerations which I believe make the attached research unit approach a better bet for minimizing friction than its alternatives, once the Sacred Cow about everybody having to produce research is sincerely challenged and effectively implemented.

First, in the absence of explicit formal arrangements (but given tacit administrative "adjustments," and the consequences of external research funding releasing faculty from some part of teaching responsibilities), every perceptive academic is aware that *these distinctions are currently made*. Highly research productive faculty with national and international visibility in the scholarly community tend statistically to have lighter teaching loads, even *within departments* operating on the present Sacred Cow principles. And within, say, a liberal arts college those fields of knowledge that (for whatever reason) have a higher prestige are treated better, both as to salary and teaching load. Scholarly prestige aside, departments whose members have no viable alternative to being college professors do not do as well as those in fields where private practice or earning one's living in government, industry, or non-teaching research institutions are live options. Sometimes these discrepancies are quite sizeable and it takes little trouble to verify them. For example, I knew a Regents' Professor of classics whose quarterly load of classroom teaching hours was regularly twice as high as that of the average member of the psychology department and three times as high as mine. This anomaly arises because social sciences have a higher prestige than classics, and because there is very little opportunity for a scholar, however eminent, to go into private practice of Greek or Latin! I am not concerned to criticize or justify these arrangements. I am only pointing out that every informed person in the academy knows that they exist. Whether a tacit understanding in defiance of any open statement is worse for morale than a formal arrangement is debatable, but it is not obvious that a formal recognition makes it worse.

Second, invidious comparisons tend to be more painful when one is in daily contact with the economic or prestigeful competitor. Not serving on departmental committees or attending departmental faculty meetings together and being in separate physical locations all tend to reduce the incidence and intensity of invidious comparisons. My experience as a member of three such units suggests to me that physical-social propinquity is an important factor in stoking up envy. In the best case, wholly free of it, research unit faculty were completely officed in a different building and had no secretarial help in the main department's building. In the worst case, the unit's offices and conference room

were a few yards down the hall, and secretarial support was partly shared. The mildly frictional instance was midway (e.g., some unit faculty had offices in both the research unit and the department). The propinquity factor was exacerbated by a pronounced difference in research productivity and national prestige between unit and non-unit faculty in the worst case. In the envy-free case, faculty not attached to the special unit were of high scholarly eminence.

Third, being a contractualist in my ethics and politics, I do not shed tears for persons who have voluntarily accepted employment under clear, openly described conditions.

Fourth, suppose I am an assistant professor of botany, a department with no research unit attached, hired as a pure knowledge transmitter. I teach 10–12 hours per week, and no publish-or-perish Sword of Damocles hangs over my head. I am aware that Assistant Professor Glotz in the department of psychology, a knowledge-generating department with a research unit attached, only teaches 5 or 6 hours weekly, and perhaps even less, having a quarter off teaching or taking single quarter leaves to do research and writing. However, since psychology runs a large PhD program, Glotz has to sit on boring preliminary and final oral examinations, write and grade written "prelims," shepherd PhD advisees through their several tasks, attend committee meetings (to select graduate students, consider how to handle the frequent "problem cases," revise PhD course requirements, integrate offerings with other departments)—none of which onerous, tedious, and time-consuming tasks fall to me. Collectively these extra jobs can easily approach an average of 5–6 hours weekly (unevenly distributed over the academic year) *all taken away from the hours Glotz needs to write research grant proposals*. His research proposals to Federal fund-granting agencies (e.g., NIDA, NIMH, NSF) have only a 10% (or less) chance of being funded. He has to write research papers which (in psychopathology) have an 80–90% chance of being rejected for publication in the "mainstream refereed journals [the ones that count!]." For non-tenured faculty at research institutions this grant-and-publish struggle constitutes a pretty ulcerogenic situation, making academia a lot less leisurely and fun than when I was a young man. I can't see how I would feel put upon under these circumstances, unless by the negative *prestige* component of not being a knowledge generator in a research department; but in that case why would I have taken a job at Meehl's Reorganized University, whose botany department has no research unit? It just doesn't make good psychological sense.

Finally, and this is, I dare say, by far the most important buffer against the invidious comparison danger, there is nothing like a big dollars-and-cents reward to make people feel better. (As pianist Liberace used to say when people asked if he felt bad that some musical snobs looked down on his "popular" piano-playing, "Yes, it makes brother and me cry all the way down to the bank.") If I have knowingly accepted a position as assistant professor in a non-research department, I am not going to worry very much about Assistant Professor Glotz, scholar in a different discipline, officed in a building on the other side of the campus (and who is under the publish-or-perish pressure, which I am not), when we are both receiving a starting salary twice as large as either of us would

have got had we gone to a "research-oriented institution" not making this administrative distinction. Consider: I have taken a job as an assistant professor, a fresh-baked PhD, at Meehl's Reformed College at a salary of $80,000, instead of starting at a research-oriented institution at $40,000; and I have every reason to believe that when I am a full professor I will be earning $180,000 rather than $90,000 (uninflated). It would take a pretty sour disposition for me to fret about someone on the other side of campus who is sweating over publish-or-perish research grant proposals.

Judging by colleagues' conversations (so predictable that no formal "opinion poll" is needed and, absent opportunity to explain, would probably be misleading) the biggest (first and strongest) objection is to *the idea of non-research departments*. The self-concept "I am a research-productive scholar" seems so important to my psychologist colleagues that they feel threatened by my challenge to Sacred Cow I, despite their easy readiness to challenge Sacred Cow IV. If you, as a reader, are troubled by this, I say: If *you* desire a research career, *you* would not even consider a job at Meehl's Reformed University, unless *your* department was one with a research unit attached. So there is no threat posed to *your* intellectual lifestyle, self-concept, or national prestige, unless you think that the mere existence of non-research departments somehow reflects adversely on you. I urge you to reflect that this latter worry is irrational, because scholars in your field who follow your published work and that of your departmental colleagues, will not "judge" you on that basis. They will neither know nor care whether Department D in some unrelated field is research productive. In fact, as "research universities" function conventionally now, your national peers do not know or care about the quantity or quality of research turned out by those in other departments of your institution. Why, then, would you care?

I am aware that this reassurance doesn't suffice with everyone. It seems that the mere abstract idea of "being at a *research* institution," whatever the qualitative research merits of unrelated departments, has a powerful hold on academic minds. The psychotherapist in me perceives this as irrational; it strikes me, to be blunt, as a sample of unexamined word-magic.[5] But you are not my patient and so I have no effective leverage. I can offer this: Given that you, individually, are deeply offended by the idea of being a professor in a university where faculty in other departments got tenure without publishing (even if you had double your present salary), is this a good objection to the idea? I am not compelling *you* to take a job at Meehl's Reformed University. Your objection would probably be phrased not about you but as something like, "No bright, scholarly person leading the life of the mind would consider such a job," a strong empirical claim about *all others* which you can't prove and which I am confident is false.

I have heard as objection to such an arrangement that promising new faculty will not be attracted to an institution which is not "first-rate," a phrase translated as "highly nationally visible across the board," meaning that every department has to function as a

---

5 [*For more discussion of word-magic, see "The Tyranny of Words" in Sacred Cow VI.*—LJY]

knowledge generating unit. I think this is an armchair objection for polemical purposes, not based upon either honest introspection or observation of how potential additions to the faculty inquire. When as a young man I was considering tempting job offers at other institutions, I wanted to know whether in an institution's psychology department (and, ideally, associated departments that bore upon my interests, such as statistics, psychiatry and philosophy) there were some high prestige knowledge generators. But never once did it occur to me to ask whether the faculty in the departments of English or geography were highly visible knowledge generators. In forty years as a faculty member, including six as department chair and sometime member of search committees, I cannot recall a single occasion when a potential addition to the psychology faculty inquired about the quality of such unrelated departments. The plain fact is that when people go to a "first-rate institution" the thing they want to be sure is "first-rate" is the department that they are going to, i.e., the colleagues with whom they are associated, the quality of graduate students they will have in their classes and as research assistants, who subsequently go out as postdoctorals to "spread the word" for their ideas. Of course they may have an interest in some associated departments for the same reasons as I did. But I can't imagine anybody deciding not to accept an appointment in the Minnesota psychology department on the ground that, say, the geography department here was not first-rate (in fact, it is presently rated top in the country). If such a strange sort of identity-problem appeared in a job-seeker, I would be concerned about the candidate's mental health, or suggest some reading of Buddha, Epictetus, Marcus Aurelius, Bertrand Russell, or Albert Ellis (the Epictetus of Manhattan).

How do we decide which departments are going to have research professors or associated research units attached? The safest, easiest, and least painful method is to select those departments already having demonstrated knowledge producing *excellence*, not merely "competent" departments. Multiple criteria are available for evaluating that question.[6] The research on scholarly productivity bearing on evaluation of departments is examined in Sacred Cow IV, and the administrative strategy for transition to Meehl's Reformed University, with objections thereto, is treated in Chapter X. It won't do to object that those criteria are fallible. All evaluative criteria, in all walks of life, are of course fallible. More pointedly, deans and higher administrators perforce rely on them *now*, in our Cow-worshipping schools. That unavoidable evaluation process is not somehow made worse by openly recognizing its existence.

A colleague responding to an earlier draft of this book suggested that I have a Sacred Cow of my own, to wit, an uncritical adulation of the small non-research oriented liberal arts college, or an irrational conviction that it provides a markedly superior kind of undergraduate education to what is given at a large "research productive" state university.

---

[6] National ratings are made by, e.g., American Council on Education, American Council of Learned Societies, Council of Graduate Schools, Educational Testing Service, *Gourman Report*, National Research Council, National Science Foundation, Social Science Research Council, *U. S. News and World Report*.

I deny this. I took my BA degree at the University of Minnesota and my beginning psychology class had 250 students in it. I'm inclined to think, especially on the basis of my six years experience serving on the American Board of Professional Psychology, that my Minnesota undergraduate education in Psychology was superior in quality to what many psychologists received at smaller schools with smaller classes. I'm not concerned here to argue anything about the superiority *or* inferiority, on the average, of getting your BA from Michigan, Minnesota, Berkeley, or Texas versus getting it at Carleton, Reed, Antioch, or Oberlin. My only point in bringing up such examples as the latter is to refute the general claim (Sacred Cow I) that a faculty *has* to be both generating new knowledge and transmitting it to do a good job of either. A few such examples (and scores exist) suffice to show the falsity of that as an empirical generalization. But I am not committed to any claim that, on the average, the education you get from a small liberal arts college whose faculty do no research is superior to what you would get at a big state university locked into the knowledge-generating dogma. On the other hand, I am not persuaded by any evidence I've seen that the superiority goes in the other direction either.

Of course, there is the factor of being exposed to "role models" (I am getting bored with that phrase but I don't have a better one handy). The role model of the new knowledge generator and the article writer is one which we take for granted at the graduate school level, but nothing I have said in challenging the Sacred Cows disputes the desirability of having such available. It is absurd to hold that the only way the scholar-knowledge producer can be "modeled" is if done with respect to *every* sub-section of subject matter in *every* area of knowledge. I cannot imagine what kind of data, or what kind of theory of learning, social perception, or identification would enable us to draw such a preposterous conclusion. Again, I invite readers to reflect on their own academic past as an undergraduate student. I had courses as an undergraduate from such role models as renowned psychologists S. R. Hathaway, B. F. Skinner, Donald G. Paterson, and philosopher Herbert Feigl—not only exciting classroom teachers but also known to me to be new-knowledge producers, who were running rats, treating patients, computing statistics, writing books and research papers. Was the impact of these role models upon me somehow mysteriously attenuated because I *also* took courses from scholarly professors like R. M. Elliott, D. Jackson, and A. Castell (see Sacred Cow IV, p. 59 for examples of excellent teachers whose publication records might not get them tenured today), who were not new-knowledge generators but "only" scholars and good teachers? Most certainly not!

Additional Reference for Sacred Cow I, not discussed in text:

Huber, R. M. (1992). *How professors play the cat guarding the cream: Why we're paying more and getting less in higher education*. Fairfax, VA: George Mason University Press.

Sacred Cow II

Formal classroom lecturing is an efficient and indispensable method of basic instruction at all levels and for all subject matters.

When a college student speaks of "taking a course" or a faculty member says "offer a class in so and so," it is universally assumed that this process consists of the student sitting in a room with other students, writing down the words spoken by the professor who sits or stands in the front of the room giving a "lecture," a scheduled group event that happens 3–5 times weekly (at my university) for 10 weeks (quarter system).[7] Considering the variety of ways in which human beings learn all sorts of different things, both in and out of the academy, not to say the history of higher education in various countries over many centuries, it is remarkable the extent to which this automatic identification of education with *formal classroom lecturing* is presupposed. In Britain, where the lecture plays a minor role, conventional terminology for studying a subject is that one *reads* (classics, or math, or sociology) rather than the American verb *takes* (a class). Despite Nicholas Murray Butler's famous crack about a lecture being a process in which the notes of the professor become the notes of the student without passing through the mind of either—a gibe that, while exaggerated, has for those of us who have been in both roles an uncomfortable ring of truth about it!—it seems that most college faculty take it for granted that if students are supposed to acquire certain kinds of knowledge this will *of course* require arranging for them to be in a certain room at scheduled times with chairs, and a lectern, and a live-body professor talking at them.

The automaticity of this assumption was called forcibly to my attention in a faculty committee discussion of a revision of the clinical psychology curriculum. It was objected that we could not require a (needed) prerequisite course because unfortunately it was offered by a professor in another area who would probably resist having it moved from the winter quarter to the fall quarter, which would have been necessary for its prerequisite function. There were 8 faculty present at the meeting and they all frowned and looked unhappy about this roadblock while I waited patiently for somebody to mention that it was psychologically possible for a doctoral candidate to achieve the minimum mastery in Course A, required for understanding Course B, without necessarily having to sit for 30 hours in a classroom taking notes. Since nobody else mentioned that unorthodox possibility, I proposed it, saying that there were a couple of excellent programmed texts dealing with the material of prerequisite Course A; and, if worse came to worst, it was conceivable that a student who underperformed on a screening examination could be instructed to buy and study books A1 and A2 or even—horrors!—go over to the library

---

[7] [*The University of Minnesota was still using the quarter system when this book was begun. A semester system course would cover 15 weeks. I did not change the number of weeks here in case doing so would impact estimates made elsewhere in the text.—LJY*]

and read them. Not to poke fun at my colleagues, I report this episode to illustrate the automaticity of the pedagogical equivalence "learning about Subject S" = "sitting in a classroom listening to a professor lecture on Subject S."

It is not perhaps surprising that professors of metallurgy or German literature would take this attitude, but I find it odd that professors in the social sciences (and specifically in psychology) have an automated, uncritical equivalence in their thinking about what is, after all, a *psychological* matter. The question whether college students can gain knowledge of a subject matter by other instructional modes than formal classroom lecture is, of course, a researchable question. (Under "knowledge" I include facts, generalizations, definitions, proofs, principles, and problem solving techniques—I set aside "attitudes and values" for the moment.) In fact, this question of learning modes has been exhaustively researched to the point that additional study at this stage would be a waste of taxpayers' money. As early as 1925, Bane (at Minnesota) found no difference between lecture versus discussion in a well done comparison of teaching methods. By 1966 there had been 40 years of research comparing different modes of college instruction. Dubin and Taveggia (1968) reviewed 91 studies comparing traditional classroom lecture, various mixes of lecture and discussion, monitored self-study, unmonitored self-study, quiz sections alone, or discussion sections alone. The conclusion is that there is *no consistent superiority of any one method over another*. This null result cannot be attributed to differential superiority of methods in different subject matter domains, or by different methods of examining achievement. That could be theoretically understandable, but it would generate a heightened non-chance dispersion of means in various investigations (despite the near-zero average difference), and no such inflated variability was found. (This summary antedated the invention of formal meta-analysis, but the authors did a good job of approximating that method.) This gigantic mass of data can be summarized by saying that *there are no differences*, *thus it doesn't matter what you do*. This should not be surprising to a psychologist or for that matter anybody who has been a college student or functioned as a college teacher and reflected on the experience critically. It is apparent to common observation that, given minimal conditions of opportunity to expose one's receptors to informative input, *how much* one learns depends upon brains, motivation, and study habits almost exclusive of anything else. There are, however, two important exceptions to the "no difference" generalization, not available in 1967: educational television and Keller's personalized system of instruction.

Educational television (*ETV*). A special case of formal classroom lecture is educational television (ETV), which can save some money by the easier scheduling logistics even if the presentation is "live," and can save a lot of money if it is videotaped with periodic minor revisions. The research indicates that ETV is at least as effective as the traditional lecture format and probably slightly superior (Johnston, 1987 [calling it visual-based instruction, VBI]; Schramm, 1977). A classic summary of research on ETV before the days of meta-analysis is Chu and Schramm (1967). Examining 207 studies involving 421 comparisons, from elementary school to college level instruction, the old-

fashioned "box score" tally showed 73% of studies ETV equal to live classroom lecturer, 15% ETV superior, 12% traditional superior. Verdict: No robust difference.

The definitive meta-analysis by Cohen, Ebeling, and Kulik (1981) warrants a similar conclusion, although the more sensitive meta-analytic approach reveals a slight superiority (6 percentile points) for ETV. The authors' abstract says:

> This article describes a statistical integration of findings from 74 studies of visual-based college teaching. In the typical study, students learned slightly more from visual-based instruction than from conventional teaching. In the typical study, visual-based instruction had no special effect on course completion, student attitudes, or the correlation between attitude and achievement. Students were equally likely to complete visual-based and conventional classes; their attitudes toward the two kinds of classes were very similar; and aptitude played a strong role in determining student achievement in each kind of class. (p. 26)

Readers unfamiliar with the ETV research should be alerted to a "pessimistic backlash" that sometimes appears in summary statements of the findings. Phrases like "has not held up on further study," "does not fulfill its early promise," "loses much of its appeal on more critical scrutiny" convey the impression that ETV "does not work," that examining the whole body of research shows it to be *inferior* to traditional live-professor-in-room instruction. This is definitely *not* the case, whether we rely on meta-analysis or on conventional narrative summaries of the empirical findings. Considering the mass of studies (in the hundreds), old and new, a skeptical, pessimistic conclusion could be that ETV is no better than the traditional format. The correct inference, as I read the record, is that it is, on the average, somewhat superior. This is what the recent meta-analysts conclude. What has happened here is a common error in evaluation research, due to careless thinking or bias. Whenever a *new method* for doing anything of practical importance—teaching, psychotherapy, drug treatments, surgery, criminal reform, marriage counseling, drunken driving programs, whatever—is invented, some of its proponents are overzealous in advocacy and make exaggerated claims. (It is also likely that enthusiasts' initial zeal about a new, exciting method contributes *objectively* to its potency, an effect that wears out with time, as with new religions, political parties, literary or artistic "movements," and even scientific theories.) When the first flush of enthusiasm fades, and critics take a closer, harder look—some of them having been hostile to the movement all along—the correct "no panacea" diagnosis is tendentiously translated as "no good" or "worse," despite the factual evidence showing "satisfactory" or "about as good as we had before." This is the ETV situation, if read by an unbiased mind. For my purposes here, the summary conclusion that ETV is *at least as good as, and probably somewhat better than, a traditional classroom lecturer* is all I need, since it costs very much less.

ETV and local faculty control. One objection to ETV material produced elsewhere (e.g., by a department at another university, or by a commercial firm like The Teaching Company™), even by educators who accept the pedagogical efficacy of ETV as a general

research finding, is that this takes "intellectual control" from the local faculty over what their department teaches. Someone else gave the lectures I am renting, and maybe I don't accept some of that lecturer's assertions or choices on what to emphasize. While this objection has a certain superficial plausibility, it is silly when closely examined. In advocating the use of non-locally-produced ETV *under certain conditions*, I am of course not suggesting that the local faculty should do so blindly, without any knowledge of what's on the recordings! That is why, in my calculation of the cost, I include one professor's full teaching time giving such a course once, that time to be spent not lecturing but vetting the ETV material. A department that prides itself on some special component of method or substance not commonly found in other departments would presumably opt for either traditional live lecture or locally produced ETV.

Even if certain lectures in an otherwise acceptable non-local ETV course were in need of counterbalancing, that does not mean it would not be useable. For example, our Minnesota clinical psychology faculty has a strong biological orientation toward the major mental disorders, both in emphasizing genetic etiology and employing psycho-physiological and neurological signs as indicators of such conditions as latent schizophrenia. Suppose we decide to focus our graduate faculty's teaching time on advanced courses and seminars, and we are aware that several excellent undergraduate ETV courses in abnormal psychology are available for rent or purchase. Vetting one video, we find it generally satisfactory and in some respects excellent; but the lecturer from another university has made what we consider a somewhat slanted (perhaps even factually inaccurate) comment about a diagnostic sign of schizophrenia, namely, the inability to smoothly track a cyclically moving object (e.g., pendulum). This SPEM (Smooth Pursuit Eye Movement) anomaly had (as of 1993) been replicated in 49 studies around the world in different laboratories, with only one failure to get the effect (Iacono & Grove, 1993; [*see also Calkins, Iacono & Ones, 2008*]). That's a batting average as good as many findings in the exact sciences at the cutting edge of knowledge, and it deserves strong emphasis, both because it is such an accurate sign (more accurate than many of the psychological symptoms which initially defined the schizophrenic syndrome) and because it bears directly upon a neurological interpretation of the disease. The non-local lecturer has in this instance not done his homework carefully, or perhaps has a bias against biological interpretations. While he dutifully describes the SPEM phenomenon, and says it has been found "several times," his summary comment is that it "remains to be validated." This slight distortion, whether from ignorance or bias, we wish to correct. What do we do? We are not pedagogically paralyzed. We take this as a welcome opportunity *to educate the undergraduate students in another respect*, not just about schizophrenia and genetics, but about problems of scientific scholarship and interpretation of data. After they have heard the non-local lecturer's brief discussion of this interesting sign, they then hear a lecture, whether live or on film, by one of our local faculty. She first explains what is wrong with the other lecturer's summary comment, and then goes into some of the common problems of scientific research, such as how it may

be affected by one's ideology or where someone got a PhD. It is obvious that this little flaw in the rented or purchased film is easily correctable, and certainly does not mean we have to reject the whole series of lectures if they are otherwise of good quality. Meanwhile, it provides an opportunity for an important kind of methodological education.

Justifying use of ETV to students. There are two erroneous views about the "typical" undergraduate student that I have found among faculty, each tending to impair an effective challenging of the Sacred Cows. Some professors seem to believe that the typical undergraduate student is highly rational, fair-minded, and moderately well-informed about such matters as the psychology of teaching and learning and achievement testing. The opposite view, found among those professors for whom undergraduate students are essentially a kind of nuisance interfering with one's research and teaching of PhD candidates, is that undergraduates are typically so uninformed and irrational that it's probably a waste of time to try to explain anything halfway complicated to them. These views stem more from personality traits and the dominant values of professors than they do from any attempt at an objective assessment of the undergraduate mind. My view is that most undergraduates are not at all well-informed—why should we expect them to be?—about the research on teaching and learning, and hardly any of them have spent time looking into these matters or even informally reflecting on them. I also think that, for some reason not clear to me, many, perhaps most American undergraduates are characterized by a kind of childishness, superficiality, and lack of general culture that has been noted by European exchange students when they can be persuaded to speak frankly. I don't have a Gallup poll on this and I'm not aware that anybody has studied it statistically, but newspaper accounts and what few conversations I've had with foreign exchange students convince me that the visiting student from Sweden or Germany or Holland finds the American undergraduate a warm, friendly, accepting person, freedom-loving, egalitarian, characterized by good social lubrication and vivacity, all admirable traits. But there is also the observation (almost universal among my faculty friends who came here from Europe in the thirties and forties) that, compared with European students, the typical American student has a kind of intellectual superficiality and an unwarranted confidence in unreflective, uninformed views which to the European mind is simply unscholarliness. Without going into the sociology of this, which is doubtless quite complicated, a consequence is that it is irrational of professors to proceed on the assumption that students will be able to "figure things out for themselves" when it comes to matters like instructional mode, class size, type of examination, or how one leads the life of the intellect. My working hypothesis, formed when I chaired the psychology department, is that the average undergraduate is poorly informed and lacks well-honed intellectual tools, but is *capable of learning* and is *open to clarifications* if they are properly presented.[8] My

[8] There will always be a small fraction of litigious, paranoid injustice collectors who wouldn't like college no matter how things were run, and it is foolish for faculty and administrators to make decisions on the basis that we must meet those people's demands or "fix them up" so they'll be happy. Being a clinical psychologist, I know that nothing can be done with a paranoid injustice collector, even in psychotherapy; that small

remarks will focus on how one should approach the "normal" undergraduate mind.

My colleague Will Grove points out that, despite the research evidence clearly showing that ETV is at least equal, and probably superior, to a live professor in the classroom, this well-established fact will not immediately "satisfy the customer" (student and parents) whose expectations were otherwise. Having no hard data, I am willing to concede Grove's point. But I would proceed on the assumption that although students and their parents are not well-informed, neither are they too stupid, irrational, or rigid to receive reasonable explanations. When televised lectures were used at the University of Minnesota, nobody ever explained to the students during that ETV period just why we were teaching beginning psychology in that fashion. I would address them in a very large room, perhaps even getting access to the old Northrup Auditorium, and set out the objective constraints within which we are operating. I would, therefore, handle the problem of "satisfying the consumer" with regard to, say, ETV in beginning psychology class by *explaining* as follows: "Something like 2,500 students enroll in our beginning psychology course every academic year. Twenty years ago, the beginning course was taught in the huge auditorium that housed the Minneapolis Symphony, with sometimes as many as 1,000 students in the audience! They were jammed together with their Minnesota winter clothing stuffed under the seats and lapboards for taking notes. General psychology has been taught by the major faculty since the department was founded in 1919. We have never turned over the beginning class to teaching assistants or half-time instructors, a practice commonly resented by the consumers, sometimes for good reasons, sometimes not. As you might expect, this was not an easy teaching task and only a couple of our full professors were articulate and dynamic enough to manage it effectively. If the public address system went out the class had to be dismissed. Despite the high quality of the lecturers that were able to hold this group, and the fact that the subject matter of psychology is interesting to many students (even if it's not quite what they expected when they signed up), there was a good deal of disaffection about the physical circumstances. We then shifted to ETV, which worked all right as far as achievement scores were concerned, but was also complained about."

Although I was not myself involved in this ETV undergraduate instruction, I am reliably informed that part of the complaints arose from an unsatisfactory physical setting, which the higher administration failed to correct despite our repeated complaints. The abandonment of ETV was based upon faculty reactions, as well as the dean's, to anecdotal evidence without any showing as to how many students complained how much

---

percentage—probably not more than 1%–2% of the student body—should be ignored. That "works," and it's the only thing that does. Concessions to irrational demands reinforces the paranoid's mute mentation and overt social behaviors. Explaining your reasons to a paranoid usually merely puts you in a defensive posture, and that is almost as reinforcing as concessions to unreasonable demands. If you argue, which nonclinicians might expect to be a punishing experience, you are also reinforcing, because verbal combat is meat and drink to the paranoid. They enjoy a good fight almost as much as they do bending you to their will. The only kind of reinforcement schedule appropriate for litigious, paranoid injustice collectors is to pay no attention to them.

or how we compared on student ratings with other introductory courses (e.g., economics, history, biology) taught in large sections by live lecturer in the room. (It is interesting the extent to which social scientists forget about the unreliability and sampling bias of anecdotes when they move from the laboratory, clinic, or library to practical matters!)

Of course, if I had hard data from student ratings, I would present those data. I would then list the possible alternatives:

1. Conduct a gigantic class using major professors as lecturers in an auditorium not constructed for that kind of use;
2. Turn the instruction over to teaching assistants in smaller sections;
3. Increase many-fold—not just three or four, but more like ten or twenty times—the fraction of parents' taxes that go for university instruction and increase the student fees many-fold so that we can continue to teach in smaller sections with major faculty as the teachers;
4. Greatly restrict the number of students permitted to take beginning psychology, thus a majority of the students would have to leave;
5. Rely, as we intend to do, on ETV.

I am confident that not a single student in the room would be able to concoct an additional way of handling this problem, and I am pretty sure that they would reject most of the options. I would be completely straightforward and make it blindingly clear that we are talking about a logistical problem and an economic problem with severe, inflexible economic constraints. I would say explicitly that *everything we have tried to handle this problem has resulted in some expression of dissatisfaction and that we understand why that would happen*. I would emphasize that we are informing students about the circumstances so they can understand our reasoned judgment. We don't want them to be taught by teaching assistants, or to be crowded into Northrup Auditorium with eight hundred others, or to have their fees and their parents' taxes increased tenfold to hire ten times as many full professors to teach small sections, or to be deprived of the opportunity to take a psychology course. What that leaves us with is ETV. I would then point out that despite ETV not being what they "reasonably expected" when they came to the university, we are pleased to be able to tell them without hiding anything under the rug that *the facts show very clearly that this is an efficient method of instruction*. I would present them with a clear summary of what the research shows. I would not try to conceal that we probably would be driven to this solution by circumstances beyond our or their control even if it were slightly less efficient. But even though we are forced to it by logistics and economics and students' distastes for the alternatives, the universe has been kind to us in that ETV works somewhat better than the kind of teaching they had anticipated. I would point out that in subsequent classes they will have live professors in the classroom who can be interrupted for questions or discussion, but that "traditional" mode of instruction is more appropriate for advanced material. It is my contention that clearly explaining the objective constraints of the situation, and showing (with carefully constructed visual aids

such as bar diagrams of achievement scores) students that they will be not only listening to a lecture in a more comfortable physical setting but that they will learn more and retain it better, would satisfy all but the paranoid injustice collectors (whom we write off as incurable). I would also urge the students to explain these circumstances to their parents and to other taxpayers who may not have any children in college. It would be worth a small investment of money to prepare a brochure that the students could pass on to their parents and others. I am confident that this procedure would work satisfactorily because it is predicated on the two notions that students and parents, while not informed about the circumstances and not aware of the very large body of research that has been conducted on the subject, are basically capable of having things explained if it's done properly, and that the initial impact of "disappointed reasonable expectations" is not a fixed rigid response.

Keller's PSI method. In recent years there has emerged an exception to the "nothing is better than anything else" generalization, and the trend is strong and clear. The "Keller Method," or "Personalized System of Instruction" (PSI), based upon programmed instructional materials and Skinner's technology of sequenced hurdles teaching (see Skinner, 1954, 1984, the latter a powerful article that is relevant here despite its focus on pre-college level instruction; Keller, 1968; Chance, 1984, interviewed Keller about his method; Cook, 1974, provides a good general description of PSI and its advantages), is usually more effective than lecture or lecture plus discussion sections (80% of published studies as of 1977). There are no studies in which it comes out worse. With the Keller PSI Method, there are a few formal classroom lectures available at varying intervals, but attendance is not required and the students are promised that nothing in those lectures will appear on the examinations. The functions of the professorial lecturer in the Keller Method are (a) as a role model and (b) as a "stimulating" motivator, rather than as a transmitter of "information" in the broad sense of the term. No teaching cost is associated with the student working from a programmed text, but there is cost involved if special materials have to be prepared in certain areas where such are not commercially available. (We can rely on the free enterprise system to produce more programmed materials to satisfy large-scale demand, i.e., when Sacred Cow II is effectively challenged by numerous institutions.) After the first round, time required at professorial levels is much less than with the lecture method, assuming faculty were doing the latter conscientiously. The main "teacher," other than in the intermittent lectures, is an undergraduate student (e.g., an honors student) who monitors the class students' progress, because in the Keller Method a student does not go on to the next block of material until he or she has achieved 100% mastery of the previous block. As a psychologist, I would consider it a miracle if the traditional classroom lecturer was as effective as this kind of instruction, there being no theoretical basis for expecting such a thing.

One need not be a "Skinnerian" to see the advantage of programmed texts or similar materials over traditional lecture note-taking. We have good data, going back to pre-

Skinner research by Gage[9] and others in the 1920s, showing the superiority of what used to be called "active recitation" over "passive absorption" such as happens when one listens to a lecturer or reads a traditional textbook in the usual way. "How to study" courses have always emphasized the desirability of active recitation in studying one's notes, for example. The reasons are familiar to psychologists and for that matter can be figured out from common sense. For example, the best way to learn to do something is to practice doing it, so that the production of a response is obviously more similar to the behavior required in the future (e.g., on an examination) than having somebody else make the response while you merely hear it or see it. Second, the active recall method, whether done in the old fashioned way or by the use of programmed texts, teaching machines, or computer software, focuses immediately upon what has not yet been mastered, so the student does not repeatedly go over material that has been mastered. Third, it permits the individual to go at her own pace, and so does not fluster and discourage the learner, nor require putting together units or dealing with higher level concepts before mastering the components. Fourth, the mental set is active, so that a low level of activation, mental distraction, and inattention (common when one listens to an unexciting lecture or reads a textbook) are largely prevented by the task-character of the method.

It is not easy to make precise comparisons over methods without tight control of how many hours of exposure are involved, because in the traditional lecture and textbook study method the student reads assignments as well as attends lectures, and may or may not be reading them attentively, just as he may be dozing or daydreaming through a lecture, none of which happens in a properly run programmed system. But *hour for hour* the programmed material will almost always show not merely a statistically significant but a practically important superiority in what is mastered compared with the traditional lecture situation. It seems clear that, in addition to learning more and retaining it better, students have more favorable attitudes toward courses taught this way.

On present research evidence and anecdotal reports one cannot say how much the pure, unmodified Keller Method (the only one that is, strictly speaking, entitled to be labeled "PSI") saves money. When considering cost effectiveness of an educational procedure, it is arguably legitimate to express it in terms of "dollars per student-knowledge-increment-retained," which is quite different from how many taxpayer dollars it costs per class or section or per head, these latter paying no attention to instructional quality as measured by appropriate outcome criteria. Professor Keller himself (personal communication, 1984) does not claim that it saves money per student, although the research studies show that it costs less per student learning increment. While this is an important point to understand, crucial in exposition to a Board of Regents or a legislature, it is not directly relevant to my general line here, that without loss of educational efficiency or other prime values we can reduce current instruction costs. Other than eliminating courses, or offering them less frequently, or dispensing with professorial contribution

[9] Nathaniel Lees Gage, 1917–2008. See e.g., Gage (1970).

(beyond preparation of a library reading list or selection of a programmed text), the main source of savings when major faculty and professorial rank are engaged in the instruction is how many total professorial hours it requires, *however spent*, to teach the class.

In his initial classic paper on PSI, Professor Keller claimed it requires as much time and energy on the part of the professor as the traditional lecture method. Professor Richard Malott, at Western Michigan University, modified the Keller Method by substituting lower paid personnel such as graduate students to do some of the work. Professor Malott (personal communication, 1984) says, like Professor Keller, that he is not prepared to argue strongly for a net reduction in cost if the outcome quality is neglected. From what he has published and explained to me about the WMU experience, I believe he is being unduly cautious, avoiding overselling his favorite method. In the Malott modification of PSI (which includes some other things besides those I will discuss here, such as mandatory frequent quizzes and less reliance on student ad lib activity) there is a pedagogical hierarchy or chain of supervision. The major professor is at the top of the pyramid and is responsible for planning the course. Under that professor is an advanced graduate student who has had previous familiarity with the process. Next in the hierarchy are teaching assistants, who in turn monitor senior undergraduates who do the firing line testing and discussing with the students. One important fact I learned from Professor Malott's communication about his system is that when he was on sabbatical leave for a year and had only infrequent contact with the advanced graduate student (a part time instructor) who really "ran things," there was no quantitative or anecdotal evidence that the system did not perform as well as when Professor Malott was "on the spot in charge." Nobody acquainted with academia will find this surprising, provided that the professor is wise in his personnel selection. The honors major undergraduates who are in direct contact with the students taking the course (this was a general psychology class of around 300 students) come free, being "paid" by academic credits; and of course the lower echelon teaching assistants come much cheaper than a professor. It doesn't take a detailed cost accounting of the modified Malott procedure to justify arguing that with this hierarchical arrangement the modified Keller Method would result in considerable savings.

Lacking hard data and given variability depending on how the Keller Method is implemented and modified, I am forced to make my own subjective estimate as to how much professorial time could be reduced. It is obvious that by far the largest amount of time must be in preparation of the programmed materials which are essential to the method. I am going to assume that this expense is largely nonrecurring, that the big time investment is in preparation for the first year the course is taught that way. I further assume that (neglecting presently available programmed texts) this professorial time expenditure is no greater than that required in preparing a series of lectures for a course being taught for the first time. This seems a pretty safe assumption; and I add the important point that if numerous colleges began using the Keller Method for a sizeable fraction (although a minority) of their course offerings, many more commercially

available programmed texts would come into being. It is a well-established principle of free-enterprise economics that whenever there is a sizeable effective demand for something, that something will be produced! Programmed material for a wide range of subjects (e.g., logic, contract law, neurology, mathematical statistics, behavior analysis, interviewing, genetics, languages) is available in texts and, increasingly, as computer software. New offerings appear in publishers' catalogues and are advertised in journals of the relevant fields.

I am therefore considering only the recurring professorial time once the course is set up and rolling, rather than the time in the first year, which I am taking to be approximately the same as on the present system, except that it's being spent selecting or preparing programmed materials instead of preparing thirty or sixty hours of lecture. I also assume, on the basis of my communications with Malott and a little common sense, that the time which the professor must spend with the half-time instructor in charge of the teaching assistants (who are in turn monitoring the honors seniors who do the bulk of the supervision) is not greater than presently spent in office hours for students who have questions about lectures. These seem reasonable assumptions to make, and I cannot see why the professorial time thus spent should be more than 25% of what it would be preparing and giving conventional lectures. One lecture per month in a three-month quarter system as "stimulus" (attendance not required) would use only 10% of the present professorial teaching time.

A colleague expert in applied operant behavior theory (Celia W. Gershenson, personal communication, May 14, 1993) agrees with my conjecture that Keller and his PSI disciples downplay the faculty-time-saved feature of the method so as to minimize "political" objections. That this public-relations concern is realistic is shown by the fact that two schools (known to her—there are probably others) abandoned the method, not because of achievement deficit or student complaint, but because of two legislatures' demands that "the professors should teach more hours [of conventional 'stand-up class time']." Obviously this stance, politically and psychologically understandable, is one the state is legally and ethically entitled to adopt. It would not arise under my proposals because, despite use of the Keller Method in some (minority) fraction of courses, 80% of faculty—the pure knowledge transmitters in non-research units—would teach twice as many hours weekly as in other state universities. The numerical paradox of heavier loads per (average) professor but fewer lectures per (average) course is due to the combination of (a) fewer professors with (b) some Keller Method courses. Given this powerful statistic, a college administrator with adequate communication skills should have little trouble defending the (currently "normal") teaching loads of the 20% faculty attached to research units, portions of whose salaries are carefully earmarked in the printed budget as *paid via the research unit*. These special units are justified to the legislature by impressive statistics as to visibility (e.g., prizes, research grants, citations in research literature, major discoveries), concrete evidence of the research unit's actual merit rather than the vague weak appeal "research is a good thing" that we now piously chant

invoking Sacred Cows I and IV. This crucial public acceptance element illustrates my general point that the Cows are interactive, such that we cannot effectively challenge one of them without attending to its (sometimes) crucial relations to the others. That "systems-theoretic" principle applies, as any competent economist, political scientist, or operations analyst will remind you, to almost all complex social phenomena. Of course, a potent selling point for the legislator and the taxpayers is that the total instructional cost is markedly reduced, as are student fees. Who will complain? Nobody.

Given the variations in estimates of professorial time and the reluctance of Keller Method advocates to emphasize fewer hours for political reasons, I will use conservative values in calculating the savings we could realize from it. Lectures at different institutions range from 45 to 55 minute hours; since I am going to be dealing with percent-reduction, that doesn't matter, so I'll simply speak of "hours." Consider a three-credit one-quarter course which at my university would involve three lectures a week for ten weeks. Questionnaire studies show that faculty claim to take typically an hour's preparation time per lecture hour, and that is what colleagues tell me over the lunch table; but I think that is exaggerated, so I'm going to knock it down to half an hour's preparation per hour lecture *on the average*. I have at times gone into class without even a glance at my lecture notes and done a perfectly creditable job, and on other occasions I have spent five or six hours rereading things, reading new things, or preparing some pass-out materials for a single lecture. This variation in preparation time is typical. Of course, I am only considering specific preparation for meeting a particular class, since there is no way to make a reasonable estimate of "general reading to keep up with the field." We are concerned here with college instructional coverage. Using the conservative estimate of a half-hour of preparation per lecture, a traditional lecture course taught this way takes $30 \times 1.5 = 45$ hours of faculty time. Switching to the Keller Method, how much professorial time is required? I am going to have the professor lecture more frequently than Keller originally proposed, delivering an hour lecture every two weeks for purposes of morale and stimulation, with the understanding that students' attendance is voluntary and that this material will not be on the examination. For a course spanning 10 weeks, that gives us $5 \times 1.5 = 7.5$ faculty hours in lecture and preparation. Let us assume that the professor meets an hour a week with the half-time instructor or TA in charge of the controlled learning phase of the course. (Some of these sessions will be attended by the senior honors majors, who do the actual hands-on monitoring and explaining.) I am sure this is more than is necessary, given Professor Malott's experience on sabbatical. That gives us another ten hours of professorial time. Hence the total professorial hours required under the Keller Method is 17.5 hours, compared with the original 45 hours for a conventional lecture course, i.e., the new professorial time is 38% of the old. Since I have assumed the lecture preparation time to be half of what professors normally claim and have arranged for more lectures than Keller advocates, or that are "necessary" for this purpose, and have scheduled more regular TA meetings than needed, I am confident that this 38% residual

faculty time[10] is a conservative estimate, that is, it could be lowered even more if one were economically pressed to achieve a minimum.

It is remarkable that college professors assume the ubiquitous need for formal lectures, considering the extent to which people study things by themselves in their hobbies and other interests and adults routinely continue to learn things in connection with their work. Most college faculty have found it necessary, because of their research activities, to acquire new competences. I have studied genetics, set theory, symbolic logic, economics, and operant behaviorism from programmed texts many years after becoming a full professor. Many full professors of my generation learned computer programming, a complex and difficult skill, in their middle age. I learned enough law from unsupervised reading to teach jurisprudence, enough theology to write 40% of a treatise on its relation to psychology, and enough philosophy of science to publish in refereed journals. Given those personal experiences, why would I assume the equivalence "learning" = "being lectured at"?

The conclusion, then, is that by a suitable mix of alternative instructional modes such as traditional library self-study (supervised or not), leaderless group discussion or group discussion sections led by teaching assistant or senior honors student (both of whom come much cheaper than professors), programmed texts, or computer software and—for certain subject matters—teaching machines, we could reduce instruction cost by a large fraction with no loss (probably an improvement) in student learning and retention.

We may be troubled by the nagging fear that some other valid aims of the higher educational process, such as the inculcation of certain intellectual or social values, may suffer if we were to replace the traditional 3 to 5 hours per week of classroom lecture by some mix of perhaps optional lecture and other modes. We consider this possible despite the research finding that those subject matter achievements that we *can* measure by course examinations or standard achievement tests are more efficiently learned by the Keller Method than by the traditional lecture format, and are equally well learned by other non-lecture modes. I do not dismiss this concern lightly, as I myself view the talented college professor's function as a role model for certain ways of thinking as probably the most important single contribution made in the classroom setting. In saying this, however, I am also aware that the average academic is not very exciting in the classroom. Does any informed person think otherwise? I further assert that there are as many cases of undesirable role models—of shoddy thinking, pompous platitudes, pseudoscience, fake sophistication, correct but dull content, and clever or modish tendentious-

---

[10] [*Meehl had a note to himself to check this figure against a separately concocted table now in Chapter VIII (where the estimate appears as 40%). Throughout the book residual faculty time estimates may vary due to the way the book was written, in segments and over many years, with calculations sometimes made on slightly different assumptions. In the final book he would have made the estimates consistent throughout. I opted to assemble the segments as found, leaving any variations in estimates, first, because they will be small, and second, because their precision doesn't alter the cost-saving message.—LJY*]

ness parading as intellectual depth—as there are models we would want to see emulated. Bypassing that, assume without proof that more often than not the college professor indirectly inculcates values of intellectual integrity, cognitive zeal ("cerebral passion"), concern for justice, and the like in "live" classroom talking and gesturing, by voice and manner as well as by the quality of cerebration and the breadth and depth of knowledge. This concern should be taken seriously, and it should be researched, despite the obvious measurement difficulties (compared with acquisition of facts, principles, or problem-solving techniques). But I submit that one cannot, prior to doing such research, use this argument as heavy artillery against a policy of shifting, *at least in some selected courses*, to less costly instructional methods, particularly PSI which has demonstrated its superior effectiveness. To urge adoption of non-lecture instructional methods for a carefully chosen subset (e.g., 25%) of courses is not tantamount to denying the importance of presenting *good* professorial role models for *some* subject matters at *some* levels. That we value Professor Fisbee's skill with Socratic dialogue (and even, perhaps, aura of sanctity!) in teaching a senior level class on Kantian ethics does not entail that calculus or introductory botany must also be taught by conventional lecture.

While open to the armchair plausibility of this role model influence, one must beware of a tempting fallacy common among academics not trained in statistical method. In this fallacy we may admit that the things we know how to measure, and that we have studied repeatedly in a variety of contexts, *do* show Method A to be more efficient than Method B. But we are comfortable with the familiar, less efficient method, so we conclude that the things that we *can't* measure are the ones in which Method B must be superior! This is not a rational way of proceeding, and puts me in mind of Bertrand Russell's answer to the argument for immortality as making up for the unsatisfactory and unjust character of the present life. He analogizes the believer's argument to that of one who, having sampled some apples at the top of a barrel and finding most of them rotten, draws the inductive inference that in the bottom of the barrel the unseen apples are all delightfully fresh and tasty. It would be as if we said we can measure educational outcomes such as students' acquisition of singular facts, factual generalizations, theoretical concepts, problem solving techniques, or methods of criticizing arguments; but because of the difficulty of measuring internalization of intellectual values and social or moral attitudes, we have not compared our instructional methods on those factors; *therefore* if we were able to study them the superiority of the methods would be reversed. This is not a sensible way to go about filling empty regions of the empirical fact space!

I do not deny some plausibility to the armchair argument that values, sentiments or tastes, and perhaps even "cognitive style," are somewhat more likely to be conveyed "in the flesh." As a psychotherapist I would find it hard to doubt that. Repeating that it should be researched and kept in mind as a valid cause for concern in any large scale liquidation of the lecture format, I offer four interim rebuttals to this objection which I hope will suffice in the present state of knowledge, if not to eliminate, at least to defuse this worry.

The first, which I have just mentioned, is that one should beware of the tendency to infer that because the things we can measure come out one way, then the things we can't measure would doubtless come out the opposite way, more to suit our taste.

Second, it is not being proposed to liquidate lectures throughout, but merely to replace some fraction of them as a means of saving money through reducing classroom lecture time. It is psychologically unplausible to imagine that if the student has, say, 70% as many hours exposure to live professorial role models, such a percentage reduction would have a material effect. A psychologist will confidently presume that the big jump in effect of presenting live role models occurs in the region from zero exposure to some small or moderate amount, with deceleration ("diminishing returns") as we go from a moderate amount to the present 100% conventional lecture time.

Third, it is never a question of all or nothing but a question of amount, of the trade-off involved. It is unwarranted to argue that the decrement in transmission of values or attitudes via the professorial role model brought about by reducing classroom lecture hours would be a greater loss than the presently known *gain in efficiency* in acquiring knowledge of facts, generalizations, and problem solving methods by substituting the Keller Method or related format. Surely the burden of proof is on somebody who says (from the armchair, without data!) that the net loss in modeling values and attitudes will offset the gain from demonstrably more effective teaching of facts, generalizations, and problem solving methods.

Finally, if we did have an accurate method of measuring all of these subtle things, the acceptability of a loss in value transmission for a gain in purely intellectual content and skills would not be an empirical but an axiological question. On the factual level, how much the typical college student (especially undergraduate) is molded by the valuational role models of faculty I do not know, although my anecdotal impression suggests that it is not much. (Incidentally, who ever proved that it's socially desirable to inculcate widespread identification with scholarly role models, let alone donnish life-styles? It is arguable that for most undergraduates, who will not earn their livings as professional scholars and who will have to choose between (a) punching the time-clock and (b) pleasing the public, a strong identification of that sort is undesirable, leading to unrealistic expectations about the world of work. Media accounts and anecdotal impressions lend some support to this.) If we assume, pending research on the subject, that there is a significant amount of worthwhile role modeling that transmits socially desirable intellectual, ethical, and esthetic values and sentiments via the live bodied professor's classroom behavior, *nobody is suggesting that this component of the educational process should be eliminated, but merely that it should be reduced* in the interest of saving money and improving the efficiency with which some other desirable things are transmitted. If adequate research on this question were conducted, a meta-analysis of the outcomes would serve to guide us in deciding within particular departments which subject matter areas are appropriate for fractional change in instructional mode. Here, as always, one

must keep in mind Milton Friedman's TANSTAAFL principle: "There ain't no such thing as a free lunch."

It is obvious that training in research or clinical skills requires an intensive preceptorship relation between professor and student, and I am taking that for granted at the graduate or professional school level. This goes on in research unit departments, as usual. But communicating factual information, generalizations, principles, techniques, or methods is probably done better in other ways, especially at the undergraduate level, and with the availability of video taping and closed circuit television a role model does not have to be present in the classroom each time a lecture is presented. The non-research departments do not have graduate degree candidates, of course, so the intimate research preceptor function does not arise here.

I cannot emphasize too strongly that I am not suggesting that we should save money by lowering the quality of the educational product. It would be rational to permit a slight loss in pedagogical outcomes to save a sufficiently large number of taxpayer and student fee dollars, but the import of the research is that we will not be sacrificing educational quality to save dollars. On the contrary, we will be increasing the flexibility of instructional modes in a direction that increases pedagogical efficiency. There is simply no theoretical reason, no common sense grounds, and no research evidence to say that the only way you can teach college students is to herd them into a room where they listen to a professor talk. I have sometimes thought that presenting several hundred quantitative investigations probably has less influence in eroding pious belief in this Sacred Cow than would regularly requiring professors to look at a few pages of student notes from a lecture they just gave, as I have done on occasion when a student inadvertently left a notebook behind. I cannot think that anybody who has done this even a few times could remain firmly committed to Sacred Cow II.

Additional References for Sacred Cow II, but not discussed in text:

Dubin, R., & Hedley, R. A. (1969). *The medium may be related to the message: College instruction by TV*. Eugene, OR: University of Oregon Press.

Friedman, C. P., Hirschi, S., Parlett, M., & Taylor, E. F. (1976). The rise and fall of PSI in physics at MIT. *American Journal of Physics*, *44*, 204-211. [PSI did not "fail" at MIT, by usual standards. It was *perceived* to fail, plus a bad logistic error in first big use. A *social psychology*, sacred cow problem!!]

Jamison, D., Suppes, P., & Wells, S. (1974). The effectiveness of alternative instructional media: A survey. *Review of Educational Research*, *44*, 1-67. [Tables show ETV ≥ traditional lecture, all subject matters, all levels through college. Very clear, numerous studies. (Most tables from Chu & Schram, 1967)]

Johnson, K. R., & Ruskin, R. S. (1977). *Behavioral instruction: An evaluative review*. Washington, DC: American Psychological Association.

Sherman, J. G. (1974). (Ed.) *Personalized system of instruction: 41 germinal papers*. Menlo Park, CA: W.A. Benjamin. [Includes Keller's "Goodbye teacher…." Example areas using PSI

include Psych, Engineering, Physics, Biology, Chemistry, Statistics, Philosophy, Library Science.]

Siegel, L., Adams, J. F., & Macomber, F. G. (1960). Retention of subject matter as a function of large group instructional procedures. *Journal of Educational Psychology*, *51*, 9–13. [Retention 1 or 2 years later, no difference between TV, large class, small classes. 9 subjects, nice spread (eg, psychology, air science, geography, social studies, English). This reference is also for chapter III.]

Taveggia, T. C. (1976). Personalized instruction: A summary of comparative research, 1967–1974. *American Journal of Physics*, *44*, 1028-1033.

Sacred Cow III

When classroom lecture is the method of instruction, small classes are much preferable to big ones over the whole range of sizes and subject matters.

The research evidence on the effects of class size on students' learning at the college level is somewhat mixed and difficult to evaluate if done by conventional narrative summary, but fortunately we have a better way to do it: *meta-analysis*. In this more sophisticated and objective approach, we summarize numerous (sometimes superficially "conflicting") research studies to determine the *typical quantitative influence* of a purported causal factor, its *effect size* (ES), then the *variation* of its ES over studies, and finally the *contextual factors* that are systematically correlated with ES. What seems warranted from meta-analysis is that the influence of class size on achievement test scores is variable and becomes practically unimportant somewhere in the region $25 < N < 30$.

The path-breaking study by Hudelson (1928) led to the conclusion that at college level class size was of negligible importance. But a more recent survey of research by Glass, Cahen, Smith, and Filby (1982), employing the more sophisticated meta-analysis method of review (Glass, McGaw, & Smith, 1981) and based on several hundred studies, reveals an unmistakable class size effect. Disaggregating studies by total instructional hours, their graph (Glass *et al.*, 1982, Figure 2.1 at p. 49) for fewer than 100 hours (the upper limit for almost all college courses) suggests that for classes size > 20 further increase produces negligible average decrements in measured achievement. *Because of the pronounced deceleration of the graph, it is misleading to use the phrase* "*class size*" *at the college level without saying what range of practicable sizes one is considering*. If a university could afford to consider having undergraduate classes of, say, a dozen students rather than three dozen, the Glass *et al.* meta-analysis provides fairly strong evidence for preferring the smaller, in an average expectable sense of better achievement scores. (Although, I cannot forbear to mention that even in this region of non-linear improvement, the 12-student advantage amounts to a modest 7–8 percentile units.) At my university, for most courses it is economically and logistically unfeasible to contemplate holding class size to a dozen students, the realistic alternatives being more like 25 versus 75, or 100 versus 300. Having myself taken (in 1938!) an excellent introductory psychology course in a class of 250, I have never been horrified by classes of that size or even larger. My beginning courses in zoology, chemistry, physics, and sociology were of size 150 or more, all superbly taught (by full professors). A gifted colleague, Kenneth MacCorquodale, lectured for several years to introductory psychology classes of more than 1000 students. He was one of the most esteemed teachers on our faculty and won an

award for excellence in undergraduate teaching. (Let us not forget John Wesley, who inaugurated a great religious revival by preaching in the fields to thousands of listeners!)

Empirical investigation of the impact of class size is difficult partly because of disagreements about the pedagogical aim and about the measure of achievement that should be used. The influence of class size has been investigated both with respect to student scores on achievement examinations and, in recent years, increased emphasis upon students' reported attitudes to the class, subject, and instructor. Overall, there is a slight tendency for smaller classes to come out a little better on both of these kinds of outcome measures, but it is not consistent over studies; and for student attitudes, a sizeable minority of studies show a non-monotone relation, that is, very small and very large classes do better than middle sized ones. The contamination of all such results by variance due to nuisance factors (e.g., departments using large classes having a tendency to use major distinguished professors as instructors in such classes) makes it almost impossible to disentangle the influences. When I was a student, I preferred attending lectures by a seasoned professor who was a distinguished scholar to lectures by an instructor or teaching assistant, even if the former involved a class size 5 or 10 times as large. I would be surprised if many readers report contrary introspections. The average correlation coefficient between class size and student "global" affective reaction to the course is only –.09 (Feldman, 1984), which tells us that, typically, 99% of the variation in student attitudinal response to a course is independent of class size, over the whole range.

A common belief among social science faculty is the assumption that you cannot conduct any sort of discussion except in very small classes, what we would usually think of as "seminar size," say, 10 or at most 20 students. Having taken and instructed courses in law school, I can testify that this is just plain false. In the law school I have participated in extended and intensive discussions of technical and recondite constitutional or philosophical issues conducted in classes of 150 students. The widespread unquestioned belief that this is impossible has presumably influenced some fraction of the empirical studies of achievement and attitude (self-fulfilling prophecy). Furthermore, many (most?) students have been brainwashed by peers, parents, news media, and even faculty to mouth, "It's terrible at my university, we have such huge classes, you're not an individual, you're just swallowed up, you have no identity as a person [blah, blah…]." Thus a mess of irrational attitudes and fuzzy ideas about "personal identity" gets scrambled into a substitute for rational assessment, and student attitude ratings (even, conceivably, student performance on examinations) will be partly reflective of these dogmas. I suppose one psychological source of my own undergraduate contentment with lectures by top-flight scholarly professors in classes of 100+ was the mental-hygienic fact that I held no silly delusions about a university's "giving me an identity" or the faculty's obligation to "love me as a person." (People are usually helped by clear thinking about the nature, functions, powers, and limitations of a social institution they deal with.)

Finally, we confront the usual problem of cost and benefits. Suppose a sufficiently careful investigation, or a sufficiently ingenious meta-analysis of the studies already

conducted, should reveal a slight superiority (say, 5% or 10%) in final examination scores of students in small classes over students in middle size or huge classes. If the main reason for dividing a class of 300 beginning psychology students into 12 sections of 25 each was this hoped-for increment in achievement, *and reminding ourselves that a 5% or 10% difference on the final exam would probably be attenuated to a considerably smaller or negligible difference after a year or two has passed*, would it be rational to substitute teaching assistants for full professors? Would it be sensible to have major faculty teaching 12 times as many high cost hours to achieve this small gain? Or should we increase faculty size 10-fold (utterly out of the question, anyway)? That would be hard to justify, given the economic pressures of higher education in which we're operating.

If the main concern here is not how much is learned (and ultimately retained) but how the students *feel* about the institution's attitude toward them and concern for them personally, the rational response is that an institution of higher learning (such as my own with more than 50,000 students and over 3,000 faculty) is simply not a sensible place to look for warm interpersonal gratifications! If you encourage people to have irrational expectations and demands, of course you will pay the price in that they will confront you with failure to meet them. A student who says, "I don't like the University of Minnesota because it's too big, I don't feel that the professors love me as an individual person," is like somebody complaining in a service station that there aren't fresh oysters on the half-shell being dispensed, or in a London haberdashery that there isn't any beer and pizza. One effective response to such childish complaints would be, "Well, why don't you and your three dorm roommates draw straws to decide which three drop out and go home?" Another would be, "Is your family willing to have their taxes doubled to solve this problem?" One should not encourage people who are old enough to vote (or start a business, or fight in a war, or marry without parental consent, or be convicted of murder) in childish attitudes that fail to ask what kinds of goods and services are dispensed by what kinds of social institutions.

Neglecting the parametric aspects of the class size problem (arguing qualitatively that "smaller is better") recently led to some unfortunate political consequences at the University of Minnesota. A brilliant, energetic, committed president and his entourage, aiming at academic excellence, fell into the habit of mentioning that one factor conducive to excellence would be reduction of the student body by some 8,000 persons. This prediction disturbed some legislators and Regents; taxpayers (influenced by tendentious journalism) saw it as elitist and unfair; outstate citizens were asking, "Does this mean my kid won't be admitted to our state university?" It was clumsy public relations and a wholly needless affront, because it made little sense economically or pedagogically. Economically, a loss of 8,000 students' fees is, of course, unaccompanied by anything like that much reduction in instructional costs (see Chapter VIII). The idea was that smaller numbers would presumably improve the quality of the educational experience. Would it? Hardly. Reducing the University of Minnesota's student body of some 56,000

by 8,000 is a decrement of 14%. Taking the average class size in liberal arts to be around 28, a reduction factor of ⅐ lowers it to 24, four fewer per class. Using the meta-analysis of Glass *et al.* (1982, Figure 2.1, p. 49), we can predict a probable improvement in achievement of less than 1%, which (if realized) would doubtless wash out a year or two after the final examination. Query whether the administrators involved had ever examined the empirical evidence on this question? I suspect not.

The "Big Educational Mill" Complaint

An anonymous professor, having seen an early draft of this book, is quoted to me by a colleague as complaining that "Meehl wants to turn the place into nothing but a big educational mill." This piece of "argument" by word-magic is a nice illustration of my contention that people with PhDs sometimes don't think any better than Archie Bunker, although they can "thob" (H. G. Wells's neologism for pseudo-thinking) with a superior vocabulary. Not having the critic available, I am not clear what the objection consists of, assuming some rational content underlies the rhetoric. I suppose the first thing to look at is sheer size—the number of professors and the number of students. If that's the point, *that is what we have at my university right no*w. The only effect that carrying out my proposals would have on that statistic is a reduction in size, both faculty and student body. So this can hardly be what the objector is complaining about, unless he hasn't read all of the book or couldn't follow it.

A more likely variant of this worry is the standard harping on the ratio of students to faculty, the unquestioned postulate being that if that ratio is large, things must be terrible. I know of no research evidence supporting that notion, although it is of course a handy line taken by administrators with legislators as a good-sounding ground for being given more tax dollars. If I parse the student:faculty ratio complaint in what I hope is a fair and rational manner, what I come up with is (a) the class size problem, (b) the relative rarity of one-to-one intellectual contacts as in the tutorial system of the great English universities, Oxford and Cambridge, and (c) the general aura or flavor of "impersonality" in big institutions that some persons find offensive.

As to class size, I reply by citing the research evidence that shows the achievement curve to be markedly decelerated when you reach classes larger than twenty or twenty-five students. There is no significant achievement difference between a class size of 25 and one of 250. Since nobody rationally thinks that we are going to get enough money from the taxpayer to keep our classes down to twenty or fewer students, complaining of this has about the status of a complaint about human mortality or the fact of earthquakes, because there is nothing conceivable we can do about it.

However, that I argue in favor of combining many class sections to reduce professorial hours and do not fret about the *existence* of classes of 100–300 students does not of course imply that there should be *no* small seminar-style classes available. Nowhere do I say or imply that. The savings achieved by implementing my proposals will bring about a

marked reduction in usual classroom hours and—by challenging the research Sacred Cow—makes a large majority of faculty available for at least twice as many classroom teaching hours as they now have at research-oriented institutions. There is no reason why some portion of those freed-up hours should not be allocated to create a carefully-chosen set of seminar-size courses, even for undergraduates, with the assumption that the faculty will also be chosen carefully. It is simply a matter of tradeoff of personnel time. From the statement, "It doesn't hurt to have large classes in most kinds of undergraduate college courses," one cannot validly infer, "There should not be any small-size classes at the undergraduate level." To move from one to the other is simply a fallacy in reasoning.

With respect to the relative scarcity of one-on-one tutorial instruction, changing this appreciably for most undergraduates at a very large university is not a live option economically—just look at the numbers—unless you are prepared to send the great majority of students back to the farm. But here again, it should not be assumed, and I have nowhere said or implied, that no portion of the faculty hours saved (which will be not some little 10% or 15%, but well over 50%, and conceivably as high as 80%) should be reallocated to having a carefully selected subgroup of professors who are sincerely interested in being available for the Oxford-Cambridge kind of once-a-week tutorial session. I have done some rough calculations on this after consultation with our associate chair, and have concluded that it would be feasible to impose a mandatory requirement of a year of such once-a-week tutorial sessions for all our undergraduate honors majors by allotting only a small portion of the saved professorial hours to this special function.

If it is maintained that every student, regardless of major, ability, and interests should have such a tutorial experience, I would be curious as to what armchair argument or empirical data could be brought to support such a statement. Most students are *not interested* in doing this. Our experience is that even senior honors majors, of very high ability, many intending to go on to graduate work or professional school after the BA, do not take the opportunity afforded by listed courses of directed study, although we do not know why. That being the social fact, why would we insist that students of middling ability and feeble intellectual interests should be coerced into talking with a grudging, bored professor for an hour every week? Anecdotal evidence (e.g., from Rhodes scholars) and academics' fictional works[11] suggest that the famed Oxbridge tutorials are not all they're conventionally cracked up to be. Until somebody brings proof that mandatory, universal tutorials for all undergraduates is a sound idea, I am not going to worry about its being economically impossible.

If the complaint is about the general "impersonality" of large institutions, that is an intrinsic fact about them whether it is the army or the church or trade unions or whatever, and unless genetic engineering can alter the facts of primate ethology and we can repeal the principles of economics, I am not going to worry about that complaint, either. As I've stated elsewhere in the book, if people come to social institutions with irrational, unreal-

[11] [*Readers who enjoy fiction about academia can find titles from Meehl's reading list in the Appendix.— LJY*]

istic expectations as to their proper functions, they will be disappointed. The cure for that is not to attempt the impossible, given the size and functions of the institutions and the available dollars, but rather to disabuse people of childish ideas and expectations.

Finally, I suppose the objector may be saying that all these "superficial" improvements in sheer efficiency of communicating knowledge will not tend to make most students passionate about leading the life of the mind. With this statement I cheerfully agree, *but that is also true of the prevailing methods*. Most students are not either capable of or interested in leading the life of the mind, and the obsession that you must somehow seduce or compel them to do so is simply a case of professorial insistence that other people should have the same values as academics do. My saying that one-on-one tutorials are mainly important for honors students, of course, subjects me to the charge of elitism. My reply is two-fold. First, it is not elitist to accept facts about human behavior; second, one of the most malignant forms of elitism is trying to make everybody else like oneself. Knowing the fact that only a small minority of honors majors in psychology take advantage of the opportunity for one-on-one Oxford-Cambridge tutorial instruction, and keeping in mind my colleague Fox's observation about the two worlds (see Chapter IX), I do not fret about confining tutorials to honors majors. If someone sees this not as a question of behavioral engineering but a matter of justice—which it simply isn't—I will not object to opening these tutorials to everybody who is getting passing grades in the major, being confident that only a minuscule number of non-honors students will take advantage of it.

Additional Reference for Sacred Cow III, but not to be discussed in text:

Siegel, L., Adams, J. F., & Macomber, F. G. (1960). Retention of subject matter as a function of large group instructional procedures. *Journal of Educational Psychology*, *51*, 9-13. [Retention 1 or 2 years later, no difference between TV, large class, small classes. 9 subjects, nice spread (eg, psychology, air science, geography, social studies, English). This reference is also for Sacred Cow II.]

Sacred Cow IV
Most research published by college professors is worthwhile.

One's position on this question will be a subjective judgment reflecting knowledge, breadth, and depth of acquaintance with one's own and other disciplines; historical perspective (some contemporary academics are remarkably ignorant of their discipline's history and as a result, in the social sciences at least, are continually re-inventing the wheel); and subtle matters of group sentiment, trade unionism, and the extent to which one's self concept and sense of personal worth involves group identification. We might expect the strongest influence on personal attitudes toward scholarly publication to be one's own scholarly visibility, but I am unaware of any quantitative data on this, and my anecdotal evidence does not suggest any clear, strong correlation. I cannot present compelling data bearing directly on this question, comparable to what can be offered concerning the influence of class size or the pedagogical efficacy of different instructional procedures. Nevertheless, it would be a mistake to treat this question as to the long-term merit—intellectual or social—of the average college professor's writings as if it were wholly a matter of taste, personal values, or subjective prophecy. Whether someone's writing has an influence on other thinkers is, after all, not beyond empirical investigation. I can only mention some lines of investigation and reflection, relying on the reader's intelligence and objectivity to determine whether to conclude, as I do, that by far the larger part of "publication yardage" by college professors, at least in liberal arts colleges, is of no lasting importance and could just as well not have been written except insofar as it contributes to the individual's achievement of salary raises and academic tenure. I can't help it if that's a cliché we hear from cynical students and unsuccessful faculty, it is nonetheless true.

I am influenced by five considerations in arriving at this melancholy conclusion. These five are easily accessible to my readers, and it would be foolish to say which of them deserves greater weight. Fortunately, I needn't assign relative weights as to their reliability or probative value; inasmuch as all five lead to the same conclusion, we need not balance them out as we would if their deliverances were in opposition. As a social scientist I naturally have a predilection for the more "objective" and "quantitative" grounds, but I do not try to impose that preference upon the reader.

The first and most subjective basis—but, I confess, one of the most compelling—is simply perusing a recent issue of almost any current refereed journal for my field (clinical psychology). I ask myself whether most of the material is interesting? scientifically convincing? highly likely to replicate? and especially whether it is a contribution of such *quality*, and the effects reported are of such *magnitude*, that the long-term history of psychology in general, or my own specialty, would be appreciably different if a particular paper had never been published? The usual answer is, "No, it would not." There is, alas

for conventional modesty, no way to avoid assigning weight to my own judgment in partial reliance upon the fact that I am a distinguished member of my profession (with a pretty good publication yardage record) and have received numerous prizes and professional awards for scholarly contribution. A statistical count of references to my work in the *Social Sciences Citation Index* tallies over 770 citations in the five year period 1986–1990, a "communication impact" that puts me in the top $^{1}/_{10}$ of 1% of *cited* psychologists. My name would be known to PhD candidates studying clinical psychology, psychometrics, or behavior genetics—and even (older) students of animal learning —at universities around the world. There's just no honest way for a person examining this sort of impressionistic, judgmental matter to escape the egocentric question, "Well, am I competent enough so that it would be sensible for people to pay any attention to a pessimistic summary by me?" And that unavoidably includes whether *I myself* should pay attention to me!

Second, I have put the question, "Approximately what percentage of articles published in the psychological journals you read do you think will have any significant impact upon the state of the field 25 or 50 years from now?" to a subset of my colleagues in the Minnesota psychology department who are perceived by one another as "superstars" in terms of brains (I employed two selection systems whose details I cannot ethically reveal, and they showed a remarkable convergence). Their estimates of the percentage of worthwhile papers vary from less than 1% to 10%, with a median of 5% (my own estimate being approximately 10%). Here again, one cannot avoid judging the judges. The Minnesota psychology department enjoys a reputation in various national rating systems (from the mid-1920s to the present time) as among the top half-dozen in the country. In three fields it is in first place, and it would be in another if ratings were made for that area. The faculty's track record, whether of citations by others or in terms of sheer publication yardage and various prizes and professional officerships, is excellent. One may reasonably hold that the opinions of its "superbrights" (roughly the ablest fifth, as anonymously starred with high consensus) in such a top flight department should be given appreciable weight, as being a judgment neither uninformed, frivolous, or "sour grapes" in motivation.

Third, a historical perspective—still relying on subjective impression, but one that I find helps some persons to take my pessimism seriously—is to take at random a bound copy of a major refereed journal in the field 25 or 50 years back and read through it. It is disconcerting to find that most of the names are unrecognized, most of the theories have not importance or relevance, and one doesn't even know whether a certain controversy was settled in one way or the other (because scholars are not talking about the subject any more!). The impact of such scanning of the older literature is so powerful that I have sometimes wondered whether it would be ethical to recommend it to graduate students because of its possible bad effect upon morale. As leading philosopher of biological science David Hull put it: "Most research programs progress for a while and then disappear, leaving a literature as calcified and causally inert as the fossils of extinct species.

The scientific process comes close to being as wasteful as its biological counterpart" (1988, p. 372).

Fourth, the opinions of eminent historians and sociologists of science are, while perhaps not uniform, clearly in this direction. While I am not an expert in the field of scientific communication networks or information retrieval—fascinating disciplines which turn out to be conceptually richer than one might have supposed (and which have technical mathematical aspects that I shall not consider here)—I make bold to draw a general conclusion from my sampling of it and reading of some of its main summary treatises. The historians, sociologists, statisticians, and communications experts who do possess technical competence show little disagreement that, so far as the sciences are concerned (physical, biological, and social), by far the larger part of publications "do not really count"; in any long run sense of impact upon other researchers and the shared corpus of methods and beliefs, they could just as well never have been published. Only a small fraction of scientists account for by far the larger portion of publications that make a detectable difference to the field.

Fifth, we have available many quantitative studies of the research literature, conducted by documentation scientists and historians and sociologists of science supporting the two big generalizations:

1. The distribution of publication productivity is extremely skew, so that a small minority of scientists generate a large preponderance of papers.
2. There is a strong correlation between quality and quantity of productivity, so that the important papers having a large and lasting impact and often cited by others tend to be authored by the scientists with many publications.

"In the early decades of [the twentieth] century, two different views on the contributions of scientists to the growth of scientific knowledge were suggested: one that a small percentage of scientists are responsible for most progress in science, the other that all scientists contribute their bit even though only a few scientists gain much in the way of recognition—Latka's law and the Ortega hypothesis, respectively…. As reassuring as the Ortega hypothesis may be, all the evidence points in the other direction" Hull, 1998, p.359). A half-century of quantitative research by historians and sociologists of science, documentation and communication scientists, and individual scientists in various fields interested in the history and processes of their own disciplines, has convinced such scholars that significant lasting contributions are almost wholly attributable to a small minority of scientists. Widespread belief to the contrary among academicians is presumably due to ideological commitments, self-concept ("identity") needs, and sheer ignorance of the research. I find by informal polling of colleagues that most of them not only don't know the substance of this research but are not even aware that it exists or the names of its distinguished contributors. I have come reluctantly to accept the conclusions of Price, the Coles, Garfield, Merton, and Co., because their statistics are persuasive and because they accord with my anecdotal impressions and the cynical judgments of my

first-rate colleagues. For a sophisticated discussion of several quantitative approaches to evaluating scientists' contribution via statistics of the scientific communication network (e.g., half-life of article citations, proportion of journal issues requested by library users) see Price (1986). Quite apart from assessments of a scientific paper's merit (citations, half-life, rated importance), the productivity data show that a small proportion of (publishing) scientists are responsible for the majority of papers.

In early research on the question, Dennis (1954) examined Lotka-like tables for publications in American psychology beginning in 1887 and found that most were produced by very few persons. For his earliest subgroup, publications of the high-producers also were cited most often in selected handbooks and advanced textbooks in psychology. Dennis (1955) got the same result for other fields (music, books, gerontology and geriatrics, geology, infantile paralysis, chemistry, and linguistics): "[T]he majority of creative contributions is made by a minority of the contributors."

The distribution of publications by author is well-fitted by Lotka's Law (Price, 1986, p. 37*ff*), that the number of scientists who in their lifetime publish n papers is inversely proportional to n^2. This fits archival data going from the first indexed volume of the *Transactions of the Royal Society* to contemporary publications and listed in chemical abstracts. "For every 100 authors who produce but a single paper in a certain period, there are 25 [who produce] 2, 11 with 3 [papers], and so on" (p. 38). Out of 100 scientists, the top 2 produce one-quarter of all papers. The average number of papers per author comes to around 3.5, constructed on the basis of exactly 100 authors who have at least one published paper. (The striking unfamiliarity of most high producers with the statistical facts is illustrated by a first-rate colleague's lament that in his exacting, time-consuming research area, he cannot expect to average more than one experimental paper annually. He was astonished when I pointed out that this output would come out to around thirteen times the typical scientist's lifetime work product.) Using Price's example of authors of various degrees of productivity (Table 2.2, p. 41), a cumulative distribution shows about $\frac{4}{5}$ of publications produced by $\frac{1}{5}$ of the authors. Half the work is done by scientists with more than 10 papers to their credit. When you get into the high-scoring group, Lotka's Law doesn't fit as well, "the number of people falling off more rapidly than the inverse square, or approximating the inverse cube" (p. 42).

Turning to citations, rather than sheer number of publications, Price (1986) estimates that, on the average, we should expect that every scientific paper ever published is cited about once a year. However, 84% of published papers are either cited not at all (35%), or only once (49%) (p. 105). The distribution of citations and their "half-life" (citation frequency as a function of time) suggests a typology of papers classifying them into "classic" and "ephemeral," the great preponderance being ephemeral (Burton & Kebler, 1960; Price, 1986, pp. 113-114). The active citation life of articles follows an exponential type function for nine physical and biological sciences, half-lives ranging from about 4 to 12 years with 95% obsolescence by 4 decades. "I am tempted to conclude that a very

large fraction of the alleged 35,000 journals now current must be reckoned as merely a distant background noise, and as very far from central or strategic in any of the knitted strips from which the cloth of science is woven" (Price, 1986, p. 118).

Price discusses the "invisible college" of scientists who exchange less formal information with one another, a phenomenon that occurred in the 1960s with a clear example from a special branch of chemistry, and which today has been greatly increased by e-mail and Internet. His example for chemistry illustrates the point that there is a research front dominated by a small core of authors, with a "large and weak transient population of collaborators." 56% of scientists can be classified as transients because they publish only once (Price, p. 211). From another database, it is found that Lotka's and Price's theoretical laws estimate single paper authors as 61% [Lotka] or 53% [Price], but the empirical data show 67% [demographic model] (Price, p. 223). When we move from publications to citations, we find 71% of scientists cited not at all, and 19% cited once, so that around 90% of scientists would be classified as "transient" using Price's standard (Price, p. 230). Price found only a single case of an author who had published as many as 10 papers, none of which were cited.

The distribution of numbers of authors cited (one year, two years, and so on, up to all seven years in a seven-year span) is U-shaped, with a heavy piling up of authors cited only one year in the period, and at the high end, another piling up of authors cited in all seven years.

An excellent discussion of *Science Citation Index* counts in a social science, sociology, is by Jonathan and Stephen Cole (1971). It provides powerful construct validation of the *Science Citation Index*[12] as a measure of the quality of sociological research, and also does a good and fair-minded job dealing with some of the worries that appropriately arise, even among non-obscurantists. I shall mention briefly some of the interesting statistics that they present. They quote my former colleague Kenneth E. Clark who found a correlation of $r = .67$ between citation count of psychologists and how many times they were named by a panel of experts as having made the most significant contributions in their field. Clark (1957) concluded that the citation count was the best available indicator of the "worth" of research work by psychologists. Nobel laureates may be safely assumed to have made significant contributions to science, and in 1961, the number of citations to their work was 58, compared to an average of 5.5 for other scientists. Only about 1% of scientists who appear at all in the *SCI* receive 58 or more citations. Among university physicists, the number of awards received, which, of course, is a number having a fairly small range at best, still correlates $r = .64$ with a citation index. Only 2% of academic physicists have 60 or more references to their work in a year, 12% between 15 and 59 citations, and 86% have fewer than 15 citations. If, for a

---

[12] [*The concept of science citation indexing was introduced by Garfield in 1955* (*see, e.g., Garfield, 1970*). *A good account of the history of the* Science Citation Index *and subsequent developments is given by Pendlebury* (*2013*) *at* http://garfield.library.upenn.edu/papers/pendleburykingresearchfronts2013.pdf —*LJY*]

given scientist, one considers only his or her three most frequently cited contributions, a measure that cannot be an artifact of quantity of publications, the correlation between the number of papers the scientist writes and the number of citations to those three is $r = .72$. Strangely enough, at least in physics, there is no relationship between the size of the specialty and the number of citations to the work of physicists in that specialty. That is reassuring, although I doubt very much that it would hold true in the field of psychology. It should be investigated, since it's one of the worries that people have quite rationally expressed (cf. the example of psychologist Carl Pfaffmann, p. 141, discussion of evaluating "Research Productivity"). Weighting citations for an author's age would seem to be substantively necessary. There is a correlation of .80 between the total of weighted and unweighted citations among physicists, and if we compared weighted and unweighted citations to physicists' publications in their three "best" years, we get a correlation of $r = .96$.

This question of the worth of most published research, while of some delicacy and having certain statistical pitfalls, can be approached from several angles; but while the different kinds of statistics computed are not equivalent or mutually translatable, they add up to a coherent picture that can be summarized as above. I give only some of the better known striking examples: Considering single articles in refereed psychological journals, it turns out that any one article chosen randomly will be read by less than 1% of the readership of these journals (American Psychological Association [APA],[13] 1963).

Studies in several fields of knowledge have indicated that around a third of published papers are not subsequently cited by anybody. Except for the very rare "sleeper" paper, non-citation in the first post-publication decade means near certainty of never being cited at all. Among publications which are ever cited by anybody, the half-life of their citations averages in the neighborhood of 7–10 years, and (surprisingly) there is no great variation among scientific disciplines in this respect (Burton & Kebler, 1960). Any scientist whose article is still being cited, say, 25 years after its publication is in the top decile among authors in the profession *who are cited at all*. Statistical studies of library records on use of bound copies of journals that are, say, 10 or 20 years old, show that in any one year the overwhelming preponderance of such bound copies are not checked out by so much as a single user. Several research studies in different fields have shown that the top 5–10% of scientists who publish contribute almost all of what has any discernible impact on the field.

A more recent study by Pendlebury (reported by Hamilton, 1991; see also Anderson, 1992) agrees with previous findings of citation rates lower than commonly assumed, and further reveals huge differences among scholarly disciplines. For example, approximately

---

[13] [*W. D. Garvey & B. C. Griffith directed a research program on scientific information exchange in psychology that was published in three volumes by APA* (*1962, 1965, 1969*)*; a number of additional publications included Garvey & Griffith* (*1971*) *and Garvey* (*1979*).*—LJY*]

37% of physics articles go uncited, compared with 41% in biological sciences, 75% in social sciences, and 98% in arts and humanities.[14]

One cannot, of course, translate this into "percentage of article-publishing scholars whose articles have more than ephemeral character," but data on that second question seem to yield about the same answer, to wit, 10% or less (the rough proportion one often sees and hears mentioned by historians and sociologists of science). Myers (1970), employing an extremely tolerant index of visibility/impact computed from the *Science Citation Index* and beautifully validated at the high-visible end against multiple non-citation criteria of eminence (election to honorary or restricted membership societies, prizes, medals and awards, election to society presidencies, etc.), found that only 6% of psychologists do work of this quality and influence. Since his sample consisted *only* of psychologists whose work was cited at least once, and since we know that about ⅓ of published articles are never cited by anyone, these findings strongly support the view that *if the majority of psychologists who are now writing papers were to quit doing so, nothing much would happen to the discipline's long-term progress*.

Mere references to a paper (as reported, e.g., in *Science Citation Index*) can hardly be taken as proving that the paper contributed materially to the advancement of knowledge, that it was a necessary or even helpful "brick" in the Great Cognitive Edifice. Scholarly etiquette requires one to have done a "literature search" and to report its seeming deliverances, however inconclusive they are or non-contributory to one's own thought or experimentation. My personal experience as an author is anecdotally confirmed by colleagues, that if Fisbee has written on Topic X, one knows there's a high probability that he (or a student of his) will serve as referee of any manuscript on Topic X. Realizing that the journal editor will insist that I cite Fisbee, I put him in my references, regardless of how I evaluate the work or whether it influenced my theoretical thinking, my experimental design, my statistical analysis, or my interpretation of the results. Authors often feel compelled to cite papers they had not even read until their manuscript was substantially completed. The irrelevancy of most dutifully cited papers is obvious to a skeptical reader once alerted to the point, the telltale sign in social science being the quick-and-easy mere *listing*, without discussion, of names and dates: "the Schreck phenomenon has been studied by Fisbee (1908), Seidlitz (1911), Fisbee and Hockheimer (1977), [etc., etc.] with equivocal results." A reader who bothers to go back to these older sources often finds that the citing author largely ignored them in conducting and interpreting the current study. Those that *mattered* to the author are summarized or discussed in detail. First-rate contributors in the social sciences can testify from introspection that the *de rigueur* "review of literature" is more often than not a boring, pedantic, unhelpful feature of the scholarly life. When reviewing memory and verbal learning literature for the *Annual Review of Psychology*, Tulving and Madigan (1970, pp.

---

[14] [*Non-citation percentages, while always depressingly high, do shift with time spans and specific data-bases used. More recent figures may be found on the Internet* (*e.g., Pendlebury, 2013*)*.—LJY*]

441-442) included a "formal" sorting of 540 publications ("slightly less than half of all relevant" ones for the time period being covered) for their "contribution to knowledge." With high agreement, they categorized approximately ⅔ as "utterly inconsequential" and less than 10% as "worthwhile" contributions to progress in the field.

A sophisticated social scientist might think it quite otherwise in physics (I had assumed so), but the surprising paper by Cole and Cole (1972) on the "Ortega hypothesis" (i.e., that the work of many average scientists contributes substantially to the advancement of science) shows that only a small fraction of publications in that field exerted any appreciable influence on the authors of seminal contributions. "[T]he physicists who produce important discoveries depend almost wholly on the research produced by a relatively small number of scientists" (Cole 1970, p. 377). As to the situation in the humanities, I should perhaps avoid speculation, but I hazard an armchair argument to explain the astonishing 98% non-citation rate in those fields. Since the cognitive *aim* of most scholarship in humanities is less "theoretically ambitious" (and typically more idiographic than nomothetic), the odds of a publication strongly influencing others are substantially lower than in fields like physics, genetics, or psychology. A fascinating study by Biglan (1973) employed multidimensional scaling to identify major dimensions on which academic disciplines differ. An incidental finding tended to corroborate Kuhn's (1962) conjecture that social sciences and humanities lack a paradigm (an accepted general theoretical framework that specifies problems, methods of solution, acceptable answers, etc.). Disciplines not oriented to the construction of strong "general theories" should presumably demonstrate less cognitive constraint from neighboring sectors. A chemist may have to worry about some minor physicist's esoteric finding of an atypical viscosity change with temperature; a clinical psychologist studying pleasure deficit in schizoids should at least ask whether so-and-so's study of reinforcement schedules with the monkey bears on the hedonic metric being used; but it seems unlikely that a scholar studying 17th century Spanish drama needs to know a colleague's work on 16th century Portuguese love poetry, although they may occupy adjacent offices in the Romance Languages Department. For discussion of that cognitive fragmentation in historiography, see the fascinating book by Peter Novick (1988, chap. 16, pp. 573-629).

The "ephemerality" problem is exacerbated in the social sciences because of certain substantive limitations, research strategies, and unhealthy intellectual habits that prevail in those disciplines. Being myself a social scientist who has published in several areas (e.g., psychopathology, psychometrics, learning theory, animal behavior), I permit myself the expression of some heretical views which are not documented by statistical studies, with the expectation that the more competent of my fellow social scientists will share them essentially, although perhaps not in the quantitative degree I here express. I think most high caliber psychologists, whether engaged in the "hard" or "soft" sectors of the discipline, while they will understandably be defensive vis-à-vis critics like former Senator Proxmire with his "Golden Fleece" award for silly research, are willing (with some reluctance) to admit *in camera* that (a) the progress of the social sciences is

discouragingly slow, (b) the validated scientific status of its leading concepts is distinctly poor in comparison with comparably central concepts of the biological and physical sciences, and (c) with certain exceptions there is a striking tendency for both experimental and theoretical contributions in the social sciences to lack that impressive cumulative character we have learned to take for granted in fields like astronomy, chemistry, and genetics. Putting it strongly: I know hardly anyone of top intellectual caliber in psychology who disagrees with the preceding remark. To readers in other disciplines I will merely recommend several papers of mine (Meehl, 1990a,b; 1978 and the papers cited therein[15]), Nobel laureate Feynman's (1986) discussion of cargo-cult science, Lykken (1991) on what's wrong with psychology, and the hard-hitting book by Andreski (1972) which ought to be required reading for all doctoral candidates in the social sciences.

There are intrinsic difficulties of the subject matter (Meehl, 1978, pp. 808-817) which make it possible—I am inclined to say likely—that the study of human personality and of society will remain permanently in the loose and stochastic state that much of medicine or of such physical sciences as geography, historical geology, or meteorology are today. But in addition to these substantive limitations imposed by the nature of our subject matter, contemporary "soft-area" psychology and sociology suffer from the ingrained tradition of almost exclusive reliance on the statistical significance test as a means of corroborating or refuting substantive theories. The effects of this intellectually dyshygienic practice, which persists as almost the sole research method in the "soft areas" (clinical, counseling, personality, and social psychology) despite the decade and a half of searching criticism by representatives of several disciplines including statistics itself (Carver, 1978; J. Cohen, 1994; Harlow & Mulaik, 1997; Morrison & Henkel, 1970; Schmidt, 1992, 1996), are grave indeed.

In the "soft" areas of psychology, there are 10 obfuscating factors at work which make it hard to evaluate a research study that relies upon refuting the null hypothesis by significance testing as a means of corroborating a substantive theory. Discussion *in extenso* of this vexatious matter has appeared (Meehl 1990b). Meanwhile, journals and publications continue to multiply (see, e.g., Broad, 1981; Standing & McKelvie, 1986).

---

[15] From Meehl (1978, pp. 806-807): "I consider it unnecessary to persuade you that most so-called "theories" in the soft areas of psychology (clinical, counseling, social, personality, community, and school psychology) are scientifically unimpressive and technologically worthless. Documenting that statement would of course require a considerable amount of time, but you can quickly get the flavor by having a look at Braun (1966); Fiske (1974); Gergen (1973); Hogan, DeSoto, and Solano (1977); McGuire (1973); Meehl (1959, 1960); Mischel (1977); Schlenker (1974); Smith (1973); and Wiggins (1973). These are merely some high visible and forceful samples; I make no claim to bibliographic completeness on the large theme of "What's wrong with 'soft' psychology." A beautiful hatchet job, which in my opinion should be required reading for all PhD candidates, is by the sociologist Andreski (1972). Perhaps the easiest way to convince yourself is by scanning the literature of soft psychology over the last 30 years and noticing what happens to theories. Most of them suffer the fate that General MacArthur ascribed to old generals—They never die, they just slowly fade away."

Sacred Cow IV: Most Research Worthwhile

Everyone acquainted with academia knows that the accumulation of "yardage" (numerous publications) is the way to get academic tenure and raises, besides being simply "the thing to do." It is done willy-nilly by faculty in "research institutions" whether or not they have much creativeness, analytic power, or deep intellectual passion. The rate of peptic ulcers in the academy would probably decrease, and family life improve, if the publish-or-perish tradition of prestigious private and large state universities were eliminated. The situation is well captured by these quotes collected by Hamilton (1990, p. 1332):

> "It indicates that too much is published. A lot of us think too much is published." [quoting "Timothy Springer, a Harvard cancer researcher"]
>
> "The obvious interpretation is that the publish or perish syndrome is still operating in force," said David Helfand [identified as "chairman of the astronomy department at Columbia University"]
>
> "In many ways, publication no longer represents a way of communicating with your scientific peers, but a way to enhance your status and accumulate points for promotion and grants." [quoting "Allen Bard, editor of the *Journal of the American Chemical Society*"]
>
> "At the state colleges and universities, where they believe publication is their road to credibility, there's still a great emphasis on the number of publications, says Vito Perrone" [identified as "a Harvard School of Education researcher"]
>
> "*If the bottom 80% of the literature 'just vanished*,' [said MIT biology professor Richard Young] "*I doubt the scientific enterprise would suffer*" [italics added by PEM]

As mentioned before (in Sacred Cow I), the personal values, expository skills, and personality traits required to be an integrator and communicator of knowledge are not the same (although they overlap in part) as those required to be a generator of new knowledge. Some of the traits and interests relevant to the two sorts of tasks even have substantial negative correlations (e.g., on the Strong Vocational Interest Blank).

This Sacred Cow is artificially fed by the administrative semantics "scholarship [= publication], teaching, and service," a tendentious use of language implying that to be "scholarly" always involves being a new-knowledge generator. Actually the publishing pressure prevents many of us from being scholars in the old-fashioned sense, as we dare not read widely *even in own our field* because it takes time and energy from our experimenting and writing. This mental set becomes so firmly fixed that most academics feel somewhat guilty about "non-professional" reading—however scholarly the substance—long after they have "made it" to full professor status.

The semantics that equates "scholarship" = "knowledge generation" = "publishing articles" has become so entrenched in verbal habits that I find colleagues are puzzled, dismayed, or offended when I point out the obvious truth that they are not at all the same, and that speaking as if they were is a corruption of the English language. It is mainly

since World War II that this automatic identification of publication yardage with scholarship has taken place. Any thoughtful person realizes, upon reflection, that an academic can be "scholarly" in all *core* respects without being a new-knowledge generator. One can be knowledgeable about the facts and principles of a domain; capable of thinking clearly and incisively (if lucky, even deeply and creatively) about it; objective, fair-minded, and rational in examining the evidence for and against various views; capable of integrating concepts from different domains; skillful in articulating what is known, and the present state of argument and evidence about what is not settled; effective in communicating these things to students; intellectually passionate and alive, so that one's enthusiasm for leading the life of the mind is obvious and, to students capable of receiving it, infectious. Being, having, manifesting all these, one may nevertheless not be an effective knowledge generator in the sense of concocting new theories, inventing new techniques, or collecting new factual data.

Let it be clear what I am asserting here and what I am *not* asserting. I am *not* saying that there is a negative correlation between laboratory skills or theory building proclivities and these other admirable traits of the intellectual life. I'm simply saying that there is not a close empirical correlation between the two, and there surely is no intrinsic logical (semantic) necessity that they go together. One cannot easily ascertain the empirical trait correlations here, the present institutional arrangements being predicated upon the automatic, unthinking identification of "scholarly" with "article generating" or "fact gathering," so that prestigious institutions do not tenure faculty of the sort I have in mind when I speak of the exceptions. I have no stake in arguing that the exceptions are common, or that if the institutional arrangements changed they would become the norm. But it is an undisputed historical fact that at many institutions of learning, over many centuries, some professors *have* been held in high repute for their scholarly attainments who were not generators of new factual knowledge or even of new conceptions, theories, or methods. As mentioned above, certain personal traits such as intelligence, verbal fluency, energy, clarity of mind, and intellectual passion will tend to contribute appreciably both to being a good knowledge generator *and* a good knowledge integrator and transmitter among those who choose, or are required, to do both. That, of course, is quite different from the question whether the two activities facilitate each other in terms of total time and energy available to a particular person (Sacred Cow I!). One would expect low positive correlations between teaching ability (say, as judged by students) and published productiveness, and that is what is found empirically; but the correlations are low (although not, as students sometimes think, negative).

I suppose that my generation of college students was one of the last to attend high-caliber state universities when an appreciable number of senior faculty were "scholarly" in the general old-fashioned sense but did not have much of a track record as new-knowledge generators. I am confident that any reader who was an undergraduate or graduate student in the period immediately preceding World War II (I received my BA in 1941 and my doctorate in 1945) will find it easy to call to mind professors who were

scholarly in a high sense of that term, but who published little or nothing containing new factual material or formal discoveries. I think, for instance, of Richard M. Elliott who was chairman of Minnesota's psychology department from 1919 to 1951 and who never published any empirical research, not even his doctoral dissertation (although he was listed *pro forma* as one of the multiple authors in the famous Minnesota Mechanical Abilities Studies because of his administrative role as chairman of the department). He did write a few book reviews and was world renowned as editor of the Century Psychology series; but he was never an empirical researcher, and he never developed any new theories, or any new statistical methods, or any new methodological principles. Nobody who took Elliott's courses or conversed with him about scholarly matters or knew him socially could have had the faintest doubt that he was a scholar in both a broad and a deep sense. His knowledge of psychology was encyclopedic, he was knowledgeable about art and music and history, a student of Buddhism, but no one would say he was a dilettante or that his lectures in the field of psychology were superficial, ill-informed, or dated. He knew enough astronomy that during World War II, when a large number of Navy recruits were on our campus and had to learn celestial navigation, he volunteered and successfully taught a section of the beginning astronomy course.

To take another example, one of the best courses I ever took in seven years of higher education at the University of Minnesota was in the mathematical theory of statistics, taught by Professor Dunham Jackson. Jackson had the same intellectual and characterological attributes as Mike Elliott (and he had done telepathy research with Elliott when they were undergraduates together at Dartmouth). Sometime in the middle 1920s he had written a couple of respectable papers on the spherical trigonometry of multiple and partial correlation, but his publication track record would not have gotten him tenured by today's ground rules. I can recall three other faculty members in the mathematics department from whom I had courses and who were superb classroom teachers, infinitely patient and ingenious in explaining difficult ideas, and who managed to transmit the intellectual excitement of mathematical concepts, but none of the three had published anything. The chairman of the Zoology Department, Dwight Minnich, had, I believe, published only one paper (based upon his doctoral dissertation) and was the same kind of scholar as was Elliott. I had fine courses and wonderfully educative conversations with philosopher Alburey Castell (the "Augustine Cassell" of B. F. Skinner's novel *Walden Two*) who had published little besides a college logic textbook (1935). "Merely a textbook" is what would be said of his publication in the dean's committee if he were up for tenure today, it being taken for granted that no matter how lucidly one exposits a difficult topic for the undergraduate mind—a task which often demands a more superior scholarly comprehension of the material than does the usual research article—it doesn't count as "scholarship" because it isn't "new knowledge."

I could go on with examples, and I don't think readers my age would have any difficulty in adding their own. Yet it has come to pass over the last 50 years that not one of these professors could now be called "scholarly" for purposes of tenure, because they

were not collecting new facts in the laboratory and publishing them in refereed journals. I have never heard a reasoned justification for this semantic legerdemain. The most you ever hear is a weak speculative argument to the effect that "surely a person who is *really* intellectually alive would *want* to produce new knowledge?" There's no "surely" about it, it's just plain empirically false. I am prepared to argue that Mike Elliott was *more scholarly* than the average academic psychologist who is grinding out, say, two standard, ordinary, pedestrian research articles annually that hardly anybody reads. It requires forgetting what we know about the psychology of learning, the psychology of motivation, and the specificity of human traits to think that nobody who has a good brain and who likes to read, and think, and understand concepts, and explain them to others, could fail to be a highly productive empirical investigator.

Additional References for Sacred Cow IV, not discussed in text:

Begley, S. (1991, Jan 14). Gridlock in the labs: Does the country really need all those scientists? *Newsweek*, p. 44. [Cites Inst. Scientific Info., 45% of papers published between 1981–1985 not cited 5 years post.]

Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science*, *13*, 285-305. [Most scientists have very low productivity. *If* scientist is noticed, will be in first 5 years after publication. Powerful summary of studies on age, area, etc. Social scientists less productive than biological and physical.]

Hartle, T. W., & Galloway, F. J. (1996, Sept-Oct). Too many PhDs? Too many MDs? *Change*, pp. 27-33. [Academic labor market grows tighter.]

MacRoberts, M. H., & MacRoberts, R. R. (1984). The negational reference: or the art of dissembling. *Social Studies of Science*, *14*, 91-94. [Example bias source in content of published articles.]

Mahoney, M. J. (1985). Open exchange and epistemic progress. *American Psychologist*, *40*, 29-39. [*Source for brief history and role of publication in science and tenure*.]

Menard, W. H. (1971). *Science: Growth and change*. Cambridge, MA: Harvard University Press. ["Most scientific papers are ignored from the time the ink is dry." (p. 21)]

Sacred Cow V

No one can assign priorities among fields because we each believe our own field is valuable.

This one is pretty silly on the face of it, because if a course in Irish-American Studies is now offered in my college, some dean or committee must have decided it should be added. By administrative symmetry, that implies some empowered individual or group is capable of deciding that it should *not* be offered or, once in the class bulletin, that it should now be deleted. Institutional social life is impossible if we operate on the strange assumption that no decisions can be made if not agreeable to all of the affected parties. Besides, is it really plausible to suppose that there would be no consensus among various groups as to priorities if economic factors *forced* us to decide which classes, areas, or departments to preserve and which to let die? I am a clinical psychologist, so I of course think that clinical psychology has something good going for it. But if I reluctantly have to vote on whether clinical psychology is as important as, say, pediatrics or chemical engineering, I will decide in favor of the latter two. Consider a list of variegated disciplines, e.g., internal medicine, Irish-American studies, metallurgy, social psychology, calculus, art history, computer programming, golf clinic, and business management (assuming each to be taught at "typical" levels of competence). Ask various groups, such as college students, their parents who pay the fees, members of the state legislature, a random sample of taxpayers, or even a random sample of college faculty or of the Board of Regents, to rank the above named fields in order of their value, letting each individual rater define "value" in whatever way. Is it seriously believed that there would be a marked disagreement on the average rankings, especially at the ends? The question is easily researchable, and it should be done; but I know how I would bet, giving good odds.

When one asks faculty to mention some (presumed) properties of subject-matter areas or courses that lead to our putting a high value on them, the justifications they come up with are few in number and there is remarkable consensus. I have done a little informal polling of colleagues but will briefly summarize qualitatively so as to avoid an appearance of committing pseudo-quantitative "social science research." Here are the usual ones:

- Does the subject contribute to the alleviation of some social problem (e.g., disease, drug addiction, divorce, child abuse, war, racial conflict, anxiety, low morale, inflation, poverty, illiteracy, pollution, crime, mental illness, economic waste, sexism, tyranny, corruption, political instability)?
- Does the subject have a fairly solid cumulative character, in which genuine knowledge develops and evolves, as contrasted with those subjects in which theories

and methods are so faddish and ephemeral that it is hard to say whether after 50 years of research we really *know* any more than we did at the beginning?

- Do we think that exposure to this course should tend to make better community members (as parents, voters, taxpayers, leaders in business or professions, opinion-shapers)?
- Is the subject something that most educated reflective persons would view as a core element in one's cultural tradition, quite apart from whether they think it has a lot of truth to it?
- Does it help one to understand and appreciate other cultures?
- Does acquaintance with the subject contribute to one's subsequent esthetic, spiritual, or intellectual life?
- Do we think that mastery of the subject requires a high order of intellectual competence (ability plus effort), or is it an easy "fake discipline" that has scholarly pretentions but very little substance that makes demands upon the student?
- Does accredited knowledge of this subject matter help the student to earn a living?
- Does it help in understanding oneself, or in reflecting on one's purposes in life and one's relationship to society or the cosmos?

Within constraints imposed by law, regulation, and custom on the power of administrators and constituencies (including students), it is surely allowable to adopt a policy of "paying special attention to better informed, wiser, more experienced persons" in weighing opinions as to the value of departments, subject matter areas within departments, and individual courses. (The same can be said, despite the terrible difficulties in taking account of nuisance variables for indicators of assessment, about evaluating faculty research productivity, discussed in Chapter IX, p. 139.) One might reasonably suppose that deans—who have, after all, been professors and who at one time were students—possess appreciably more competence to evaluate the merits of a subject than would a second quarter freshman. Further, however it is done in detail, there is the inescapable constraint involved when we consider who is paying the freight. At the height of the student power movement in the late 1960s and early 1970s, one frequently heard the argument, "We students, being the consumers, ought not merely to have an input or some fraction of the vote, we ought to have the chief say in what is done." Since students collectively paid for only 25% of the University's operation (at the professional and graduate school level students pay less than 10% of the cost of their education), I am disinclined to accept this as a valid argument.

It is interesting to consider whether the "rationality" of evaluations reached by various groups (undergraduate students, graduate students, students in professional schools, professors, higher echelon academic administrators, members of the Board of

Regents, members of the State Legislature, parents of college students, and non-parent taxpayers generally) could be assessed in any way. One criterion of rationality is coherency, that is, whether the representative individual in a group has thought through his ideas enough to "get his act together" in such a way that the attributes that he, upon reflection, deems important in determining a course's value fit closely with his global evaluation of a course or do not.

Starting with the above list of attributes that contribute to the valuation of a subject matter domain or an individual course, one could obtain judgments from persons who have taken a course as to how much of each attribute the course possesses. Less pertinent but still worth getting would be judgments by persons who have not taken the course but who have had an opportunity to look at the course description together with, say, a syllabus and copy of its final examination. Ideally, such judgment would involve rating or ranking courses only on a single attribute at a time, so that the judge is not influenced by background knowledge of other competing attributes in the list. This rating procedure would yield a set of crude "scores" for each course (e.g., sewage disposal might get a high rank in social utility but a low rank in contribution to general intellectual culture of the student). A second set of judges, not overlapping with the first set, but drawn randomly or representatively from the same population (e.g., graduate students), is not asked to evaluate specific areas or courses, but is asked to judge the relative importance (by rating, ranking, or pairwise comparison) of the relevant properties. These raters are therefore presented with the list, and while they are urged to reflect on various courses they may have taken (or taught), such reflections serve merely as an aid in assigning "importance" weights to the properties, not to rating specified courses as such. The result of this judgment process is a set of relative weights, so that a crude index for evaluating any particular course or subject matter is available. For example, it may turn out that contribution to solving a social problem gets a very high weight, helping the student to get a better job a slightly lower weight, and perhaps familiarity with an important historical contribution to Western culture (e.g., reading Plato's dialogues) gets a small weight, and so on. A third group, who also have either taken the courses or are somehow familiarized with their content, is asked to rate the courses or subject matter areas by whatever criteria they choose to apply. They are not asked to state their subjective criteria. Again, it doesn't matter much whether this is done by rating, ranking, paired comparisons, or by more elegant psychometric methods. Part of this third group could be provided with the *list* of criteria, but not the *weights* we have derived. The other part makes judgments without having such a list.

The point of using three subgroups is that we don't want the individual raters to achieve an artificial coherency via their own strain to consistency. For that reason, one part judges the properties of the courses without having in mind the importance of a property or what other properties it might compete with; the second part evaluates the relative importance ("value") of the properties themselves; and the third part, without either of these quantitative criteria available, evaluates the courses globally. If we sample

sufficiently large numbers to achieve reliability (there is a standard formula in psychometrics for doing this) from the same group, the extent to which we can predict the global ratings of the third part from an equation based upon the results of the first two subgroups is a crude indicator of the extent to which (collectively) they "know what they are doing," the extent to which they are coherent in making their global judgments.

The sacredness of this Cow is seen in the almost reflex defensive reaction to the gentlest, timidest, tentative suggestion that perhaps Area X need not be maintained (or improved) in one's college or department. This defensive response is not wholly attributable to faculty identification with their own areas, however; it is partly a sort of general trade union solidarity against what is perceived as a philistine attack on the values of the academy. "Well, Professor Glotz is retiring next year, and perhaps we should re-examine the question whether it's imperative to offer his classes in interpersonal ceramics," is likely to be met with, "Oh, if we gave that up it would gravely weaken our program—surely you're not suggesting this is not a valid domain of scholarship?" If one is so rash as to mention the widow charwoman's tax dollar as perhaps not fruitfully spent in this way, one is likely to be put in the role of an anti-intellectual holding the values of Sparta against those of Athens. (This amuses me, since my value system is predominantly Athenian, and more so than some of my brethren in clinical psychology—an *applied* discipline—approve of!)

That this reaction is kneejerk defensive seems to me attested by two considerations. First, we all know that there is hardly a single institution which maintains even modest strength in *all* scholarly areas. One accepted basis on which prospective graduate students choose a university (along with stipends, geography, family considerations, etc.) is whether a department is nationally known as emphasizing or de-emphasizing a particular subject matter. This is taken for granted, and I have not heard anybody in my profession maintain that it is imperative that everything be offered everywhere.

Second, the transition from "It may not be imperative to maintain excellence in Area X at Minnesota" to "Oh, you don't value this discipline, you don't think that it's worth anybody's scholarly pursuit" involves major slippage in a sequence of logical steps required to make the argument. Consider the statements needed in a chain of inference to get from the first to the second statement:

1. Subject S is worthy of scholarly pursuit by someone.
2. Subject S ought to be pursued by a sizeable number of scholars.
3. The scholars who pursue Subject S should be college professors.
4. The college professors who pursue Subject S should be employed by some tax supported universities as well as by private ones.
5. Every state college or university should have professors who pursue subject S.
6. Every state college or university should have professors who pursue subject S at a high level of excellence.

It is obvious on reading this list that not a single one of these statements is logically entailed by its predecessor.

I offer an example close to home. My teaching and publishing career has included work in more different areas of psychology than is usual, including animal behavior, learning theory, psychodiagnosis, psychometrics, psychotherapy, behavior genetics, political psychology, personality assessment, and forensic psychology. With the possible exception of the last one, the traditional metaphysical "mind-body problem" of the philosophers has had absolutely zero bearing on my research, teaching, or clinical activity in any of these fields. Nothing that any psychologist does as experimenter, theoretician, or practitioner is affected one iota by where he stands on the metaphysical mind-body problem. I have, however, had a moderately strong interest in it since I was an undergraduate, and have published several papers in philosophical books and journals dealing with it. I am the only one of a psychology faculty of 33 [around the time of this draft] full time members who has done so, or who can be seduced into a serious discussion regarding it. It is one of the "Great Problems" of philosophy, and I suppose now that I am retired, a few—a *very* few!—psychology students will lament the absence of a professor who has an interest in this ancient esoteric topic. But it would be foolish for the psychology department to seek a replacement for Professor Meehl with one of the job requirements being competence and interest in the mind-body problem at the psychology/philosophy interface. It's an interesting subject, and I suppose most educated persons, whether academics or not, would be pleased that a *few* good intellects continue to devote themselves to it. But there are far more persons writing on this subject than ought to be, most of what is written on it (especially by the ordinary language philosophers) being worthless. If the number of faculty writing on it were cut down to a few who hold positions in philosophy departments at private universities, the intellectual life of humankind would not suffer an irreparable loss.

While I personally am glad the Minnesota taxpayer did not object to my spending a small fraction of my scholarly time on this fascinating metaphysical mystery, I do not entertain the egocentric illusion that I was making any appreciable contribution to scientific psychology, or to the welfare of Minnesota's citizens, or even to the illumination of more than a minuscule percentage of doctoral students in psychology because I have that oddball metaphysical interest. So I hold that the Minnesota psychology department does not need somebody to take this up now that I am no longer teaching. If there were *no* psychologists who took the mind-body problem as one of their minor interests employed by *any* state university, no social (or, I wager, "purely intellectual") harm would be done. If the majority of private institutions of learning had no psychologists who did so, there would be no great loss. One can list this topic, along with Etruscan Art or Buddhist Theology or 12th century Portuguese Ballads, as worthy of serious scholarly investigation by a few professors, without concluding that hundreds should be employed for doing it—certainly without concluding that those employed for doing it should be working at state universities on taxpayer money. The readiness with which a challenge of

this sort is met with the accusation of Spartan philistinism testifies to the sacredness of this Cow, as well as to the general trade union defensiveness of many academics.

An especially costly and wasteful personnel policy arises when a variant or corollary of Cow V (priorities are unassignable) is implemented under joint constraints imposed by Cows II (lecture format is optimal), I (teaching and research are complementary), and IV (most research is worthwhile). That everyone naturally believes in the value of his or her own field easily leads (via a mixture of amity, comity, academic etiquette, and frenzied egalitarianism) to each of us valuing—or pretending to value—all fields that are widely "accepted" as at least within the pale of scholarly respectability. For example, while my colleagues in experimental psychology would not countenance a Jungian teaching our students astrology, they might take my word for it that our department "really needs" a class in Freudian dream interpretation. (I do not here argue the merits, and have chosen an example questionable in my own area, lest the reader assume I am snobbishly exercising my prejudices against some others. I used to teach psychoanalytic therapy, including interpretation of dreams.)

Sacred Cow V gets an assist from the academician's status motive being institutionally channeled, in that one wants to be perceived, whatever the reception of one's own published work, as "in a good department." If clinician Meehl insists that Freudian dream theory is a "must" for a good department, the thought that I may be right will make my non-Freudian colleagues institutionally nervous about omitting it. The feeling is one of having to cover the scholarly waterfront, on pain of losing face, of being second-rate, or "first-rate, but *narrow*."

From this point on, the mini-stampede of Cows V-II-I-IV drives us inexorably into costly foolishness. If students must learn Freudian dream theory (V), then there must be a lecture format class in it (II), but then we must hire someone who is a knowledge generator as well (I and IV?). Result: A new high-priced budget item!

How would a rational department handle this kind of problem? It would begin by recognizing explicitly and unabashedly what everyone knows to be true today (even while worshiping the Cows), namely, psychology departments are not all "equal" in the amount and quality of their offerings in subdomains, and in reality they don't even claim to be, or to try hard at it. Whom are we kidding? Ourselves? Students? Taxpayers? Other schools? To be all things to all persons is both impossible and unnecessary, so why make a pretense of working at it? The plain fact is that faculty and students at good schools have a pretty clear notion of "what really goes on" at other places, and if one doesn't know, it takes little effort to find out quickly. One observes this in talking to new graduate students coming from elsewhere. Only the most uninformed or feckless psychology BA would choose Minnesota for graduate study in psychoanalysis, or Adelphi for mathematical taxometrics, or Michigan for behavior genetics. Accredited psychology departments make available a core of course offerings, at undergraduate to graduate levels, without which they would have trouble remaining accredited. But the differences

among them outside that central core are very great and closely correlated with the faculty's publication pattern and the scholarly ethos, intellectual heroes, and conversational flavor of each school.

Making no bones about this, a rational Minnesota faculty would have several inexpensive options. It could say, "Meehl is a clinician with 30% Freudian sympathies, so we understand his view. But we don't have to buy it. Our consensus (not unanimous, but strong majority) is that we do not *need* a dream course, although if dollars were plentiful, it would be nice. But we don't *need* it. Students with a BA or PhD are not warped, stunted, or mis-educated for not having had such a class. Of course an enterprising student can read up on it, with a small library assist from Meehl, since he thinks well of it. Everybody knows the Minnesota emphasis, and Freud is not it."

Alternatively, it could decide to offer such a course by hiring a local psychoanalytic practitioner on a part-time basis. Since pay is pro-rated, such special jobs are invariably done much cheaper this way than by adding a full time line item to the department budget. But you can't see it that way if you are worshiping Cows I and IV, because your part-time instructor will not meanwhile be bringing glory to the department by publishing papers on dream interpretation.

Suppose the class urged on us is less peripheral than that example, say, a general senior college course in abnormal psychology. A department with dominant hard-science ethos and without a clinical training program at the graduate level might (I would say *should*) think that some acquaintance with the breakdowns of psychological function is part of being an educated psychologist. For undergraduates not going on to graduate or professional school (e.g., law, medicine), one may plausibly argue that familiarity with basic descriptive psychopathology is useful to them as future spouses, parents, voters, jury members. Several thousand preventable suicides occur annually because relatives of depressed patients do not know how to recognize the behavioral and vegetative signs of an endogenous unipolar or bipolar depression, for example. Trade unions can be led astray into grievous self-harm by a leader whose paranoid grandiosity would be apparent to even a fledgling clinician. So our hard science experimental department concedes, however reluctantly, that an undergraduate level offering in abnormal psychology is "necessary." But they don't have to squander a professorial slot on it, if they are free of cow-worship. They can choose among several alternative modes of instruction. ETV is a particularly good option here because a lecturer's verbal characterization of how depressed, schizophrenic, psychopathic, hysteroid, or coarse brain syndrome patients sound and look is a poor substitute for hearing and seeing several of each. To the extent that a live lecturer is desired, we again have the less expensive part-timers readily available. The main point is: The two propositions, "Students should be permitted to learn basic abnormal psychology in our department" and "We must add a full-time professor to lecture on abnormal psychology and publish articles about it" are a long, long way from being equivalent.

Sacred Cow VI

Going to college is good for almost everybody, hence the more persons who get a college education the better; and, ideally, almost everybody should go to college.

I have seen survey data showing that about 75% of young married couples hope and expect that their children will go to college. Moving down to the 25th percentile on the normal distribution curve takes us to IQ 90, the mean IQ of those above this cut being 106. Students in the IQ range 90 to 115 can profit little from studying Kantian ethics or Greek history, let alone integral calculus, organic chemistry, or Keynesian economics. It has often been noted that the American mind presents a paradox here, in that it values education *as an idea* and parents desire education for their offspring as a means to upward social mobility, but the dominant American ethos is anti-intellectual, probably more so than in any other developed nation (Hofstadter, 1963; Veysey, 1965). A degree in liberal arts is not such a safe ticket to a well paying job or one carrying prestige, interest, and inherent work satisfactions as formerly, but the evidence is clear that it still correlates with subsequent earning power enough to be worth doing (cf. Herrnstein & Murray, 1994). However, all we can assert with confidence is that "going to college" is a statistical predictor of economic level. As for determining the relative extents to which this long-term economic benefit is attributable to (a) the sheer social status upon job entry of *being* a college graduate; (b) knowledge, skills, interpersonal style, etc., acquired in college; (c) self-selection (as to who enters college and survives) for life-helping traits (e.g., IQ, social potency, persistence, energy)—teasing apart the causal chains underlying cross-sectional correlations is, I believe, beyond our powers, even with sophisticated statistical procedures like path analysis. Only the wealthy can afford to go to college with the mental set of the nineteenth century upper class, that it is a sort of social-cultural-intellectual "finishing school" where you may also meet your future spouse. One is likely to be called an elitist for saying that most people shouldn't go to college, either for their own good or for the good of society. But one need not be an elitist (in the disreputable sense) to say that this Sacred Cow remains unproved and has a good deal of both common sense and current experience speaking against it. An industrial psychologist colleague informs me of a growing tendency in manufacturing companies to hire bright high school graduates to fill certain kinds of technical jobs, because it usually turns out that the BS or BA degree does not prepare applicants to function adequately without considerable in-service training. Since the company has to educate the new employee, why not select non-college applicants who expect less and come cheaper?

Although the correlation between IQ and family of origin is modest ($r \approx .35$), correlations of this size should not be dismissed when social policy involving millions of persons is our topic (cf. Lubinski & Humphreys, 1996). The correlation is relevant to

Sacred Cow VI: College for Everybody

Sacred Cow VI because the high cost of college attendance is more burdensome for lower socioeconomic class students, and is the main cause of the marked lengthening of time to get the BA degree in recent years. "Getting an education" is the magic phrase for upward mobility in lower and lower-middle class families where parents and offspring aspire to rise in the social system. For many such students the vocational competencies acquired in college would be far more quickly and cheaply acquired without the numerous BA requirements, and most of them are not even faintly interested in the life of the mind. President Clark Kerr (Berkeley) said, "The university has three constituencies: The students, who are interested in sex; the alumni, who are interested in football; and the faculty, who are interested in parking," a cynical witticism which is, alas, close to the truth. In fairness to a distinguished administrator, I should not quote that insightful witticism without mentioning President Kerr's (1964) fine book, *The Uses of the University*, based on his Godkin Lectures at Harvard, for a presentation of his serious views.

Unlike the other cows, this one I do not know how to cost account with empirically defensible figures, so I confine myself to some plausible qualitative comments, some of which bear, at least indirectly, on money. Academic colleagues (especially ex-deans and department chairs) have objected to my challenge of Sacred Cows II (formal lecturing) and III (small classes) on the ground that the competence of many high school graduates has become so poor in recent years that the motivational factor of lectures is necessary ("they can't—or won't—study on their own") and classes must be small enough to permit much individual attention, such as teachers are expected to provide to small children in elementary school. They also point out that special "remedial" classes have become necessary because even freshman courses in English composition, mathematics, foreign languages, and the various sciences presuppose knowledge and learning skills that in past times were acquired in high school but are no longer. Aside from costs to the institution due to these factors, the economic and social costs to large numbers of students who should not be in college (because they lack the requisite academic ability, motivation, interest, or work attitudes and habits) represent needless psychological stress and social waste. Of course challenging this cow is not only an institutional desideratum, as it requires re-thinking by employers, parents, and secondary school administrators. Substantial reform of universities in this respect would be hampered by pervasive, rigid irrationalities in the larger society for which I am not foolish enough to prescribe cures. But one may entertain hope that purely economic forces, almost always potent in eroding ideological impediments, may be somewhat therapeutic in the long run.

The proliferation of "froth" courses—even froth departments—is too well known to require proof.[16] The social causes underlying this widespread academic degeneration are complex. To analyze them properly is beyond my competence and, I assume, most readers' interests. But one factor is the influx of students whose talent or motivation is

[16] [*It was planned to provide a couple of book cites and a few examples, e.g., a course on wine tasting, but details were not noted on this draft.—LJY*]

inadequate for studying intellectually demanding scholarly materials. In classes where student participation plays a significant role, the presence of much less capable persons tends to degrade the process. (Even at the graduate school level, one is aware that most students prefer to "hear the prof" rather than their typical peers.) One might assume that having more undergraduates taking froth courses does not cost more money, since they are merely listening to a froth professor rather than a scholarly one. But it does not work out that way, because hiring three froth professors does not permit getting rid of a professor of physics, Latin, history, biology, or logic. The non-froth departments continue to exist, and some students continue to take courses in them. Even if a calculus class enrollment drops from 50 to 35, we do not reduce the math faculty size.

A delicate question (not raised in the chapter on Sacred Cow V [prioritize fields]) is that the average IQ of academics may have undergone a decline, such that classes of low scholarly quality have to be available for such faculty to teach. There are many college professors who simply could not master intellectually demanding concepts, persons who belong to the "chattering class" and enjoy the relative ease and security of academic life. I have no quantitative evidence directly bearing on this matter, but research of census data[17] could provide suggestive statistics. The proportion of American employed adults who are college faculty has increased over the years. Unless the attraction of academia (compared with work in business, industry, government, professional practice, etc.) for persons in the high IQ range has greatly increased—and academic life is clearly *not* as much relaxed, carefree fun as it used to be—this rise in numbers must involve dipping down lower in the IQ distribution. In that region of the curve, where it is relatively flat, a 4- or 5-fold area increment requires a non-trivial abscissa move, quite possibly a standard deviation or more.

Another anecdotal hint: I know, with their permission, the Miller Analogies scores of my peer group of clinical psychology PhD candidates in the period 1941–1945. Twenty years later, with 5 to 6 times as many trainees in the program, the average score had declined by a full standard deviation, estimated by Miller's scale equivalents as around 20 IQ points. The brightest clinical psychology trainees we have today are as bright as my peer group, but the average of the group as a whole has declined considerably. One exemplary statistic is that the *lowest* score among my 1941–1945 peer group was 10 IQ-equivalents above the later median. Since these PhD candidates make up the faculty of the next generation, I am willing to count this anecdotal evidence.

Increased student enrollment

It is appropriate here to say something about the role of sheer student enrollment increase in raising costs, a factor much less important than the expansion and diversification of course offerings in the qualitative sense. Persons outside the academy—

---

[17] [*A census data search was started but not completed; changes in census categories over time complicate the task; but online access to census data would make it easier now.—LJY*]

including, I have been surprised to learn, some legislators—are unaware of the great difference in the financial impact between two kinds of enrollment increase. Even college professors deceive themselves on this score, I'm inclined to think for defensive reasons. In most subject matter areas (certainly in psychology), a mere 10% increase in number of students has a negligible effect on instructional cost and garners student fees greatly in excess of the cost increment; whereas if that 10% rise in student head count comes about because of expanded offerings by the creation of a new specialty area, the opposite is the case, i.e., the university usually loses money.

For example [*illustrating here with graduate students—LJY*], suppose a department has 100 PhD candidates in graduate school and—either deliberately or by allowed fluctuations—this number rises to 110, a 10% increase in enrollment. At the University of Minnesota, if seven of these 10 are non-resident students (who pay a higher tuition) the increased enrollment brings in approximately $100,950 in additional fees. What is the increment in cost of instruction necessitated by this 10% increase in PhD candidates? We don't hire any new professors, increase any cross charge of professors participating in our teaching program, hire another half-time teaching assistant or another fraction of a clerk typist or secretary. If I give a lecture course in psychopathology that had 50 students in it the previous year, it will now have 55 if I get a random share of the 10% increment. Nothing that I do in preparing for class, how I conduct myself in class, or anything else changes discernibly. (I neglect the possible influence of an eighth sacred cow, to wit, the dogma that unreliably scored essay examinations are more valid than reliably scored carefully constructed multiple choice examinations, to be discussed in Chapter IX). So far as I can figure out, the only increase in cost is around $120 for additional copying of course syllabi and examinations.

Contrast that situation with one in which the 10% increased enrollment at the doctoral level in a department is attributable to attracting 10 graduate students who would not have come to the University of Minnesota had it not been for the department's decision to embark on a new area of PhD specialization, say, a doctoral degree in interpersonal ceramics. In order to "develop strength" in that area and attract high-quality students—once the department has decided to commit itself to such an enterprise—it is thought that we need a critical mass of at least two faculty. Starting from zero in the domain, we want to hire a major professor who already has eminence. We manage to attract a "star" in the field of interpersonal ceramics from a high paying Ivy League university at a starting salary of, say, $100,000. In order to get the critical mass (so the new professor will have somebody to talk with and to engage in cooperative research) we hire a fresh-baked PhD, perhaps one of the new professor's own recent doctoral products, at, say, $30,000 at the assistant professor level. In order to handle the new courses offered we require a half time teaching assistant at around $7,700,[18] and we need a starting-level secretary for the new

[18] Information 3/3/97: TA gets $10.77/hr × 20 hr for half-time = 215.40/wk × 4 wk = 861.60/mo × 9 mo = 7,754.40/acad yr. Round down because probably doesn't work holiday breaks, etc.

program, who costs us around $16,800.[19] So all told, we need about $155,000 per year to get the new program going. Putting this increment in costs together with the new fees, the $100,950 additional student fees means a net change of around −$54,000 [rounding]. This $54,000 increased cost is attributable not to the increased head count, which taken simply as an enrollment increase would have netted us $100,950 in additional fees; the reason for the increased cost is the expansion of offerings into a new area.

Anyone who has been around academia for long knows that in the great majority of instances such expansion of offerings is not primarily based upon student demand, or the necessities of preparation for subsequent profession as a psychologist, but by the prestige needs of the department faculty. They fear (irrationally) to be nationally looked down on as "slipping," "old fashioned," or "out of touch" because they have not added staff—preferably highly visible staff—in this new growing area of scholarship. Yet there are first-class high-prestige psychology departments that do not even offer a PhD in the "big field" of clinical psychology (e.g., Stanford, Harvard, Brown). Stanford dropped counseling psychology, despite the counselor's favorite instrument, the Strong Vocational Interest Blank, having been developed in its department. Does Stanford lose face thereby, become perceived as going into scholarly decline? Not in the least. Harvard dropped "accredited" clinical psychology in favor of its special high-quality brand of experimental psychopathology. Do we despise Harvard? Hardly. If some academics are so rash as to conclude anything from such non-offerings, it's that counseling and clinical psychology are in decline!

The Tyranny of Words[20]

We academics are inclined to flatter ourselves that, on the average, we "think better"—that is, reason more logically, with less prejudice and emotional distortion, relying on valid arguments and properly verified factual evidence—than nonacademics do. I am not aware of any good quantitative evidence either for or against this pleasing self-image, but speaking anecdotally (itself often a weakness in reasoning, which I hesitate to commit), I incline to doubt it.

One piece of evidence that bears on this notion, although indirectly, is that there is a correlation between IQ and various more direct measures of problem solving ability, and there is also a correlation between IQ and one psychometric measure of fair-

---

[19] Beginning secretary gets $8.76/hr × 40 = 350.40/wk × 4 = 1,401.60/mo × 12 = 16,819.20/yr. Use 12 mos for secretary because they work summers too. NB: does not consider cost to U in insurance, retirement, etc.

[20] [*At the time of this draft, it was still undecided where or whether to place this section in the text or let it stand alone as an appendix. It would have needed nuanced editing with either placement. It doesn't fit closely with the focus of this chapter, but I have left it as Meehl wrote it and hope this note alleviates any dissonance a reader may feel.—LJY*]

mindedness.[21] This latter correlation is not as high as one might have expected, but since it was computed on a college population, it is attenuated by a restriction of range. If corrected it might be as high as $r = .50$ or .60, that is, somewhat higher than the correlation between college grades (in the old days when they meant something) and measures of academic ability. But this relationship may then merely be a group trend associated with individual differences in intellect rather than with whether one was in or outside of the academy. There are a lot of bright people in this world who prefer not to be college professors, and that's not only because they like to make money. If one were to collect a sample of nonacademic lawyers, physicians, engineers, army officers, accountants, and business executives, matched for IQ and (having had some college education) many of whom will also have advanced degrees (MD, MBA, PhD, etc.), it would be interesting to study how their tendency to think rationally compared with a group of academics. Absent any such comparative study known to me, I can only record my personalistic probability that there would be little if any difference. If there were at least a statistically significant difference, however small, and I were forced to bet money on it, I would bet in favor of the nonacademics thinking more rationally.

Despite reading books and leading "the life of the mind," academics as a group (lumping together all fields) suffer from three grave cultural disadvantages militating against being clear in the head about hard facts. One disadvantage is that even if your thinking is muddled, nothing bad is likely to happen to you. A congressman once asked Secretary of State, Henry Kissinger, who came from the academy (where I believe he was not especially distinguished), why there was so much petty animosity, back-biting, envy, and tedious argument over trivia among college professors. Kissinger's tough-minded response was, "Well, it's partly because it doesn't make any practical difference anyway." I believe there is a profound truth in this simple comment. A general who consistently thinks irrationally loses battles and will be relieved of command. A business-man who thinks irrationally about the competition or the consumer will go broke. A politician may think irrationally in floor debate or pass stupid legislation, but if he thinks irrationally during an election year, he is likely to lose. A surgeon who gets a reputation for irrational decisions will be protected by the physicians' old-boy solidarity, but many colleagues will be reluctant to refer patients to him. Whereas in the academy, if I write a stupid, tendentious paper, what is likely to happen is that other academics will be moved to write replies, the result being that I have boosted my number of citations in the scholarly literature.

Secondly, speaking in terms of the psychometricians' well established hierarchical structure of human capacities, we are mostly 'word people' and we make our living talking. A rather small minority are number people and spatial people. A good vocab-

---

[21] [*This sounds like Meehl had a particular cite in mind. I am grateful to David Lubinski for suggesting it may have come from Cattell* (*1950, e.g., pp. 98-99; that location is cited by Jensen, 1998, p. 572*)*. That's the best we can offer.—LJY*]

ulary, verbal fluency, and lots of practice holding forth develops a facility in avoiding—and even for concealing—the truth with plausible-sounding words.[22]

Thirdly, most academics associate both on and off campus with other academics who tend to share their views on such fundamental questions as ethics, politics, religion, science, etc., and as a result, they are rarely put in the position of having to defend the academic subculture's received opinions, because those opinions are not challenged. Thus, for example, almost any American psychologist is absolutely certain that there can be no such thing as telepathy, because everybody around him has that certainty. It is a sacred cow (I am not here raising the question of its truth) that the theory of evolution is correct, and it is taken for granted that nobody could possibly entertain any scientific or philosophical doubts about it except a barefoot fundamentalist from the backwoods with a low IQ. Almost all social scientists take it as axiomatic that rent control and the minimum wage are good things—all, that is, except the economists! I conjecture that these three cognitive disadvantages conduce to irrational thought and may be more than sufficient to counteract whatever advantage reading books and engaging in intellectual conversation may possess over the "real world" of nonacademic life.

Irrational thinking by college professors, of course, covers the whole range of formal and material fallacies, including some big ones that unfortunately do not have recognized standard names. I think the most common single fallacy both in scientific writing and in faculty meetings is treating a piece of evidence or an argument that may have logical weight as if it were dispositive; meaning that one ignores or downplays equally good arguments and facts pointing the other way. Logic textbooks label this widespread error the Fallacy of Neglected Aspect; it is a failure to obey the logician's Total Evidence Rule.

Another kind of error that lacks a standard name is one to which academics are, I believe, especially prone because of our being word people, and that is to be overly influenced by affectively charged or ideological words, and to try to solve problems and make decisions, and posit goals and subgoals, by word-magic. I don't say that college professors are more given to this ubiquitous error than preachers, politicians, or journalists; but I am pretty sure that we are more subject to it than accountants, engineers, physicians, and practicing attorneys.

Count Alfred Korzybski, author of the famous *Science and Sanity* (1933) and founder of the movement known as "General Semantics," has had a bad press. The distinguished psychologist, Charlotte Buehler once dismissed him, saying, "Vell, ve in Wienna knew vhat to tink of a *Polish count*." Doubtless, Count Korzybski was grandiose, perhaps a little crazy, and he did himself no good by saying some stupid and careless things about logicians and Aristotle in particular. Having no academic connections and no PhD, he

---

[22] [*Meehl made a note to himself:* "Here quote, I think, Machiavelli: 'Occasionally words must serve to veil the facts. But this must happen in such a way that no one become aware of it; or, if it should be noticed, excuses must be at hand, to be produced immediately' ('Instructions to Raffaello Girolami' in *The Prince*, quoted by Koestler in *Darkness at Noon*)."]

was readily ignored by academic philosophers. I believe that, due to these unfavorable auspices, the importance of his core message has been underestimated. I would not urge readers unfamiliar with his work to burden themselves with trying to fight through that massive 1933 book or to worry about whether his plates-and-string apparatus, "the semantic differential," was an effective treatment for schizophrenia. Instead, I would suggest a reading of smaller works by his followers: Hayakawa, *Language in Action*; Stuart Chase, *The Tyranny of Words*; and Wendell Johnson, *People in Quandaries*. Human beings are smart enough to use language, but not smart enough to avoid misusing it. People tend to react to words instead of real things. A large part of human misery is wrought not by earthquakes and droughts, but by people upon other people, much of which can be explained by the tyranny of words. To avoid being seduced by affectively charged phrases that are weak or devoid of objective semantic content requires a sophisticated hyperawareness of this danger, something possessed by only a tiny minority of persons. I do not, of course, suggest that Korzybski was the first thinker to notice this, and if the reader happens to know about and dislike General Semanticists we could as well go back to the Platonic dialogues to see an awareness of this danger (which did not, unfortunately, prevent Socrates from often falling into it himself).

I do not bring this topic up just for the fun of it, but because some of the irrationalities of higher education stem at least in part, and some almost wholly, from academics—especially academic administrators and boards of regents—being hypnotized by word-magic. Examples are all around us for anyone who takes the trouble to notice them. I will only exemplify with two magic words that, taken together, tend to generate a good deal of costly foolishness. They are currently in frequent use, and they are more or less sacred, being rarely critically examined as to what, if anything, they mean—or why whatever they may mean is important. The words are "community" and "diversity."

Lest I be accused of sinning against Korzybski and Co., using my own unexamined buzzwords in talking about "buzzwords," let me explain that I am not. The main evil of a buzzword is not that it carries some motivational-affective wallop. Almost all words, and certainly all words from ordinary language, have this characteristic, and there is no reason why we should attempt (fruitlessly) to get rid of it. Humans are not pure computers and we would not want to be. The good old term "rhetoric" has been abused by the media, but it once was a respectable idea. What's wrong with buzzwords, as I use that derogatory term, is that they have only a very confused semantic content, or they denote something that is conceptually clear but has no real existence, so that their main communicative function is the emotional loading. This combination of strong emotional loading with weak objective semantic content is what does us in, leading to irrational behavior and inefficient problem solving efforts. I do not believe I have employed terms in this book that fail to denote something real, and the term 'buzzword' certainly denotes something real. But I have not hesitated to use a word when, in addition to its cognitive content, it may impel the reader by its attitudinal overtones to see something new or think about something differently.

Sacred Cow VI: College for Everybody

Taking "diversity" first, I have trouble figuring out just what is meant, or why it is a good thing. The kind of diversity people seem to have in mind is diversity of skin color, or what one's first language was, or what foods you prefer, or where on the earth's surface your grandparents were born. It also seems to be something that should be "celebrated" (another buzzword). I recently saw a huge poster on a college bulletin board that said, "CELEBRATE DIVERSITY." I understand what it means to celebrate the Fourth of July, to celebrate Mass, or even one's birthday. But given that people have different colored skins or eat different favorite foods, that this is an occasion for celebration puzzles me. Nor do those who talk about it make clear whether people should make an active effort to think and act like those who differ from them in ethnic, class, or other origin; or whether they should cultivate their own subgroups' distinctive features and take pride in them; or perhaps both? Who knows. I am at a loss to understand that last possibility as well. Does it mean that because I am of Norse extraction I should cultivate a liking for lutefisk, which I despise? Or that, if you are an Irishman, or a Jew, you have some sort of obligation to learn to like lutefisk? Or should you rejoice that I do —or don't—like it? Should I take pride in my "roots"? Should I try to identify with them? Should I rape my neighbor's wife, murder him, and steal his silverware, thereby emulating my Viking ancestors? I must say that none of this makes any sense to me, and I view it as largely faddish foolishness, except for the obvious desirability of people not looking upon those who are different from themselves with hatred, fear, or contempt. Surely that is desirable if we are to live together in peace. But I have the impression that those who talk about "celebrating diversity" want something more than this important *negative* advice.

In any case, harping on people having different skin color or food utensils and the like is focusing on the *wrong kind* of diversity, and we would be better off if we quit talking about it. There is, however, an important kind of diversity, which forces itself upon the attention of any observant, rational person and which cuts across such biological and social variables as race, sex, geography, language, or class. It is this other kind of diversity, which we don't usually celebrate and which some muddleheads prefer to sweep under the rug, that really matters and has an unavoidable impact upon what happens in institutions of higher education. This real, important, non-buzzword diversity is the plain ineluctable fact of individual differences in interests, values, passions, aspirations, hopes and fears, and—I almost fear to say it—talents and weaknesses. It is this ubiquitous and utterly unchangeable, powerful kind of diversity that is relevant to how the university should be run. Harping on the diversity of racial origins is focusing on the wrong thing.

Unfortunately, the unimportant, irrelevant kind of diversity sometimes has appreciable correlations with the important kind, which leads unclear thinkers to think it wicked to recognize the important kind. Examples abound. Females have superior verbal fluency, males better spatial cognition. Orientals have higher IQs than Caucasians, and Oriental neonates are more "placid." Little boys are more physically aggressive than little girls. Creative scientists are less sociable than people in general. Political conservatives attach

a higher value to family life and pride in family than liberals. There's hardly any variable of the socially defined and overemphasized diversity that lacks correlates of the individual differences kind. Those muddleheads who are also fanatics hope to devise some social machinery to abolish all such correlations. I am not clear why they think it important to do this.

Over the last 25 years, much of my research has been concerned with devising new statistical procedures for the classification and genetic analysis of mental disorder. Because of this, I have occasion to spend some time in the mathematics department library several times a year. Having noticed early on these visits that there seem to be quite a few Oriental faces reading in that library, I decided, being a psychologist, to practice tallying them. Over at least 50 visits, I found that there was invariably at least one yellow face in the math library, and the proportion varied from 10% to 70% of the readers. Should I be fretful about this? Should I think, "Oh, my goodness, there must be something terribly wrong here. Less than 2% of the population of Minnesota are Chinese or Japanese, so why is there this big ethnic difference?" In those 50 visits, I only once saw a black face, and that person was reading the student newspaper rather than a book or journal, so maybe he came into the math library for the air conditioning and comfy chairs. Is this cause for horror? From 20–25% of the graduate students in physics and mathematics are Orientals. So what? If it is a social fact that people with yellow skin have somehow concluded that mathematics is their thing, that they like it and they're good at it (more often than is true of us Caucasians), whereas Black students have concluded that it's not their thing, that they don't care for it and are not good at it, is this cause for alarm? I do not see why.

I read somewhere that every single tugboat in the New York Harbor is owned and captained by a Norwegian. Should I conclude that the political appointee directors of the New York Port Authority (regularly Irish, Italian, or Jewish) have some sort of bias in favor of Norwegians? But enough. I hope no reader takes me to be saying there are no such things in this world as racial and religious prejudices, and that there has not been such a thing as racial discrimination. I am not an idiot or uninformed. What I am complaining of is the reflex assumption by a certain kind of frenzied egalitarian that whenever there is a statistical difference—ethnic or national or class or religious groups, or that fewer females with high IQ want to be engineers—this is something dreadful that must be altered, if necessary, by force.

Sidestepping frenzied egalitarianism and the advocates of the Ant Society, let us look at the bearing of important kinds of diversity; the phenomenon of individual differences in personality, taste, aspirations, ability, and the like that are relevant to how you run a college. Consider some familiar examples. I shall permit myself a bit of stereotyping, but very little, with which I am sure anyone at all familiar with the scene at a big state university will agree. I describe several idealized types to make the point.

— Al Bumstead is the son of a successful small businessman. He regularly works summers in his father's business, and will play an increasing role in managing it, so father can retire early and play golf after Al gets through college. Al is bright, but not superbright, is unbookish, has little interest in the life of the mind, but will have no trouble getting respectable grades toward his BA in business administration. That's the kind of thing he needs to know to do an adequate job running the shop.

— Chuck Devlin was born to considerable wealth. His parents wanted him to attend Cornell because father and grandfather did so, but he had too many intimate friends who were attending the University of Minnesota; so Chuck's parents reluctantly agreed for him to go to the University of Minnesota, which, while not an Ivy League school, is rated fairly highly nationally. His intellectual interests are absolutely zero, which nobody complains about; and he plans to spend the four years having joined his father's and grandfather's fraternity Rho Dammit Rho, chasing girls, drinking beer with the boys, and hollering at the football games. Although totally unintellectual, he is not stupid and has no trouble getting a gentleman's C.

— Eddie Fisbee is the son of upwardly mobile, lower-middle class, blue collar parents whose IQs are much better than their educational level. His grandfather had to quit college during the Great Depression, and his father passively assumed he would be a blue-collar worker, grateful for economic security and his trade union. But today, the main thing for Eddie, as for his parents and their social group, is that he should rise in the socioeconomic hierarchy, that he should "be a success," which means not bring his lunch in a pail, he should wear a white collar, join the Episcopal church rather than the Baptists, and make a lot of money. It is not statistically astonishing to find that his IQ is considerably higher than either Al's or Chuck's, since the correlation between intelligence and social class of origin is negligible, although that between intelligence and achieved social class is moderate. Persons of proletarian origin who attend college tend to be self-selected for academic ability. At the moment, Eddie and his folks are thinking in terms of a Bachelor's degree in business, but in the back of their minds is the thought that if he does well and can get a graduate fellowship of some kind, he's bright enough to go on and get an MBA, maybe even taking it at Harvard rather than at Minnesota. He does not intend to conduct himself like a total grind and have no social fun, but first and foremost is to hit the books and get the kind of grades that he needs to get where he wants to get.

— Georgianna Hillstrom is the daughter of a physician and a nurse, and since she inherits the bright genes from both sides, and enjoyed science classes in chemistry and biology, she intends to be a physician, which, with her brains and work habits, she will have no trouble achieving. She is not particularly good-looking, and too blunt and matter-of-fact to be a big social charmer, so that if she plans to marry (not a big consideration at the moment because she is not terribly interested in boys), it is likely that it will be another medical student.

— Her sister, Irene, is of somewhat lower intelligence (although not dull), but is highly extroverted, could without exaggeration be described as a beauty, and she will frankly tell you, if she likes you, that she is going to college to "get a little polish," and mainly to find a mate. Finding a suitable mate is one of the several accessory reasons for being in college for both sexes, but particularly females in our society; and since the coefficient of assortative mating for IQ is higher than it is for any other trait, physical or mental, or even social class, this is not a function of college that should be ignored or denigrated. Irene didn't like science courses and found dissecting the frog "icky," so neither she nor her parents considered a medical career for her.

— Now, finally we come to young Philip Martin, an intellectual like the young Paul Meehl of 1938 who decided at age 15 to be a college professor, before he had decided what to profess. This choice was based partly upon the appeal of an autonomous, easy-going life with economic security—he was a teenager during the Great Depression—but mainly because he wanted to lead the life of the mind, and academia was one of the few places you could make your living at it.

A big state university of high quality has a niche for all six of these diverse people. There is no reason why any of them should find the milieu wanting in respect to the satisfaction of their dominant needs. But now we come to the buzzword "community." Years ago, I read a report of a conference of college and university presidents in which they were discussing their problems as administrators. Like most conferences, it was a waste of time; but at least (if memory serves) it was not paid for by the taxpayer, but by a private charitable agency (perhaps Carnegie?). The presidents were asked what were some of the difficult problems they faced and which they found the most troublesome. The biggest concern was difficulties they experienced in fostering a sense of "community" among the students. I take it that to most persons, "community" is self-evidently a good thing, so that if an administrator finds himself unable to foster its development, he is somehow failing at the job. I don't share this attitude. I think this is buzzword baloney. Some may say this is because of some spiritual defect on my part, but I think it is because my Korzybskian way of thinking enables me to see that it's a buzzword that does not stand for anything real. The reason that these fretful college presidents cannot foster an "academic community" is the hard fact of wide individual diversity, which they talk about, but do not understand. There is no way to arrange the institutional forms of a big state university, with 50,000 students and 3,000 faculty who are truly diverse in the relevant sense of that word, that will yield a "community" in the sense that these people seem to desire it. Fortunately, this unattainable goal is also unnecessary. Absence of community (in this nondenoting, spooky sense) in this kind of institution *has nothing wrong with it*, and is not something a rational person would be fretting about.

Consider an analogy: Suppose I am the owner and manager of a cafeteria, or perhaps a more genteel establishment, a smorgasbord. I set out a wide variety of goodies with convenient logistics for people in different lines. The place is tastefully furnished and the chairs are comfy, there's no smoking allowed, nice paintings are on the walls, and

civilized music is piped in. Some vegetarians have been told positive things about my establishment, so I cater to them by having a variety of very high-quality nonmeat dishes. But customer Henry Glotz loads himself up on high-cholesterol, high-fat meatballs and mashed potatoes with high-sodium gravy. Does it bother him that his daughter eats what he calls "rabbit food"? Not in the least, nor should it. Should I, as the entrepreneurial dispenser of these varied dishes, be fretful about a "lack of culinary community"? Of course not. It's silly even to think about it. Those college presidents were just producing hot air.

The most profound and sincere kind of egalitarianism and the most coherent acceptance of diversity is to let people be what they want to be, do what they wish to do, and associate with those they want to associate with. If you are coercive about these matters, like sort of a supernatural social worker making sure everybody is like *you* think they *ought* to be, your love of equality and diversity is a fake. I am aware that in saying this, I betray my own individualism and libertarian sympathies, but what Meehl values is not the point. We are talking about how much it costs to go to college, and what can be done to lessen this cost and satisfy more of the parties involved (students and faculty and taxpayers). Being seduced by these buzzwords gets in the way of rational, efficient problem solving. If someone said, "Meehl, you've just forgotten what it was like to be a poor little student," my answer is, "No, I have not. I remember it very well indeed." It didn't occur to me to ask whether the social frat guys or the jocks or those hell-bent for becoming a millionaire were seeking the same things on the campus as I and my like-minded friends were seeking. I took it for granted that there were all sorts of people different from me. *And the University of Minnesota, way back in 1938, provided goodies for all of us*.

Sacred Cow VII:

Providing persons with a college education is socially valuable, a collective good; for that reason, the taxpayer should be the main source of support.

This question is only partly empirical, although there are statistical data interpretable as showing that the lower and lower-middle classes are regressively taxed to subsidize education of the middle and upper classes at state universities. How much tax support seems appropriate for state institutions of higher learning depends on one's overall political and economic philosophy. Being myself somewhat libertarian, I operate on the general principle that, if it can possibly be arranged, those who *get* something are the ones who ought to *pay* for it. Under the present system, a doctoral candidate in clinical psychology is taught by a professor who is paid by the taxpayer, including the widow charwoman whose child has neither the IQ or desire to go to college. This doctoral candidate is also subsidized (trainee stipend) by federal tax money to take graduate work. Whether he ultimately goes into private practice or works in a public clinic, the widow charwoman pays him for professional services should she require them, via fee for service, HMO, insurance, or taxes. If she is lucky (odds > 4:1) in never needing the services of a psychologist, she still pays for somebody else's use. This is arguably an inequitable arrangement.

Despite differences in political philosophy (beyond the scope of this book), it seems pretty clear that this much can be said: The person who benefits *soonest, most certainly, in largest amount, over the longest time period, and in the greatest variety of ways* from getting a college education is the one who gets it. That being so, distributive justice would suggest that this person is the one who should mainly pay for it. Since we all benefit indirectly from living in an "educated society" with professionals such as psychologists handy (even if we don't all use them), a plausible way of slicing the cost pie would be, for instance, for the taxpayer to pay for the bricks and mortar but the students should pay the freight of instruction. At the present time the largest share of instructional support comes from those students who get the least payoff (e.g., in liberal arts), whereas the students who get the greatest payoff (e.g., physicians, lawyers, engineers, psychologists) pay less than 5% of the cost of their education, with the taxpayer subsidizing the rest.

If the taxpayers' support were considerably reduced, on the theory that those who benefit most from higher education should be the ones who pay for it, there is an obvious objection that students are already having difficulty paying the fees. This is not a silly objection and this book begins with it as a major social problem. But I have two replies, which, while not eliminating the problem, should soften the blow considerably.

Sacred Cow VII: Taxpayers Should Pay

One purpose of challenging the sacred cows is to reduce instructional costs, and we shall see later (Chapter VIII) that they are reducible to about one-fourth of what they are today. For illustration later on, I shall assign 25% portions of the 75% savings equally for the benefit of taxpayers, students, and raising professors' salaries. One obvious amendment to my proposals would be that we shouldn't double professors' salaries, but only increase them by some rewarding fraction, and hold tax support at the present level, so that nearly all of the savings could go to reducing student fees. This is a judgmental matter that depends on considerations of ethics, politics, and economics; those allocated proportions are obviously adjustable, depending upon empirical experience with my proposals. For instance, we might assign, say, 10% of the saved dollars to increasing faculty salaries, and another 10% to reduced taxpayer contribution, which would mean 55% of original costs is assignable to reducing student fees.

However, I believe a more fundamental reply is important, although I shall not develop the argument at length. In my view, it is rational to look upon one's education as a major investment, which one expects to "pay off" in various ways and over much of one's lifetime. I do not think it unreasonable to analogize getting a college education—undergraduate or graduate, depending upon one's vocational aims and the job market—to taking a mortgage on a home. When you buy a house, you anticipate paying off the debt for many years, and you do not, if I recall my own experience as a homeowner, do this grudgingly as if it were some kind of terrible hardship or unjust treatment by the world. But even adopting a rational attitude, there is a major problem presently in the fact that some persons (including lawyers, physicians, engineers, business executives) who become very well paid (or even, by usual standards, wealthy) are so lacking in ethics as not to repay student loans. I am not technically competent to address the question about what to do about this. But it is inconceivable to me that in our society—where everybody has a social security number on income tax forms *and on filed documents for license to practice*—we can't identify such deadbeats and, by a combination of measures, force them to pay. If that is not managed at present, it means either gross inefficiency of the bureaucracy in identifying persons, or a lack of will to coerce them, or both.

I realize that the mental set that I am here describing toward educational investment as analogous to one's investment in buying a nice home, or buying life or liability insurance, or social security, or catastrophic medical protection, involves changing some attitudes, which is never easy. All I'm saying is that the analogy to these other kinds of investment in one's future is a fair one and does not involve imposing an unconscionable burden on the recipients of higher education, who, after all, do it as a matter of choice. I read recently that in Minnesota physicians average an annual income of $175,000. Medical students finish their training owing $55,000, a shocking-looking debt. But what is so terrible about it, rationally contemplated? Paying it off in 5–6 years at $10,000 per year leaves poor Dr. Kildare with only $165,000 annual income, so my heart doesn't bleed for him. Or possibly it may be paid up in 10 years or so, as it takes a while to get started.

There arises unavoidably a problem of what to do about those who (by lack of academic ability, inadequate motivation, or poor work habits) do not survive the educational process, for whom it would be a considerable hardship (not an injustice!) to pay back the money when they do not experience the anticipated economic advantages of the educational degree. This presents economic, political and even ethical issues beyond the scope of my book. A rational solution to these difficult problems involves giving due weight to several distinguishable and sometimes competing social interests. We deal with a mix of value commitments and empirical generalizations from replicable social science research. Unfortunately, this is one of those areas where considerations of justice (in the sense of even-handed treatment, holding persons accountable for the consequences of their free actions) and compassion may be in conflict, not because somebody is being irrational or careless, but because of the complex nature of human relations. Most of us are willing to bend principles or rules stemming from Aristotle's concept of distributive justice in order to achieve certain desirable kinds of social change, sometimes "social engineering," and out of compassion for persons subjected, through no fault of their own, to one or another kind of disadvantage. But, I would insist that any compromise among these forces should be so framed that persons are, if not mentally ill, accountable for their actions. We treat persons—metaphysical determinist disputes aside—as free agents who *choose* and *act*. Such free agency includes making choices that take risks in the hope of long-term advantages. Whether those who borrow money for two years of college and then drop out and get a job should have to pay it back depends not only upon the reasons for their dropping out, but also how well they do financially, despite having dropped out.

The idea that there is somehow an intrinsic human "right" to a college education at someone else's expense is not one I would accept. But I don't want my analysis to hinge upon agreement with me as to that, because I know there are many rational and informed persons who disagree. It could be argued that the best way to look at this (making compromises among aims of distributive justice, commutative justice, feasibility, and long-term social engineering) would be, by analogy, not with a mortgage, but with various kinds of insurance, where we collectivize individual risk. You may end up paying for my automobile accident, for which I was quite clearly at fault (and that's why the insurance company consented to pay without litigation), but no "injustice" is done you. We had no way of knowing that in advance; it could have worked the other way around, and both of us voluntarily paid our premiums to avoid individual catastrophe. In that sense, one might say that we all pay through taxes to create an adequate student loan bank, just as we all pay for fire and police protection, Social Security, Medicare, and the like. But I insist that despite the complexity and difficulty of solving these questions in a manner equitable, compassionate, and efficient, the basic idea that *those who benefit most from higher education should sooner or later pay most of its cost* should be implemented one way or another.

Chapter VIII

Taking the Sacred Cows to Market

In saying that there is important *interaction* of the Seven Sacred Cows, I am not using the word "interact" in the vague jargon of some social scientists, where it merely connotes (in a familiar and reassuring semantics) that (a) all sorts of things are happening and (b) we would not wish to think linearly or "atomistically," or to attribute more importance to one thing than another. By interaction here I mean the technical concept of the statistician (R. A. Fisher); that is, the quantitative impact upon instructional cost of challenging one of these Sacred Cows is itself a function of the extent to which another one is concurrently challenged, so that the net effect of jointly challenging two Sacred Cows is both additive and multiplicative. Believing in one of these Sacred Cows potentiates the influence of belief in another on increasing cost; challenging one of them potentiates the cost reducing impact of challenging another one. If one reflects upon the various concrete proposals that come to mind when challenging one of them, and the objections to those proposals, there are many examples of interaction effects.

Example: Suppose we decide that a certain specialty area is not of central importance to our department (challenging Sacred Cow V). We may still want to offer something in the area, without expecting to attract doctoral candidates for whom it would be the main concentration. The residual offering (reduced quantitatively and, usually but not always, qualitatively) is to be covered by a professor for whom it is a secondary interest, someone who would only claim moderate competency rather than excellence. To this we hear the objection that, in order to brush up her knowledge of that secondary field and add one course to her load, too much time will be taken away from her research. This objection is answered by concurrently challenging Sacred Cows I (both teaching and research must be done by all) and IV (most research is worthwhile). If they are not concurrently challenged, we cannot give an adequate reply to the objection.

Example: The challenging of Sacred Cow II about lecturing is impeded by the background presupposition of Sacred Cow VI, that as many people as possible should go to college. Suppose we eliminate a couple of lecture courses and a seminar in a low priority area, but retain the course domain under "Readings" or "Independent library study" (challenging Sacred Cow II). It will be objected that this may be all right for a few graduate students who want to explore the area; but it won't work for our (cancelled) undergraduate lecture course because you can't expect undergraduates to be sufficiently motivated and competent to do such reading on their own (implicitly invoking Sacred Cow VI, everybody should go to college). Sometimes this objection is expressed as, "That may be all very well for students in Germany or Scotland, but American students aren't able to do that." I should be saddened to believe that the American culture has

gone so far down the drain that we can assume so little initiative, energy, or competence in a college student; but if that is the case—as I do not myself believe it to be—somebody should look at what we're doing in their previous schooling to make them so passive, lazy, uninformed, and helpless. Presumably one reason for the stereotype that a college student can't read anything independently and comprehend it, or use a dictionary to look up a word, is all sorts of people coming to college who lack either the intellectual competence or the value orientation and motivational system appropriate for college tasks. Further, to the extent that the society (e.g., future employer at job entry) relies upon a person's possession of a bachelor degree as proving something about the individual, I maintain that a certain amount of self-study in the library is not unreasonable to require. It is a psychological truism (not found in many psychology books, except implicit in a chapter about "reinforcement"): *A competent person, not impaired by physical or mental disability, will do pretty much what is required in order to achieve something strongly desired*. I don't believe it takes a PhD in learning theory to know this; my grandmother, who never went past the third grade, was familiar with the principle. If the only way to get credit for a course that is not offered as lecture, but for which a syllabus of readings is specified and a final exam or other means of assessment exists, is to trudge over to the library and read two books carefully, then students who want those two credits on their grade record will do these things. Of course, if you *assume* that nobody can learn anything unless spoon-fed, that's the result you will produce in a sizeable number of students.

Example: Increases in teaching loads for non-research faculty (challenging Cows I and IV) have the realistic constraint of imperfect faculty fungibility (discussed below), and hence will not work for some departments unless Sacred Cow II (formal lecture as the optimal teaching method) is concurrently challenged, which we have just seen involves Sacred Cow VI as well. Hence we have higher than "first-order interaction effects" in the statistician's sense, here an interaction of four factors. Although I haven't cooked up an example, it seems likely that there are some areas, departments, and constellations of faculty interest in which high-order interactions (up to the sixth) would be realized, in which every one of these Seven Sacred Cows potentiates the interaction effect of the others taken conjointly.

The Two Money-Saving Strategies

First, ***reduce total classroom teaching hours required*** by a combination of:

(1) *Dropping courses*: The first phase eliminates course offerings, under any guise. We decide that interpersonal ceramics in the psychology department is a nice thing to have if the money is flowing freely, although even under cushy circumstances there seem to be some doubts about its long-term scientific merits; but nobody, not even its own practitioners, is prepared to define it as part of an absolute minimum core of education in psychology. Readers can easily pick their own examples in almost any

discipline, although I grant that the peculiar heterogeneity of psychology makes it a bit easier than in others (Meehl, 1986b). Let's say we are going to get rid of 10% of teaching time by ceasing to offer certain courses, perhaps ceasing to offer *any* courses in a certain field.

(2) *Managing lecture sessions more efficiently*: We can eliminate more of the remaining "live professor" classroom hours by offering selected courses less frequently, combining sections into larger classes, lecturing to multiple sections via closed circuit TV, and using video recordings for a few of the "standard, slow-changing" courses. At Minnesota we used the same General Psychology videotapes for 10 years, in effect reducing professorial classroom time by 90% for that course. A basic, undergraduate course whose content does not remain largely stable for 10–20 years is suspect, but in any case should not be taught by video that is quickly dated. If a clearly proved major "breakthrough" of knowledge occurs (e.g., in psychology, the 1954 discovery of positive and negative reward centers in the brain), a special lecture and associated readings can be prepared with negligible expense. A department handling one course in 10 by video has already achieved a 9% reduction by this tactic alone, without combining sections or offering courses less frequently. Giving one fourth of courses this way yields a 22% saving.

(3) *Substituting instructional modes.* Remember, it is not necessary to prove that these other instructional modes are *superior* to traditional lecturing, since they are all cheaper and they are easier to handle from the standpoint of faculty logistics and student scheduling. We have seen that the empirical data indicate a substantial equality of most other modes of instruction to classroom lecture, and a probable superiority over lecture for two of them (ETV and Keller's PSI method). It would not be necessary even to show equality; *close to equality* would suffice in the present straitened circumstances. Even a small decrement in student performance, say, 5%, resulting from the substitution for traditional lecture of an instructional mode that saved 30–40% of costs and increased curricular and logistical convenience would be acceptable to anybody but a perfectionist or pedagogical fanatic. However, the research shows that if there is any effect it will be an improvement.

I maintain there is nothing horrible, scandalous, or educationally hurtful in these consecutive reductions. I suppose the outright killing of 10% of courses is likely to be the most shocking to one who has not yet accepted the brute fact that something has to go because the taxpayer won't provide more money. But such a person has to reflect that unless the faculty are willing to have sizeable increases in their teaching loads, there *must* be a dropping of some things that had been previously offered if their salaries are to be markedly increased. The most casual inspection of a list of graduate and undergraduate offerings in a medium-to-large department should convince a rational mind that some things can be dropped. Cursory inspection of our Course Bulletin enabled me to come up immediately with a half dozen three-credit courses that are nice to have but dispensable

in my department. Two of them were courses of my own which have been dropped since my retirement; they were good courses, but they *are* dispensable.

Second, ***eliminate the publish-or-perish requirement for most faculty*** to increase professorial teaching time available.

Those classroom teaching time reductions collectively would suffice to achieve big savings. However, the most powerful leverage would come from a definitive, principled abandonment of the obsessive idea that one's institution must be a "research institution," meaning a new-knowledge generating institution, in turn translated as a scholarly-article-publishing institution, and—the most important part—*that this requirement must hold for all departments*. If that guiding fiction were abandoned and it were made formal policy that the great majority (say, 80%) of the faculty were purely knowledge-transmitters (rather than publication-generating), average faculty teaching loads (outside of research units) could easily be doubled. (A more rational figure for that division would be that 90–95% of faculty become pure knowledge transmitters, given the data discussed in Sacred Cow IV. That is, I think my estimate of 20% teachers-and-researchers is at least twice as much, and maybe four times as much, as is warranted by the impact of track record under the "publish-or-perish" tradition.)

It is pointless to play out the various numerical possibilities because they depend completely upon the department's field of concentration, the distribution of faculty over areas, the rank distribution, the amount of national demand, the amount of non-taxpayer support for portions of time that are unequal over domains of investigation, and the like. But to say, "If we were deprived of any specified proportion of faculty teaching hours available, we simply could not carry out the core educational function of the department," is unplausible. One could not defend such a statement without presupposing that *every* student in *every* program takes *every* course in the department, something which, as we all know, never happens and which is neither possible nor desirable.

Faculty Fungibility

Readers of earlier drafts have argued that I achieve dramatic reductions in instructional costs by assuming complete fungibility of faculty members with regard to what they are competent and willing to teach. My expository problem here is that calculating hours of class time saved and new hours available (most faculty being freed of pressure to publish) must be done in that crude way, because precisely how courses and hours will be allocated within a department must be individuated on the basis of the department's size, faculty composition, and major scholarly emphasis. It would be counterproductive, for example, to make the kinds of changes rendered possible by challenging all Seven Sacred Cows concurrently, as if we were in complete ignorance of which professors are better at large lecture section teaching and like to do it, versus those who are not good in that situation but who are highly stimulating to advanced, high caliber students in a seminar context.

Another sacred cow, not listed among my seven, requires challenging in order to handle the imperfect fungibility problem, namely, the fixed idea that every respectable department must have (a) top flight research productive eminent scholars and (b) brilliant classroom teaching *in all subdomains*. That semi-sacred cow is best challenged by pointing out that no department achieves this at present, under the conventional inefficient system. In fact, I have never heard of a psychology department that even attempted to do such a perfectionistic job. It is a matter of common knowledge that where psychology students go to take graduate work depends on what they want to specialize in. This is to some extent true even for undergraduate majors, provided they have the financial means to live away from home and to pay high fees at select undergraduate institutions. In clinical psychology, for example, no student seeking a PhD and primarily interested in psychoanalysis and projective methods would come to Minnesota. Combining departmental "official statements" about the program with a cursory look at the catalog of course offerings, plus the published research of major faculty, plus scuttlebutt from one's undergraduate advisor or other students who have gone here—these inputs in the aggregate are an almost perfect guarantee that the matching of a student's intellectual passions and ultimate vocational goal with a particular department will be fairly good. Clinical psychology students mainly interested in behavior modification would not elect to study at Michigan, or at Adelphi if mainly interested in mathematical taxometrics, or at Berkeley if mainly interested in behavior genetics. There are highly specialized but important areas of psychology which are literally not available in some departments. For example, biomechanics, "the scientific study of the human in the machine," is not available at the great majority of psychology departments in the United States at the present time. Only a few schools offer specialty training in behavior genetics. Several first-class departments (e.g., Harvard, Stanford) do not have an APA accredited training program in clinical psychology.

We start with a social fact about the differentiation of institutional emphases that is commonly not in dispute: In the case of state institutions, it is economically out of the question to make sure that you are offering top-flight instruction at graduate and undergraduate levels including exposure to internationally known scholars in every domain of psychology. However, there is an unfortunate (because unrealistic and costly) tendency to lean in that direction; and when this tendency is combined with subscription to the Seven Sacred Cows, it makes the fungibility problem intractable.

Suppose the cows are all challenged concurrently, and every department with limited tax support (and the political necessity to hold down fees, at least for state residents) announces explicitly *what everybody understands whether they talk about it or not*: "This department is going to keep its strengths in areas A, B, and C; it is going to offer what is minimally necessary in D and E; and it will offer almost nothing in specialty area F." Such an announcement is simply honesty in advertising. (It can be avoided because, as one prof-bashing book states bluntly, being a college professor is the only occupation one can succeed at without satisfying the consumer!) *Example*: A department with a strong

applied psychology emphasis—a heavy focus on clinical, counseling, and industrial psychology—will nevertheless have offerings in such basic science fields as sensation and perception. Presumably a department that did not even offer a senior college course in perception would not remain accredited. But that does not mean it must have a full professor whose field is audition, and another full professor whose field is vision. If we make clear that students should not come to our school to get a PhD in the perception domain, there isn't any reason why one professor could not teach both undergraduate courses in these areas. The Sacred Cows about publication being repudiated, this professor is not one of our knowledge generators, thus can take time to achieve an advanced undergraduate instructional level mastery in both sensory modalities. When I was an undergraduate and graduate student (1938–1945), Minnesota's psychology department (rated among the top half-dozen in the country since 1925), offered only a single (senior college) course called "Sensation and Perception." Admittedly that was spreading it pretty thin, and I am doubtless somewhat uneducated thereby. But this domain weakness has not prevented my becoming a competent clinical psychologist. *Example*: A different sort of department of the "traditional, brass instrument" orientation does not offer graduate training in clinical or counseling psychology. All of its faculty are engaged in experimental psychology in the "hard science" areas such as Perception, Memory, Animal Behavior, Learning, Neuroscience, and the like. It is, however, recognized that there should be available at least a three-credit course in descriptive psychopathology for undergraduate majors. Some graduate students might be interested in the bearing of auditory hallucinations on auditory perception theory. Challenging the Seven Sacred Cows concurrently, the problem is easily solved in a variety of alternative ways or combinations thereof. One way is to use videos prepared at other institutions. Another way is to combine a few visits to a state hospital with videos of various kinds of mental patients being interviewed, and the students doing monitored or unmonitored library readings of a good textbook in abnormal psychology. Or you can rely on programmed texts, from which the student will learn and retain more than from attending a traditional lecture course. If your institution has a medical school, you can pay a cross-charge to one of the younger psychiatry faculty to teach descriptive psychopathology. (Psychology students enjoy hearing from a psychiatrist, and will get exposed to another discipline's way of operating.) Still another way is to hire a local practitioner on a part time basis, which is easy to do because practitioners often want the combination of prestige, library privileges, contact with academic scholars, and such perks as advance football tickets. Even if you pay the practitioner a pretty good fee to make up for loss of time in private practice, the cost of such a person teaching a three credit course in abnormal psychology will be small compared to hiring a full professor to "cover the area." Fungibility is a problem, but all it requires for solution is challenging the cows with flexibility, rationality, firmness, and a smidgeon of imagination.

Savings Achieved by Mass Slaughter of All Sacred Cows[23]

There is a simple equation for computing reduced professorial teaching cost when the several cows are conjointly repudiated. For instance, when reductions in teaching hours to be covered are combined with confining the research function to a minority (e.g., one-fifth) of faculty, so that faculty freed from research requirements can have larger teaching loads, professorial cost can be reduced by 75% or more. This earth-shaking cost reduction, which most academics find incredible on first hearing, is accomplished by challenging all of the Sacred Cows and acting accordingly, together with the simple household principle that pennies add up to dollars.

Let us make some reasonable estimates and see what total savings can be achieved by adopting all of my suggestions for saving faculty time required to cover the classes. A reader who is reluctant to adopt certain of the changes, or to apply them to as large a proportion of the courses as I advocate, may calculate savings on the basis of some other distribution of changes. For example, replacing conventional lecture mode by the more efficient Keller Method can easily be defended—both as to time saved and better student achievement—for the majority of undergraduate classes, whereas my assigned value for this illustration will be only a hyper-conservative 20% of classes. Only those rare subject matters requiring frequent recourse to Socratic dialogue (e.g., Kantian ethics, recent controversial research on behavior genetics) cannot be best taught by Keller's PSI method. While I haven't been draconian in assigning the proportions of courses to change, the residual *faculty time fraction associated with each of the changes singly* has not been chosen optimistically but rather as what I consider a least plausible saving.

Table A illustrates estimates for proportions (P_i) of courses to be handled in various ways. For each changed instruction method, we also have an estimate of the amount of professorial time that will still be required (R_i), expressed as a fraction of the original professorial time the course required. Thus, either combining sections to double class size or offering a class half as often will require only 50% of the original instruction time. If we adopt the Keller Method for a class, I have estimated that the revised offering will require only 40% of the time a professor spent previously. Not changing the way a course is handled (20% of courses in this example) will require the same amount (100%) of time as before. If P_i is the proportion of courses to be handled in a specified way, and R_i is the residual fraction of faculty time required after the changes, then the new total amount of faculty time needed is $\Sigma P_i R_i$: the sum of the products of the proportions (P_i) of courses handled each way, multiplied by the residual faculty time required (R_i) after the changes.

---

[23] [*I am grateful to Niels Waller for help in making this section more easily readable.—LJY*]

TABLE A. Reduced teaching time required when courses are handled differently.

Assume that initially 100% of classes are lectures by professors on a regular schedule.

New ways to handle instruction:

P_{EL} : Eliminate courses.

P_{FS} : Fewer Sessions by increased class size, use of closed-circuit TV, or classes offered less often, to average ½ as many sessions (or possibly fewer).

P_{IS} : Independent Study, requiring minimal monitoring.

P_{ETV}: Educational Television produced locally, rented, or purchased.

P_{KM} : Keller Method; lecture biweekly, meet one hour per week with instructors, TA, and honors student monitors.

P_{NC} : No Change to traditional way of handling course. Selected on basis of level, role of discussion, lecturer talent and interest, new knowledge on topic, department's interest, orientation, community resources, ideology, graduate programs (if any).

Proportion of courses to be handled each way: P_i	×	Revised instruction time required: R_i	Comments on R_i estimates	Proportion of original time still needed after change
P_{EL} = .10	×	R_{EL} = .00	Eliminated classes will require zero time	= .00
P_{FS} = .10	×	R_{FS} = .50	Session frequencies reduced by ½ on average	= .05
P_{IS} = .20	×	R_{IS} = .15	Minimal monitoring required	= .03
P_{ETV} = .20	×	R_{ETV} = .15	Perhaps a high estimate of time required	= .03
P_{KM} = .20	×	R_{KM} = .40	Based on experience with method so far	= .08
P_{NC} = .20	×	R_{NC} = 1.00	Same time required as before	= .20
Total = 100%				

Teaching time to be covered as a fraction of initial faculty teaching time = $\Sigma P_i R_i$ = .39

NOTE: To use other estimates, alter percentages P_i of courses to be handled in different ways (always keeping total courses at 100%), or try other R_i estimates of the fraction of time a new instruction method might require (compared with previous time needed; time required for P_{EL} courses will always be .00, and for P_{NC} courses it will always be 1.00). Then calculate your new $\Sigma P_i R_i$.

Slaughtering all seven Sacred Cows at once, I changed the handling of original lecture session courses with the following proportions: 10% are Eliminated; 10% are offered more rarely or by doubling class size to yield Fewer Sections; 20% are offered by Independent Study with minimal faculty monitoring; 20% will use rented, purchased, or locally produced ETV; 20% will be taught by the Keller Method; and the remaining 20%

of courses remain taught as they presently are (No Change). After making these changes, the professorial time required to cover the total course load has become $\Sigma P_i R_i = .39$ of the time initially required. We have reduced faculty total teaching time required by 61%.

Given the reduced amount of instruction time *needed*, we then proceed to make more teaching time *available* by getting rid of the publish-or-perish requirement for some fraction of the faculty. I make the conservative estimate that 80% of the faculty will be pure knowledge transmitters and 20% of faculty are in departments associated with research units; and I take a doubling of teaching load as a safe lower bound to use for non-research faculty members. This teaching load is less than the ratio we get if we compare current loads in research universities with those before World War II (even where there was some pressure to publish) or with small liberal arts colleges that have not tried to change into research institutions. Thus, in Table B we have multiplied the proportion of faculty who will become knowledge transmitters (teaching only) by .50 (only half will be needed if they are teaching twice as much), and added that to the proportion of teaching + research faculty who will remain both knowledge generators and transmitters and for whom the teaching load remains the same as before. The result is .60 of the old faculty available to cover the teaching load, a 40% reduction in original faculty cost.

TABLE B: Reduced faculty cost when publication requirement is eliminated.

Assume 80% of faculty will be *knowledge transmitters*.
These professors will teach only, and will have double the initial teaching load.

Assume 20% of faculty will be *knowledge transmitters and producers*, in departments with research units. These professors will have research and publication requirements and will have the same teaching load as before.

Then the **faculty fraction available to cover the teaching load** will be:

$(.80 \times .50)$		$(.20 \times 1.00)$		
Teach only	+	Teach and research	=	**.60**

Finally, the new cost of instruction is the product of the new faculty time needed to cover teaching and the reduced faculty cost of providing that time:

Proportion of original teaching time to be covered		**Proportion of original faculty available to cover teaching time**		**Proportion of original teaching cost still needed**
.39	×	.60	=	**.23**

With the estimates used in this example, the new cost of instruction is less than ¼ of the present cost! *We have saved 77% of the initial cost of instruction*.

For those who like to see everything presented in one grand equation, the new *reduced faculty teaching cost* is given by:

$$\begin{aligned}\text{New Cost} = [&P_{\text{Eliminate}}(.00) + P_{\text{Fewer Sections}}(R_{\text{Fewer Sections}}) + P_{\text{Independent Study}}(R_{\text{Independent Study}}) \\ &+ P_{\text{ETV}}(R_{\text{ETV}}) + P_{\text{Keller Method}}(R_{\text{Keller Method}}) + P_{\text{No Change}}(1.00)] \\ &\times [F_{\text{Teach}}(.50) + F_{\text{Teach \& Research}}(1.00)]\end{aligned}$$

Where

$P_{\text{Eliminate}}$ and $P_{\text{No Change}}$ are assumed to require zero and 100%, respectively, of the initial teaching time.

F_{Teach} assumes faculty who only teach will have their teaching loads doubled (multiply by .50), and $F_{\text{Teach \& Research}}$ faculty will continue with their initial teaching loads (multiply their proportion by 1.00).

Then the amount saved will be 1.00 minus the new, reduced teaching cost.

Because there is considerable (I think irrational) resistance to courses being taught by "outsiders," I have not included the savings achievable that way, although in applied fields (e.g., clinical psychology) such community practitioners often make a unique contribution.

I do not discuss the allocation of this huge saving to the three classes of plausible beneficiaries. This is a policy question and the answers could vary quite a bit from one institution to another. But to give one easy division, in a state university we might decide to split the savings equally among three affected groups. Or, one quarter of the savings could be taken out of the tax support, and another quarter applied to lowering student fees. Having done that, we still have 50% of the original dollars, whereas the new cost is 25%, so we now have twice as much money to pay professors. We could *double* faculty salaries while lowering fees and taxes! Students, their parents, and the taxpayer would be overjoyed. As I have said, despite the sacredness of the slain cows, I think all but the most fanatical cow-worshippers in the professoriate would find the prospect of doubling their salaries appealing. Who is harmed by these drastic appearing changes? I can't see that anybody is. Poll figures on academic attitudes toward publication pressure suggest that around $^{2}/_{3}$ of faculty of colleges and universities would welcome the changes.

Chapter IX

Other Bum Steers and Anointed Bovines That Waste Money

A heterogeneous collection of additional academic beliefs and practices are wasteful of faculty time and, hence, the taxpayers' and students' money.[24] Some (e.g., reverse discrimination) cost only a little bit, and their cost is indirect; others (e.g., feckless unnecessary meetings) waste a sizeable amount of a professor's working week. In the aggregate, I am confident they consume *at least* 10% of faculty time, and I think most experienced academicians would say that is an underestimate. Elimination of this waste motion would reduce the 25% residual cost computed in Chapter VIII to a mere 15% of the initial expenditure, if faculty time units were completely fungible. I have not tried to take account of this 10% (or more) savings in the master savings equation because I do not know how to include it; my equation throughout aims to be conservative, and professors don't come in divisible subunits and are only imperfectly fungible on the teaching side. If we were to assume every professor in a 35-member psychology faculty has 10% of time freed up by reducing these smaller inefficiencies, we still have to pay each professor's full salary, so it looks as if we aren't saving any money.

But it would be a mistake to infer from the nonfungibility and nondivisibility of faculty that there would be no savings from a diminution in these diverse inefficiencies. One can easily see three cost reductions which, although indirect, should not be ignored. The first arises from team teaching, where the nature of a course is such that different portions of it are taught by different faculty members. For instance, I taught several sessions of an applied psychology course in which eight faculty participated. When I chaired psychology, I ran a three-quarter "systematics" course in which each quarter was split between two faculty members. Some departments rotate their large introductory class among faculty so each has responsibility to teach it every several years. In such situations, there is partial fungibility, so that adding up fractions of several professors can mean the equivalent of a full-time faculty member. Or if the time saved summed to a half-time member, another half might be handled by a cross-charge on a related department or a part-time lecturer from the outside community.

The easiest way to save money via that 10% average reduction in wasted hours is with respect to *adding* faculty for a specific teaching task not presently handled but deemed important. If we are not focusing on covering or expanding a research domain, but think we need a couple of courses in a domain where we already have faculty members who are or can readily become competent (not for advanced seminars but, say,

---

[24] [*The ordering of topics in this chapter does not imply any particular priority among them. Probably the most determining factor was when they were transcribed or editing was sufficiently final.—LJY*]

respectable senior college courses), pooling the 10% savings of several faculty to handle those two courses, especially if they are alternated so each one is offered only every other year, obviates the need to hire a new full-time faculty member.

Even on the research side, there is an indirect possibility of significant savings. In designating certain departments whose faculty are high research producers (quantity) and whose research has high impact and visibility (quality) as "research departments" (i.e., with research units attached), and doubling the teaching loads of non-research departments, I had not assumed that nobody in the pure teaching departments does any research. Prior to World War II, faculty in first-class universities were publishing research despite having average teaching loads twice or three times what is now considered reasonable. We may reasonably assume that the faculty in non-research departments who do publish research will tend, statistically, to be competent and highly motivated. Hence the potential "research loss" attributable to the research/non-research administrative division will not be a total loss, as if we assumed that nobody in non-research departments ever published anything. That being plausible, our anxiety about the basic policy should be lessened. We might therefore feel somewhat more comfortable moving the cutting score for classifying a department as a research unit higher on the productivity-impact dimension. Moving a professorial slot from research to non-research status always saves money.

Examining the contents of these additional inefficiencies and time-wasters, a case can be made that they are more faculty-fungible than lecture classes, research, tutorials, or counseling students. *Example*: Shifting from essay to multiple choice exams will save *h* hours of grading time whether the shifter is Professor X or Professor Y, and aside from whatever else they spend their time doing. However, I have no trustworthy way of computing the savings for these additional factors, and—more importantly—I believe that most of these inefficiencies are rooted in some basic human traits that are highly resistant to rational intervention. They are also "institutionally" fostered, academia having all the main characteristics of a bureaucracy. One may be able to convince people that essay exams are usually unreliable (although *three generations* of research has not succeeded at it); but persuading people who love to blather in committee meetings—and would feel guilty for not earning their keep if this time-filler were unavailable—to cease wasting so much time this way is, I fear, a hopeless undertaking.

1. Quaker Consensus

I have alluded before to the frequent failure of psychologists to think like psychologists when they are not in the laboratory, library, seminar room, or clinic (e.g., p. 33). (It is well known among scientifically trained clinicians that a large number of their clinical colleagues, not so scientifically trained, do not think like psychologists in the clinic either!) One context where this puzzling lack of "transfer of training" wastes time

and reduces the quality of the decision maker's work product is in faculty or committee meetings discussing matters of curriculum: e.g., What should be our requirements for the BA degree major in psychology? Should there be any required courses for the PhD that are considered "core" across all subspecialties? What domains should be covered on the written PhD prelims? I have seen instances in which the group's actions were paralyzed when confronted by the ubiquitous fact of human disagreement, either about empirical matters that are not clearly settled by evidence, or about value commitments, or both. An extreme example of this occurred when I was a member of a "blue ribbon" committee of the American Psychological Association in which the content of the training for clinical practice was under discussion, and naturally the problem arose of accrediting departments taking the committee's offerings (especially their required offerings) for the doctorate into account. Several voices—fortunately a minority, but not a negligibly small one—were raised strongly objecting to the setting up of *any* curricular requirements for the doctorate in clinical or counseling psychology, the main reason being that "not everybody, even in this committee, let alone in departments, will agree about everything." To offer this objection to listing a set of educational standards for students who are going to take professional responsibility for the diagnosis and treatment of mental illness, testify in court about criminal responsibility, give advice to judges about sentencing of dangerous offenders, write reports on the suitability of foster parents, and the like (including even matters of life and death—some mental patients are at risk for suicide or murder) on the ground that the degree of consensus among informed practitioners and scientists in the field is less than Quaker unanimity is, of course, absurd. I could rest this argument on the elementary point that *all* social decisions involve less than one hundred percent unanimity, so if Quaker consensus were a necessary condition for social action no legislature could enact or repeal a statute, and the whole process of social life would come to a grinding halt.

Aside from the political absurdity of requiring Quaker consensus in complicated matters, what is striking about this is that psychologists, of all people, are most familiar with a variety of ways in which human judgments can be investigated and combined for certain practical purposes. *Example*: The Spearman-Brown Prophecy Formula was developed before World War I for predicting the increased reliability of a mental test that can be expected when the test is lengthened by adding more items which are qualitatively homogeneous and similar in their statistical properties to the previously existing items. Already in the 1920s, it was shown that under certain conditions human judgments about other humans' traits (e.g., leadership, intelligence, shyness) are fairly well predicted by the Spearman-Brown formula. Thus, for example, if the average correlation between two clinicians' ratings of mental patients' anxiety level runs around $r = .60$, then the formula tells us that if we want to attain a pooled judgment with reliability $r > .90$, we can achieve this by averaging the judgment of seven clinicians. A vast body of evidence has been gathered by psychologists in the applied areas (industrial, military, clinical, counseling psychology) about how best to extract reliable judgments' statistical

properties. These methods can be applied, *mutatis mutandis*, to other imperfect consensus decision problems.

If we ask what a PhD in psychology from the University of Minnesota department should be required to know, we proceed as follows. A set of knowledge domains is prepared and each briefly characterized by a sentence, phrase, or term typed on a 3×5 card (e.g., 'visual perception,' 'intelligence testing,' 'neuroanatomy'). The initial batch of topics or domains is easy to construct by asking faculty and students to submit items, scanning the current course offerings (whether required or not), and reviewing previous lists whether they've been amended or not. Similar lists of knowledge domains have been prepared by national committees on training, by site visits, accreditation committees, and by certification boards like the American Board of Professional Psychology and state licensing boards. Possible new areas not presently available are elicited by an open-ended question to students, faculty, and clinical practitioners. At this early stage, nothing—*however eccentric or unfeasible*—is screened out. The main point at this stage is not to do any editing or culling except for obvious overlap of quasi-synonyms. If there is any doubt as to whether certain labels are synonymous (thus, for instance, in some departments no course in individual differences or differential psychology is offered but, if questioned, the department alleges that its course in mental testing covers this domain; we discover that sometimes it does and sometimes it doesn't) there is no need to argue, because at this stage *the rational thing to do is to separate similar topics rather than wrongly lumping them*. If you lump topics, you don't have the data to treat them separated, whereas if you separate them and later find that their semantic content and statistical behavior make them sufficiently close to each other, then you can lump. This is a basic principle in adjudicating controversies between "lumpers" and "splitters" in all fields of scientific knowledge, and yet one finds psychologists fretting about whether they should count a course in classical psychometrics as being the equivalent of a course in differential psychology. The rational answer is no, *not at this stage*.

Each faculty member is then provided with this set of cards and instructed to sort them on, say, a ten-step scale (I was taught to cut it fine, but I have no objection if someone wants to argue for fewer steps) as to how "core," "fundamental," or "central" to being a psychologist this knowledge domain is. I do not discuss the merits of free, forced, or semi-forced (my preference) distribution of the judgments, but here again we have a sizeable body of research evidence concerning the pros and cons of the various procedures. We then correlate individual judges with one another pairwise, and the resulting Q-correlation[25] tells the extent to which Professor Jones and Professor Smith agree in the relative importance they assign the domains. This inter-judge correlation matrix is factor analyzed, and we assign to each judge a weight that is a function of that judge's loading on the first big factor that emerges from the analysis. At this step we

[25] This terminology for correlating the judges is used to contrast it with the conventional R-correlation in which placements of the several items are correlated with each other over individuals.

already have some empirical evidence that bears upon the controversy whether there should be any core curriculum or not, since if this matrix of Q-correlations fails to reveal a large *first* component, we have empirical support for the complaint that there is too little consensus—whether on a factual educational or a value question; at this point we needn't decide that—to warrant compelling all students, regardless of area, to meet any requirement. Given the big first factor, we assign weights to the judges on the basis of their loadings on that factor. The rationale here is that the existence of such a large first factor *in the inter-judge correlation matrix* testifies to the truth of the conjecture that the judges are by and large, although with perhaps some deviate outlier exceptions, "doing roughly the same kind of thing"; but if there is an appreciable spread of the factor loadings it means that some of them are "doing it better" than others. This doesn't mean that some judges are mistaken, it just means that they are outliers with respect to what group consensus empirically exists. For each item we can then construct a composite evaluation of its importance by pooling these weighted ratings. This composite evaluation yields a distribution of the group opinion as to merit per item over the set of items. Taking account of the logistical constraints, which can be set beforehand on some reasonable basis—if necessary, by a similar kind of Q-correlation—we simply go down the list until we have covered as many items as the constraints allow. No PhD program will be able to require as many formal lecture course credits as demanded for the BA, for example.

If our previous considerations of student time available and of feasibility and faculty contentment set a rough limit of, say, twelve domains, we simply count down our ordered list and stop at twelve. I do not mean to suggest that everybody will be completely happy with the result of this procedure. I am assuming that faculty are sufficiently rational, flexible, and conscious of their own fallibility so that something close to Quaker consensus was initially achieved with regard to carrying out this procedure. Thus, for example, I was required to take a six-credit course on individual and group (gender, class, age, race) differences in various traits, particularly intelligence. That emphasis upon the study of human traits as they vary over persons and groups was an emphasis for which Minnesota was world-renowned but deviant. (Even today I think it is accurate to say that while all psychology departments offer a course in mental testing, none of them require it of all of their students; and over half of departments, while they offer a course in psychometrics, a course in statistics, and a course in mental testing, do not offer a course labeled "differential psychology.") I would like to see this course included as part of a PhD core, but I am quite relaxed about the prediction that it would not be so judged by the application of Q-correlation to our current faculty. On the other hand, failure to include a course in elementary statistics would bother me a great deal; yet, even there, while I would be surprised at such an outcome, if I had antecedently committed myself to the idea of a core curriculum and the Q-sort procedure for deciding it, I would voice no objection.

I do not mean to suggest, of course, that empirical correlations can take care of all differences between scholars that reflect basic value commitments or speculative

opinions about how to foretell scientific developments. For example, I am convinced that the future of a scientific psychopathology hinges on the development of three domains, namely, psychometrics (including taxometrics, the mathematical theory of classification), psychophysiology (including "soft neurology"), and behavior genetics. The tendency of our abler graduate students today to be attracted to one or more (sometimes all) of these three approaches to studying mental illness is suggestive, and I find it reassuring. But it would be arrogant and dogmatic of me to insist upon this conjecture about the future course of psychopathology. Therefore, while I would give a high rating to a senior college course at least (and preferably a graduate seminar) in behavior genetics as core to clinical psychology training, and would do my best to electioneer for that during the rating period, I would be resigned in advance to being defeated on that point. If that happened, I could confine myself, as a PhD advisor and classroom teacher, to pushing my view with the hope that students who were not required by a departmental regulation to take such a course would be moved to do so voluntarily. And, of course, there would be nothing about the departmental curriculum requirements that would prevent me as an individual PhD advisor from requiring my own candidates to learn some genetics. That element of "academic freedom" would have to be clearly guaranteed as a precondition for my agreeing to abide by results of the Q-sort procedure in the first place. "Core" means an advisor cannot dispense students from learning a core domain, not that the advisor cannot prescribe anything "non-core" for his or her candidates.

There is a difference between the professor who is a "solo" deviant because of an idiosyncratic, perhaps neurotic, fixation on some particular subject matter or research method that nobody else considers important, and a minority opinion consisting of several faculty who are outliers with respect to the majority but who have a well-reasoned, articulated position that the larger group finds intellectually responsible and somewhat persuasive but still not dispositive as to what should go into the core. The history of science shows clearly that such deviant subsets may, from time to time, be more insightful than the compact majority, but are shown to be so only in the long run. One rational and socially acceptable way of handling this problem, since we do not have access to the verdict of history when making our curricular decisions, is to achieve quasi-Quaker consensus as to a procedural matter that does not presuppose consensus as to the substance. I don't understand why it is so difficult to get social scientists to think this way when lawyers, historians, and political scientists take it for granted. *Example*: Having set a rough upper bound on how many courses can feasibly be defined as core, and recognizing explicitly that individual PhD advisors can impose one or two additional requirements upon any doctoral candidate who wishes to work with them—a basic position taken for granted by all departments as well as by the graduate school—we then cut back the number that will be defined as "core by high consensus" sufficiently to allow for a small number of courses to be defined as "core without high consensus," in the event that such a well-reasoned minority report emerges. This approach can be thought of partly as a matter of necessary compromise in order to keep people happy, but that is not

its only rationale. The other element required in order to handle the disagreement in this fashion is that the members of the majority group realize in advance of collecting the data—perhaps they know it from general observation of faculty conversations, and have an expectation as to who the strong minority will consist of—that this minority group *may in fact be correct* as judged by the verdict of history. What we are then saying is there is high consensus about the central core, and lack of consensus about a few courses less central, but that we want to take account of the possibility that the minority are correct in holding that these should be considered core. So, we are doing two things. We are lubricating the social machinery with respect to a non-tiny minority who feel very strongly about a certain domain of competence; and the rest of us, while we do not feel so strongly, recognize that they may very well be right and that we are therefore playing it safe.

This is playing it safe from the standpoint of the graduate students, whether they begrudge it or not, because studying two or three classes that it turns out you did not need to have will not damage your scholarly career, whereas not mastering a domain that turns out to be of great importance is a much more serious mistake. I can illustrate this by an example where I am a member of a minority (consisting of more than a couple of persons) that has never been able to persuade most of the department that psychologists should study some undergraduate mathematics, at least through calculus. Social scientists learn statistics (though frequently not the most useful kind, in my opinion), but they do not learn much mathematics, and in some of the "soft" fields such as clinical and social psychology, most of them don't know any mathematics. Since the power of the advanced high-quality sciences is regularly associated with two factors—a *developed technology of measurement of the observational variables*, and a *powerful formalism*—it is arguable that improvement in the status of the "soft" areas of psychology will hinge partly upon some degree of mathematical competence in its investigators. This is not the place to develop that argument, so I will only point out that we currently have a kind of vicious circle, in that PhD candidates in the soft areas don't think they need to know any mathematics because their advisor is a tenured professor who doesn't know any mathematics, so why bother? Besides, mathematics is difficult, a good deal more difficult than almost anything else one studies in psychology. If the published work in one's field does not use mathematics, one is not hampered in reading it by one's mathematical ignorance. Of course, if nobody in the soft areas knows much of any mathematics, then this vicious circle cannot be broken. We cannot come to a reasoned conclusion about the potential of mathematics in the soft areas without some bright people having a try at it. But even bright people in the soft areas can't have a try at using mathematics if they never studied anything but cookbook statistics. When I was chair of the psychology department at Minnesota in the 1950s, I made a heroic effort, supported by two or three of my dozen colleagues, to set a firm mathematical requirement for at least one of our three undergraduate majors, but I was unsuccessful. There is a place for a kind of "general education" BA in psychology for students who do not intend to do graduate work, and such a

major would probably be as helpful in being a better parent, voter, citizen, taxpayer, etc., as most other majors in a liberal arts college. For these students, I would insist upon basic statistics, such as how to look at the misleading bar diagrams and graphs the media publish, but I would not insist upon their learning mathematics as such. For students contemplating graduate study in psychology, either in the applied or theoretical domains, I would impose a requirement of at least elementary calculus. Today, it is possible to learn about algebra, analytic geometry, basic matrix algebra, and differential and integral calculus without taking the fifteen or twenty credits required when I was an undergraduate psychology major.

In implementing the idea of a strong but not quite persuasive minority position, we compute the usual R-correlation between pairs of courses or perhaps cluster analyze them, which would identify a "mathematics cluster" of the ratings. We then identify those professors who are statistical outliers in the Q-correlation analysis, which will enable us to distinguish between those who are solo deviates from those who are somewhat outlying but along with others who are outliers in the same way (namely, they give a moderate-to-high rating to mathematics courses). By combining the information from the course clusters with the Q-correlation information about a subset of faculty outliers, we can come to a reasoned conclusion that there is a strong minority report as to the importance of mathematics. This might lead us to insert a small mathematics requirement into the core, and undergraduate or graduate students who satisfy this will therefore have prerequisites adequate for individual doctoral advisors who belong to this minority to specify whatever more advanced math courses they deem appropriate.

The main point is that one settles in advance upon an evaluative procedure, and part of that initial overarching settlement would be an agreement to drop the whole idea of a core in the event that either (a) the first factor accounted for below a certain agreed upon percent of the variance in the judgments, or (b) the minority of deviant judges showed some considerable clustering among themselves and their ratings were sufficiently extreme to show that they had very strong feelings about the matter. It would be rational to say in advance that if there is not even quasi-consensus when judges are weighted by their first factor loading, or if when given such a consensus, the deviant minority judges are more than one in two in number and their ratings are very extreme in opposition to the factor-loaded pooled judgments, then it is not a good policy—in some sense not "reasonable" or "fair"—to impose such a core curriculum on all PhD candidates.

I hope it is clear from this that I am not saying that all questions can be statistically decided. One can invoke political, economic, or ethical considerations *in advance of the empirical results of the judgment process* so that hardly anybody—ideally, not a single faculty member—will have just cause for complaint. (I leave aside the extent to which graduate student judgments should be included in this process, which I consider an open and complex question that I don't wish to go into here.) If it is objected that the proposed procedure is *time-consuming*, my answer is, first, that setting up requirements that all students must meet and that individual advisors will have to put up with, even if they

were in the minority, is a sufficiently serious business all around that it deserves a considerable amount of time to arrive at judgments that have demonstrable validity; and secondly—a point that many tend to forget, especially if they enjoy meetings—I doubt that the total person hours involved in this judgmental procedure would in fact be appreciably greater than the number of hours wasted in discussion in faculty meetings and committees and over the lunch table which take time and arouse passions but do not eventuate in a defensible group decision.

2. Excellence

Several years ago, I was invited by a former president of my university to participate in a panel discussion, the other participants being both local faculty and professors or administrators from other institutions, on how to achieve and maintain "academic excellence." Doubtless my declining this invitation reflected my personal distaste for conferences and my firm conviction that ninety percent of them are a waste of time (see Section 5 of this chapter, on meetings), but I did have a more specific reason for not participating in this one, which I attempted to convey in my letter declining. I do not know whether the panel ever took place, but I never read anything about it or saw a report so perhaps others were uninterested as well. The main thing I said to the president was that one ought not to spend the taxpayer's money on a conference flying people in from all over the world and paying their expenses for several days without first ascertaining from readily available, costless, local personnel whether there was any substantial disagreement about the question to be discussed. Of course, it could be that views about how to get scholarly excellence would to some degree reflect some other sacred cows of the academy that I have not here examined; so I was not suggesting that consensus, even if Quaker unanimity prevailed, would be conclusive proof of anything. However, there is a difference between the sacred cows criticized in this book and a quasi-unanimous conviction by department chairs or other eminent faculty, because there are some objective criteria of academic excellence we can turn to (cf. Sacred Cow I, p. 25).

Let me be clear that I am not objecting to a scientific (empirical) research study of scholarly production in relation to policies and other administrative variables that might be subjected to some kind of measurement. Like other psychologists, I always prefer systematically collected quantitative data, *provided they are of high quality and clearly relevant*, to anecdotal impressions. But a big conference, a talk-fest of purported "experts," chosen by virtue of their former administrative roles in distinguished departments, is not a research enterprise, rather it is a (perhaps biased) Gallup poll of opinions. My point was that it is wasteful to spend money collecting a bunch of mere opinions if you have/find quasi-Quaker consensus among more accessible and less expensive personnel locally available.

The psychology department at Minnesota has ranked among the top dozen or so in the country since the first rankings were made in the middle 1920s, in recent years has

moved up, and in three areas (clinical, counseling, and industrial) has been for some time tied for first or second places in the country. I am familiar with the views of the nine department chairs, beginning with Professor Richard M. Elliott, the first chairman when the department was created in 1919 (and including myself as chair 1951-57), and there is no question about their position. I have also discussed academic excellence with several chairs of other first-rate departments on this campus. I have been an adjunct professor in a half-dozen departments. I have seen numerous professors from other departments as psychotherapy patients. On the basis of these experiences, I have formed a definite opinion that the views of the nine psychology chairs are the same as those of others. The practices of first-class departments when contrasted with mediocre or weak departments are easily discerned without making a formal statistical study. When nine chairmen of a consistently excellent department state confidently the conditions for academic excellence, despite themselves representing different fields of psychology and having markedly different administrative styles, I look upon it as pretty good evidence despite its informal anecdotal character. In order to have a first-class department in terms of intellectual stimulation of students, turning out PhDs who themselves achieve distinction in the field, and writing papers that achieve high citation rates in the scholarly literature, there are three conditions that should be met.

The first is that the *evaluators of personnel*—in assessing job candidates, giving raises, promoting to tenure and to full professor rank—*must themselves have scholarly excellence* at least in a statistical sense. That means more than that a bare majority should have it, although it is relatively harmless if there is a sufficiently small minority who lack it. Since in our department a two-thirds vote of the faculty is necessary to grant tenure, a minority of one-third mediocrities is too big to safely ensure continued excellence. The reason why this condition is necessary is quite simple, namely, "it takes one to catch one." A competent but not first-class professor of physics at East Overshoe College is of course capable *today* of saying that Einstein was first-class. But he might very well not have been able to discern that before 1905, as is suggested by the fact that Einstein was unable to get an academic job and when he conceived and published the theory of relativity he was working as a clerk in the Swiss patent office! Persons of minimal competence are not reliable judges of the difference between first- and second-rate minds, and furthermore, they are frequently ambivalent in their attitude toward brilliant people. Of course, since in the academy "brains is the name of the game," hardly anybody is going to come out openly and speak against intellect. For a college professor to be against brains is like being against God or motherhood or the American flag. But it is a fact repeatedly observed by academics of first-class intellect, that second-raters may have mixed feelings toward people with intellectual brilliance and are rather often to be found voting against them or raising various nonintellectual objections against them. I have no statistical data on this point, but my observation over many years convinces me that while merely competent faculty do not usually resent a first-class colleague who is several years older than they, and they may or may not have ambivalent feelings toward

one of their own PhDs who is brilliant (one's narcissism permits an extension there, in thinking of 'my brilliant student so-and-so'), when it comes to faculty their own age or younger and in their own research field, they are frequently ambivalent and sometimes negative in attitude. This should not surprise anybody since envy is a basic and widespread human tendency, and the competition for prestige as well as money in the academy has, especially in recent years, been intense.

Secondly, connected with the first condition but distinguishable from it, the *empowered decision makers have to regard academic excellence as the primary goal of personnel selection and promotion*, to the exclusion of other considerations until it is met. In prediction or selection, industrial psychologists distinguish various kinds of statistical models. The two main ones include a compensatory or regressive model in which different desirable characteristics, qualitative or quantitative, are combined in an additive way with different weights so that sufficient superiority in one is able to compensate for inferiority in another. But another model, known as the conjunctive or successive hurdles model, says that no amount of trait Y can make up for a deficiency in attribute X, which in practice amounts to saying that attribute X must be present if it's qualitative, or if it's a quantitative characteristic must have a score above a certain value known as the cutting score or critical score. With regard to excellence, the decision makers (which include the evaluators discussed under the first condition but also higher administrators such as deans) must hold to a successive hurdles model. If a job candidate or a candidate for tenure or promotion to full professor does not pass that first hurdle of possessing scholarly excellence (ideally evaluated by multiple criteria, some of which can be made quite objective and others are admittedly judgmental in character), nothing else is attended to. One does not inquire into the gender, politics, religion, race, personal style, work habits, or anything else about the candidate until the excellence hurdle has been adequately satisfied.

The third condition for a first-class department is, of course, *money*. By and large, academics are not as strongly money-oriented as other persons in our culture of roughly equal intelligence and education, although I have the impression that there has been some change in that respect in recent years. However, in these times, it is simply not possible for a college professor to ignore an offer from another institution that involves a sizeable salary increase. Most academics have children who will be going to college and even state institution education costs are a heavy burden to bear. So, if Princeton offers a midwest psychologist a $10,000 increase in salary, even supposing that such a person, "a retention case," values the intellectual companionship of his present colleagues, enjoys the Minnesota opportunities for fishing and hunting, likes the quality of life in the Twin Cities, and has friends and relatives locally—he simply cannot afford to ignore the fact that he will ten years hence have earned $100,000 less if he declines the competing offer.

It is obvious that even if an institution is fairly well off as regards dollars, which few state universities are today and few will be in the foreseeable future for a variety of reasons discussed by economists, it is not possible to maintain excellence over the board

in all departments. Whatever one may hold ethically with respect to how much market considerations should prevail—I have never been able to make up my mind about this question—if there is a policy that all departments regardless of their quality should receive roughly the same amount of support, there is the foregone conclusion that the excellent departments will over a period of time undergo a decline in quality.

If these three conditions are met, one is almost certain to achieve academic excellence and to maintain it. I cannot say that they are literally necessary conditions (at least in the short run), because there are some famous examples, at least in my field, of departments in which two or three unknowns happened to spark and develop programs that achieved high visibility rapidly and as a result began to attract other faculty perceiving the place as "on the make." Signs of decline in excellence, because one of these three conditions is no longer met, have a malignant autocatalytic effect, which is easy to explain if one understands how the academy works. Suppose, for example, a department of reputed and real excellence loses a series of retention case episodes not easily explicable on some grounds such as that the department has ceased to put a value upon a certain sub-area and doesn't really mind if faculty in that area leave. Academic gossip is often invidious as well as administratively useful to competing schools. It starts to be said, "You know, Minnesota was a pretty good place, but lately something seems to have gone wrong there, look at what happened in the last five years, they've lost Professor Jones and Professor Smith and Professor Robinson—all stars—because they failed to meet a retention offer." Such rumors, whether well-founded or based upon fragmentary or inaccurate data, will lead to a reluctance on the part of job-seekers to come to the institution, and faculty replacements will have to be from a pool of persons who, while perhaps quite competent, do not possess scholarly excellence. A further consequence is that super-bright graduate students will go elsewhere because their undergraduate advisor has said there seems to be something going wrong at Minnesota and maybe it's not all it's cracked up to be or all it used to be in the good old days. This means a less stimulating intellectual environment for other students, and it also means that faculty who rely upon doctoral candidates as research assistants find their research impaired because they have to rely on less competent assistants. All of these processes unfortunately possess positive feedback to the declining state and are difficult to correct unless one does it very quickly. It is well known that national reputations of departments and of even specialties within departments have a considerable inertia, so that you can coast awhile on your reputation; but the other side of that coin is that once the reputation starts to slip, the slipping has an inertia of its own.

It should not be necessary to state although, given the current climate of frenzied egalitarianism and political correctness, let me say for the record that I am not asserting that scholarly excellence *should* prevail over all other considerations. That is my own preference, my "taste" if you will, but I do not seek to impose it on others who put a higher value upon other considerations. I am inclined to think that if you ask the taxpayer, you would find that excellence comes ahead of goals like affirmative action, at

least when the two conflict, but I do not know this. This entire section is a conditional argument. It says what you have to do if you want to achieve and preserve academic excellence. If someone is not interested in excellence when it conflicts with, say, ethnic diversity or some other cherished moral or social value, then obviously my means-end argument would not be of interest. All I am saying in this section is *if* you value academic excellence and sincerely state it as your primary goal, then these are the three conditions that you will have to meet to attain it.

3. Exams

Social scientists often complain that not enough money, private or public, is available for social science research as compared, say, with research in physics or biology. I have never complained about this—even when I was a department chair—for three reasons. First, one should expect astronomy and physics and parts of biology as now studied to cost a lot more than psychology or sociology for obvious reasons of instrumentation. Even a small cloud chamber on a laboratory table is bound to cost many times more than a mimeographed personality inventory. Secondly, a very large part, in my opinion well over 50%, of research studies in the social sciences are worthless (see Sacred Cow IV). Thirdly, a fact I find more discouraging than either of the other two, even when the social scientist manages to conduct genuinely scientific research which does show adequate replication of a robust result in different populations and using different methods, so that it is possible to formulate an empirical generalization (or, more rarely, a strongly corroborated explanatory theory) with as much confidence as we ordinarily expect in empirical knowledge, there is a tendency for the social group to ignore it. The present section involves a striking and discouraging example of that sociological phenomenon.

The controversy concerning what kind of tests should be used to assess student learning at the college level has gone on now for almost three-quarters of a century, having begun in the early 1920s with the introduction of the "objective" ("multiple choice," "response-selection") type of test. As epidemiologists and insurance companies count, this means that disagreement on this matter has persisted for almost three generations of college professors. Research on the unreliability of grading the usual essay examination antedates World War I, the classic studies being those of Starch and Elliott (e.g., 1912, 1913a,b) showing the pronounced disagreement among teachers in grading examinations in history, English, and—surprisingly, considering the relatively objective subject matter—even plane geometry. Many aspects of the controversy have been researched and the results are as clear-cut and consistent as we can hope to get in psychology (see, e.g., surveys of the research literature[26] by Coffman, 1969, 1971;

[26] [*A great many articles related to this section had been collected and read by Meehl, but not sorted for citation. It was decided simply to let the summarizing literature surveys here serve as sources for research findings that are mentioned but had not yet been formally cited in this section.—LJY*]

Coffman & Kurfman, 1968; Dwyer, 1982; Gronlund, 1988; Hopkins & Stanley, 1981, chapter 8; Mehrens & Lehmann, 1984, chapter 5; Page, 1966, 1985; Page, Tillett, & Ajay, 1989; Tyler, 1934). Nevertheless, one finds that college professors, students, and administrators make dogmatic pronouncements as if no such empirical investigations had ever been conducted, pontificating with assurance and usually relying on cliché.

If a test of scholastic achievement—"What has this student learned about this course's subject matter from reading, lectures, and discussion?"—is largely invalid, does not rank student performance with even moderate accuracy, then it is inefficient from a social engineering assessment standpoint and, more importantly, commits an injustice. Even since the grade inflation of recent years, students still don't end up all having the same grade-point average. A student's job opportunities, the kinds of letters of recommendation that some professor scanning the grade record may write in his behalf, whether he is admitted to graduate school or a professional school, not to mention impact upon self-concept and feeling of personal worth—all are affected, sometimes markedly, by the grade the student got in a course. I need not argue further that, assuming some kind of grading is to be done (something that has, of course, been challenged in recent years at some institutions), it should be done as accurately as our knowledge, techniques, and financial resources permit. *Given what can hinge upon a student's academic record, to employ an inaccurate method of assessment when a more accurate one is available is unethical*.

On the other hand, it is inefficient if we have available two procedures that are of approximately equal accuracy for assessing what has been learned, but one costs a great deal more of faculty or graduate student time to conduct. Arguably it also has unethical features because *anytime you waste money, somebody pays*. Obviously here it is not the professors or teaching assistants that pay, but the students who pay the fees and the taxpayers who support the institution. The grading of essay examinations is commonly considered among the most irksome and tedious tasks of faculty members and teaching assistants. Recently I mentioned to a retired professor in our department how much I was enjoying my retirement and asked how was he doing; before saying anything else on the positive side he said, "It's such a relief not to have to read those damned bluebooks" (called so at our institution because essay exams are commonly written on theme paper bound in blue covers). With regard to teaching assistants, it is at least arguable that the student's grades should depend upon evaluation by the professors who wrote the essay exams and know what they have in mind for acceptable responses rather than by graduate students who often have not even taken the course themselves; and for such graduate students, the irksome and time-consuming task of grading bluebooks often occurs at the same time when they themselves are under pressure to take end-of-term examinations. Less highly visible consequences may also be of great significance to individuals; for instance, perhaps a certain grade is required for Course X as a prerequisite for enrollment in Course Y, and if a student is unreliably evaluated with respect to that first criterion there can be trouble for her in the next course, which in turn means additional time

and worry for faculty and teaching assistants. Thus, a discussion of the strengths and weaknesses of the two kinds of examinations is important for both moral and economic reasons.

We start with an explicit recognition that both the essay and multiple choice exam have their characteristic defects, some of which are curable with sufficient effort (rarely expended, however) and others of which are (so far as human ingenuity for the last seventy years has been able to work out) not curable and probably intrinsic to each type. That being so, the only rational approach is to evaluate the efficacy of cures for the characteristic defects of each, assess the cost of such cures, and then—something that people dislike doing but rationality requires—if that cost seems excessive, to assess the disutility to the student, teacher, and taxpayer of leaving a particular defect uncured. If one sets up the utopian demand that we have to have an exam of a sort that will be perfect, no matter what it costs, that is unreasonable. Operating within constraints of time, money, logistics, and human patience, we have to evaluate, on the best available evidence, the curable and incurable defects of the two kinds of tests.

One should not, as is commonly done, compare the best of one type of exam with the worst of the other and draw conclusions from such an unfair comparison. It is rational first to investigate the comparative merits of the essay and multiple choice test when each has been constructed with extreme care, even if that degree of care is not usually achievable in practice because people aren't sufficiently motivated to exercise it. It is also, unfortunately, desirable to ascertain how the two compare when they are constructed in a relatively sloppy manner, as is commonly done for each. Throughout one must keep in mind the question, "If well-informed, what will faculty actually in practice be willing to do?" We should not spend large amounts of money and irk large numbers of persons for the sake of small differences in efficiency or even small differences in justice. We take for granted that any fallible procedure, even in courts of criminal law, will be bound to result in individual injustices. Assessment of student performance should operate on the lawyer's *de minimis* principle.

The first important clarifying observation is that the defects of the usual carelessly built multiple choice examination are obvious to any thoughtful person; whereas the defects of the ordinary typical essay examination are not obvious to most people, and are not appreciated without some technical knowledge of psychometrics and acquaintance with the empirical research studies. This asymmetry in how much you have to know and how sophisticated you are about problems of psychological assessment is what leads to the widespread opinion among faculty that the essay exam is unquestionably superior to the multiple choice format, and to a similar feeling on the part of many students. (However, when asked which type of exam students *prefer*, the studies are not consistent. There is a tendency for students to say they prefer the multiple choice tests even when they complain that multiple choice tests are no good as a measure of what you know. That is not an inconsistency; it is quite possible for me to prefer being examined by one form even though I think the other is superior for some reason.)

Let me first consider the alleged and actual deficiencies of the multiple choice test. (For a guide to construction of good multiple choice tests, see Sechrest, Kilstrom, & Bootsin [1993] and additional comments on that article by Logan [1993].) The most common complaint made by faculty who resist their use is that multiple choice tests merely assess the student's ability to regurgitate facts and that they can't require the student to "think." This cliché is false when formulated as an incurable defect, although it is sometimes true when a multiple choice test is carelessly built. As an undergraduate I took the national examination in physics because our physics department was cooperating with the Educational Testing Service in norming it. While I received the top score out of some 250 students in the class, I can confidently assert that it was harder, forced me to think more, than almost any examination, multiple choice or essay, that I took in seven years of higher education. There is no doubt in my mind that it forced me to think harder than the essay format examinations of the PhD prelims in psychology. This is not difficult to achieve in any subject matter which is sufficiently developed so that it is possible to say what is the right and the wrong answer to a complex question. A problem is stated, and the student must set up the equations based upon his conceptual understanding of the components of the physical problem, and carry out the required calculations to get an answer. He then scans the six alternatives offered in the multiple choice format to see whether one of them agrees with his answer. The mental processes required are of the same kind as in an essay exam, but the outcome of those mental processes is then to be matched with an answer permitting a completely objective, or "clerical" scoring of the test. All of the answers are plausible; if you haven't correctly worked the problem, you will not be able to get more than one in six right on the average, so your expected "chance" score on a hundred-item test will be around sixteen items correct. We need not be dealing with an exact science like physics to construct an examination that requires some thinking, although admittedly it is more difficult to do in a less scientifically advanced domain. Clinical psychology can hardly be counted among the advanced sciences, but when I taught the introductory course in that for twenty-five years at Minnesota, I had constructed a multiple choice test that all of the students that I ever talked to agreed made them think hard. Students who took the exam without having taken the course averaged close to chance scores. Construction of a good multiple choice exam takes work, but it can be done.

Furthermore, one should beware of the cliché about "regurgitation." That's one of those remarks that both professors and students like to utter without having thought very hard about it. Knowledge of a subject matter consists of a variety of elements, including facts, formulas, generalizations such as scientific laws, tactics of problem solution, definitions of terms, names of important contributors, and so forth. There is nothing wrong with asking a student who has taken a course in algebra to identify in a multiple choice test the quadratic algorithm. Admittedly, we would prefer that the student be able to produce it rather than select it from a set of alternatives, and, in this instance, it is possible to use a "short answer" format which is neither multiple choice nor essay. But in

a test of abnormal psychology it is quite possible to make multiple choice items that assess which symptoms of various mental disorders students have learned and whether they recall that one of them gets somewhat more diagnostic weight than another, and so on. *Multiple choice examinations which do not require thought are not intrinsic to the format but are the product of professorial laziness*. So while a carelessly constructed multiple choice test often has the weakness that it does not force the student to think, but only to remember, it is wrong to count this as an intrinsic ineradicable defect of the format.

A second defect is that because the format only requires the student to *select* a response rather than to *produce* one, the test cannot tap the student's ability to organize and exposit thoughts. This is an intrinsic defect, since obviously I cannot sample expository ability if I ask somebody simply to choose rather than expound. We must chalk this up as an intrinsic and ineradicable deficiency of the multiple choice exam. However, even that defect must be very carefully defined if it is to be in accord with the research evidence on the subject. We must distinguish between an appraisal of a student's expository talent which proceeds by requiring exposition, and an appraisal of that same expository talent done by indirection. The distinction here, about which non-psychologists are usually not informed and without which one cannot properly evaluate psychometric questions, is the distinction between test items as *signs* or as *samples*.

If we assess a student's spelling ability by taking a random sample of words from the dictionary and counting how many can be spelled correctly, the test items are *samples* of the domain of our defined intrinsic interest. That is, we are literally sampling the very behavior we wish to appraise. On the other hand, one may employ a psychophysiological test to detect tendency to schizophrenia, in which case the test is not a sample (since the defining properties of schizophrenia as a mental disorder are molar level psychological, such as affect, thought-disorder, delusions, and hallucinations); the neurological test is functioning as a *sign*, rather than as a sample. Similarly, if we administer an intelligence test to decide whether a person should be legally classified as mentally deficient, we have a variety of vocational, educational, and social capabilities and disabilities in mind when we administer the test. One of the most highly valid subtests of the Wechsler Adult Intelligence Scale is the block design test, which samples how well the subject can arrange colored blocks to match the pattern of a presented design. This is not a task which the person will be required to perform in any of the life domains that we're concerned with clinically or administratively, but the fact is that it is well-saturated with the general intelligence factor *g*. (In my psychology internship I administered only the block design to patients who spoke only Finnish, not knowing any Finnish myself.) If it were for some reason decided to appraise a geography student's expository skills rather than technical knowledge in geography, however clumsily and unattractively presented, it turns out that it is possible to assess expository talent with a multiple choice test carefully constructed for that very purpose. So, one must ask, "Do I care about the expository skill in this particular class?" It is obvious that in some courses you would care about that; and

in others, a little bit but not much; and in still others, not at all. If I want to assess psychology students' knowledge of learning theory or psychometrics or abnormal psychology, I prefer that assessment not be contaminated by the expository skill dimension. But if we are dealing with a law student in a course on appellate pleading, expository skill is an essential feature of what we're trying to appraise in her behavior repertoire. This component of expository skill which, if we wish to assess it, may count as a defect in the multiple choice test, has another side to it in that it counts as a possible defect also in the essay test as a source of bias in the grading, which will be discussed below.

What we are left with is that a common defect of multiple choice tests requiring "mere unthinking regurgitation of a fact" is a consequence of lazy and careless construction and not intrinsic to the format. The question of assessing expository skills has to be parsed for (1) whether you are intrinsically interested in appraising it, and (2) whether you care that you are requiring exposition in the course of appraising it. A multiple choice exam can appraise students' levels of expository skill (test as "sign") without requiring exposition to be performed (test as "sample").

The essay examination as ordinarily conducted has a very grave defect which, despite all of the research demonstrating it, seems to be one that most faculty are blissfully unaware of. That defect is unreliability of grading.[27] The research on this subject is not a matter of dozens, or scores, but hundreds of research studies going back to Starch and Elliott. The studies are so large in number and varied in subject matter that educational psychologists writing summary papers or book chapters have often in recent years urged a cessation of such research studies because, unless some new idea appears, there is no point in repeatedly researching something that has now become blindingly clear to anyone acquainted with the investigations. The point here is very simple: a typical professor outside of the field of psychology who is unacquainted with psychometrics and with the research on the subject of achievement testing takes it for granted (as we all tend to do when we do anything in our work) that she can do it competently, which would include reliably. The research shows that two graders who have scholarly competence in a subject matter do not show satisfactory agreement with one another if they independently grade the same set of examinations. Even the same person grading the same set of examinations weeks or months later shows an unsatisfactory agreement with her- or himself. I can't help it if that makes the reader uncomfortable, that's the way the facts come out. There is no excuse for people interested in this topic to proceed without even inquiring into that question of grader reliability. The reliability coefficients vary widely over the studies from horribly low values in the .20s to maximum values in the .80s, with a representative value of the reliability coefficients being somewhere around .50–.60. The ordinary global impressionistic grading of essay exams will not achieve reliabilities in the .80s or usually

[27] Edgeworth (1888, 1890) published probably the earliest papers on grading essay exams (although Galton would be a possible predecessor), mostly discussing the theory of error (with some real data from "marks" on exams for Civil Service, Cambridge Honors, etc.) and the lack of "reliability" in grading.

even in the .70s with most graders. Now, a grading reliability of .60 cannot be considered adequate for appraising the performance of individuals, and employing it is a working of injustice. A interscorer reliability of .60 means that only 36% of the variation in scores achieved by the students is based upon properties of their performance, and that two-thirds of the variation is a matter of "chance," that you could take as equivalent to flipping pennies or throwing dice.

We usually do not consider it appropriate to employ a psychological test for assessing individuals (as contrasted with comparing groups) unless the reliability approaches $r = .90$, although we often are forced to settle for somewhat less than that. A good individually administered intelligence test or a carefully constructed group test of achievement (such as spelling or arithmetic) will have a reliability approaching .90. The trouble here is that people unfamiliar with the research literature simply take it for granted that *they*—being honest, conscientious people—grade examinations reliably. If they are told research studies show that essay exams as normally graded are graded unreliably—that they would not agree satisfactorily with another independent grader familiar with the course material, and as a matter of fact in all likelihood would not agree very well with their own previous grading—they simply don't accept that it applies to them. This is an example of a universal human failing: assuming that you can do something better than you can do it; and if confronted with the fact that by and large people can't do it well, to consider oneself to be one of the rare exceptions who can. I know something about this because I have used both multiple choice and essay examinations in various courses over my years as a professor. On the two occasions that I investigated the reliability of my own essay grading, on one occasion I did fairly well, and in the other instance—when I was initially equally confident—I did very poorly. The plain fact of the matter is that we cannot judge our own performance reliability on the basis of our subjective feelings of confidence. I don't expect non-psychologists to accept these generalizations readily, but I am astonished when I find that many psychologists and other social scientists show the same reluctance to accept the well-established findings, or accept them but take it for granted that they themselves are an exception. I do not dispute that there are individuals who can score reliably, and pairs of individuals who will agree well. That doesn't do us any good with the problem of efficiency or the problem of justice unless we have taken the necessary steps to calibrate the particular individuals concerned and find out who constitutes an exception to the usual rule that global grading of essay exams is unreliable.

Let me emphasize that what I'm explaining here has nothing to do with any particular theory of the mind, or of human abilities, or any controversial problems in classical psychometric theory, nor does it have anything to do with questions of educational philosophy or values. It is simply a matter of high school algebra that the reliability coefficient imposes an upper bound on validity. More precisely, the validity of a test cannot exceed the square root of its reliability coefficient, so that it does not matter how beautifully and precisely the task samples the achievement of the student, that if God

were grading the essay exam it would come out perfect in appraising performance. Whatever the behavior product is as a sample of what the student can do—even if it's a flawless indicator and ranks students in perfect rank order of how they would be ranked if we had Omniscient Jones' knowledge of the innards of their minds—if the behavior product cannot be reliably judged, that intrinsic validity does us no good, and our judging procedure commits injustice.

Fortunately, this grave defect of the essay exam is curable. Just as in the case of the multiple choice test, when if you take the examining task seriously as an ethical problem of just appraisal you can avoid the charge of "mere atomistic regurgitation," so with the essay examination format it is possible to conduct it in such a way that grader reliability approaches the satisfactory .90 value; but, as with the multiple choice test, you have to work at it. Let me remind the reader again, it's no good comparing a lazy use of one type of test with a hard-working, well-informed examiner's use of the other kind. If you want to make comparisons properly, it is necessary to compare the two kinds, each of which is being approached by someone sophisticated about test construction and motivated to do an adequate job of it.

One method of achieving satisfactory reliability is to increase the number of graders. The Spearman-Brown Prophecy Formula, originally developed to predict the improvement in reliability of mental tests achieved by lengthening them, applies to a wide variety of human judgments in which the individual judge's response is counted as if it were a test item. For example, if the grading reliability of an essay test on a certain subject matter by a pair of subject-qualified professors is around the typical value of $r = .60$, the Spearman-Brown formula tells us that we could achieve a pooled rating reliability $r = .91$ by having the essay exams read independently by seven professors. I need hardly say that, while such investigations have been conducted for research purposes, this is not a solution that is very appealing in our present context of trying to save money. I suspect that you would not be able to find 1% of departments (even of psychology, who are aware of these problems) where the faculty would be willing to spend this much time grading each other's exams. Even in a very large department, for most subject matters it would be hard to find seven faculty members who would be competent to serve as readers. In my own department, we have almost never managed to get more than four independent readers of our PhD written preliminary examinations and on many of them we have only two. Nevertheless, we keep in mind that the achievability of satisfactory reliability by increasing the number of readers is part of the available toolkit.

A more realistic approach is to use *analytic grading*, in which the grader is provided with a carefully constructed list of items that are checked off if they appear in the essay with tolerable clarity. These items can be qualitatively of any sort, ranging from a simple atomistic fact (e.g., the Armada was sunk in 1588) to a more complex content (e.g., does the student correctly formulate Kant's categorical imperative in each of its three forms?). If desired this checklist can be supplemented by a global evaluation of the sort that is ordinarily the sole method of appraisal of an essay exam. The student's grade is then

obtained by summing the items, perhaps assigning greater numerical weight to some than others (although in that case it is desirable that the grader be uninformed of these weights), and if a global evaluation is included adding that as well, to obtain a total score. In the military higher educational institutions (e.g., Army Command and General Staff College, the War College) I believe this approach is known by a different phrase, the *school solution*. The phrase refers to the fact that it's the solution to a military dilemma or decision problem that the faculty of the school consider acceptable, including the reasons for and against the various alternatives that should have occurred to a commander in a described situation. The advanced student in one of these institutions is presented with a situation where he is, say, commanding a brigade in a certain position, with such-and-such weather conditions, such-and-such information from air reconnaissance, statements extracted from prisoners of war, and orders to do X or Y depending on his appraisal of the situation, but not to do Z without further communication with the higher command, and so on. The faculty that concocted this problem usually agree on the "right answer," but they may also have a minority view so that some credit is given to a non-preferred answer, and zero credit is given for a "bad answer." But the examinee is also supposed to mention the various considerations taken into account in arriving at the best action.

I can illustrate the school solution approach by an example from personal experience which also provides a suitable horror story about low reliability of the customary global grading. When I served on the American Board of Professional Psychology in the 1950s, I was the first "non-grandfather" to serve in that capacity and therefore had myself taken the Board's examinations, written and oral. I took a dim view of the exam and, aided and abetted by a couple of Ohio State PhDs (where there was a strong tradition of emphasizing reliability questions), insisted that one year every essay exam on research should be graded by two independent, randomly assigned graders. It turned out, to our horror (but not to my surprise, having taken my doctorate at Minnesota), that this essay had an inter-grader reliability coefficient of .21. I ask the reader to reflect on what this means. These examinees are psychologists applying to be diplomates in a specialty (such as clinical or counseling or industrial psychology, analogous to the specialties of medicine such as obstetrics, internal medicine, and pediatrics) who have put their egos on the line, spent several months studying, and paid a nonrefundable examination fee. Each has a PhD from an accredited institution and five years of practical experience with patients or clients, some of which time may be doing research but most has to be doing treatment or giving advice to business firms post-doctorally. The essay exam was on research design and interpretation, which one would expect to be one of the more "objective" kinds of topics to write about. What does that reliability coefficient of .21 mean? It means that how the candidate is graded depends about 4% on the characteristics of his or her behavior product—the essay—and 96% on who happens by chance to be the grader. It is remarkable that a board consisting of psychologists would have been conducting this sort of exam for several years until three of us made a fuss about it. Had examinees who failed been aware of this miserable grading reliability, they might have had a pretty good

legal case against the Board for damages. Since, having taken this exam myself, I was the most vociferous in complaining, I suffered the usual fate of being invited to devise a school solution examination on research. It didn't take as much work as I had feared, and the inter-grader reliability coefficient on my school solution exam reached the satisfactory value, $r = .86$, only a little less than the equivalent-forms reliability of a good individual intelligence test, like the Stanford-Binet. Moral: if one has reasons for preferring the essay format, it is possible to eliminate its greatest (and, as usually administered, well-nigh universal) grave defect of inadequate scoring reliability.

A second defect—unfortunately, incurable by any amount of work and conscientiousness—is *domain sampling* and it is also an aspect of what the psychometrician calls reliability. There are technical problems both conceptually and mathematically concerning the traditional distinction in psychometrics between the reliability of a test (roughly, how accurately it measures whatever it does measure) and its validity (how accurately it measures what it purports to measure). I want to avoid those controversies, which are not directly germane here, and simply say that from the standpoint of tests as samples, if, as is almost always the case, there is considerable qualitative heterogeneity in the subject matter of a college course (i.e., different sorts of items of information and skill ranging over a variety of subdomains), it is impossible to sample the domain adequately with a small number of essay items. Since the chief impetus for adopting the essay format lies in its requiring the student to organize thoughts and produce a logical, clear, effective exposition about some topic, that qualitative aim is defeated if the time available to write such an organized essay is less than a half-hour. (Some would say that is too short a time to get the distinctive advantages of the essay format.) A final examination in a college course almost never has more than two hours available, and if we tried to extend that time enough to deal with the present problem, things would get out of hand (e.g., we would be in trouble with the administration as regards logistics, and in trouble with students for obvious reasons). That means the examination can only require four essays. In a three- or five-credit course in medieval history, there would be scores of topics of importance that were treated in some detail in the textbook or in lectures, and from which these four essay topics could be gleaned. Whatever four topics we select, there is a very large element of chance involved from student to student, and again an inadequate sampling of the domain of knowledge, hence many instances of injustice. The professor may have an essay question on the investiture controversy between the Holy Roman Emperor Henry IV and Pope Gregory VII, and student Jones, being a Roman Catholic with more than the average interest in the history of his religion, finds this fascinating and studies it intensely. Student Pappas, who is Greek Orthodox, doesn't care about the pope, but she spent a lot of time on the schism between the Eastern and Western church and the life and personality of the patriarch of Constantinople, Michael Cerularius. All sorts of differences—in interests, attitude, what other academic pressures there were at the time a certain topic was reached in class or in the book, and so on—contribute to the near-certainty of a modest or low correlation between how much students know about topic A,

topic B, topic C, and so forth. It's as if we used a spelling achievement test in which instead of sampling one hundred words from the *Oxford English Dictionary*, we sampled only four. The chance factor affecting which students will do well and badly in this instance is very large. One partial solution which is often used in PhD prelim examinations is to present more essay questions than need to be answered and let students pick the ones they prefer to answer. This improves matters but it doesn't repair them, partly because there is research to show that students are not as accurate as one might hope in knowing which of the list they will do better at. The only conclusion to draw about the domain sampling problem is that it is incurable given the limitations of testing time. No amount of ingenuity will get rid of the fact that you cannot sample a domain of dozens of topics that vary a great deal in their content and interest by essays on only four of them. Whether it is possible by great effort and ingenuity to construct some sort of integrated essay questions that will only be answerable if the student is proficient in all of the subtopics I do not know, and I have not seen any research showing whether that can be achieved. For most subject matters, I am confident it can't be done.

Another defect of essay exams is the influence of irrelevant biasing factors on the grading. For example, there is a correlation between grades assigned to essay exams and the clearness or aesthetic merits of handwriting, even though examiners don't approve of taking that into account and may have even been instructed specifically to avoid it (Marshall & Powers, 1969). Furthermore, there is evidence to show that those graders whose own handwriting is attractive unconsciously give greater weight to the examinee's handwriting than do graders whose own handwriting is not aesthetic. Most advocates of the essay examination are insufficiently aware of the extent to which irrelevant features of style and the ability to do a verbally fluent scholarly-sounding "snow job" can fool the grader.

One interesting line of research has shown that a simple composite of a limited set of essay text features *scorable by computer* (e.g., range of vocabulary and sentence length) correlates as well with human judges as the judges do with one another (Page, 1966, 1985; [*for a bibliography on this topic, see Haswell, 2008*]). While this finding might solve some of the problems covered in this section, it also raises the question whether judges who think they are responding to conceptual depth and organizational ability are responding more to "cosmetic" features of the verbal production than to what they think they're paying attention to. Of course in many subject matters outside of science, the position that the student adopts on a controversial matter is subject to impact in terms of the bias of the grader. I have myself seen graded essay examinations in some of the "softer" areas of the liberal arts, with marginal comments in which the personal social, economic, religious, or other prejudices of the teaching assistant who did the grading were quite obviously present. It is likely on the available evidence that some of the features of the essay format that are commonly admired are themselves inextricably entwined with such adverse properties as verbal fluency, buttering up the readers' prejudices, and aspects of verbal performance that are essentially nothing but indirect

reflections of the general intelligence factor. We cannot object to the empirical fact that an achievement test reflects intelligence, which of course it is bound to do if it is measuring anything worthwhile. But there is a difference between a performance that is good on its merits partly because of the student's intelligence (along with work habits, motivation, and conscientiousness) and a feature of the verbal output which is simply another indirect way of getting at the intelligence component itself rather than a measure of what has been learned.

An extreme, disconcerting example of the impact that rater unreliability and domain sampling can have on achievement appraisal is the assessment of writing skill. Accepting multiple and large sample work products as the best available criterion of writing ability (content knowledge aside), it turns out that an "indirect" measure, a multiple-choice "Test of Standard Written English," predicts the criterion about as well as a small writing sample (Breland & Gaynor, 1979; Breland, 1977). This counter-intuitive finding, that a sign does as well as an actual sample of the behavior domain we aim to assess, arises from unreliability factors. I am not advocating such an indirect approach, which does violence to the intuitions even of psychologists, mine included. I mention it only to highlight the unreliability issue. In knowledge domains where we consider skillful exposition as intrinsic to competence (e.g., a legal brief, a philosophical discussion of Kant's ethical theory—unlike knowledge of trigonometry, neuroanatomy, or psychometrics), there is no need to assess the writing skill separately from the content.

What do these clear-cut, massive, consistent research results, going way back to the 1920s, lead to as a practical conclusion? My conclusion is that unless some new evidence is forthcoming (which I see no signs of happening or anybody attempting; after this many years, all of the plausible excuses have been looked into pretty well), the burden of proof now rests on the professor who chooses the essay format. If considerations of justice and cost were in conflict, then we would have a mixed moral, economic, and political problem to worry about. But in this instance they are not in conflict. If I as a professor do a typical, easy-going, slap-dash job of building a multiple choice test, it will have a satisfactory reliability but will not assess the "higher" cognitive functions in which I may be interested. However, if I do an equally lazy slap-dash job of making and grading an essay exam, it will have such an unsatisfactory scoring reliability, combined with the unavoidable domain sampling unreliability, that I cannot accurately assess *either* the lower or higher cognitive functions in which I am interested. That I have a pleasant feeling that I am doing those things when I grade the essay exam simply shows the powers of human self-deception and the ability to deny research facts that don't fit one's prejudices.

On the other hand, if I am prepared to build a reliably gradable essay exam, I will have to put in a great deal of time and cannot leave this task to my teaching assistant. Then the question arises, if I put in an equal amount of high-quality time in building items for a multiple choice test, how well will it do? The answer is that it will do about as well as the essay does in tapping the higher functions and will be superior in domain

coverage given equal grading reliability. I do not know any way to escape this line of argument and have not seen anybody attempt to do so who was informed as to the research data. If I construct a globally scored essay exam, I should have some valid reason (not just my armchair conjecture or my cozy feelings, but some kind of hard data calibrating myself or my teaching assistant if she does the grading) showing that we have somehow managed to concoct essay questions that require integration of the subdomains the course has covered, and that we have reliable grading. But in that case the burden, which is no longer one of proving something about accuracy and justice, is still present to show that I am not wasting the student's fees and the taxpayer's dollar by doing something in one way that I do could do as well, or almost as well, in another way. I am unaware of any good research evidence, or even any convincing anecdotal "evidence" from myself or others, to show that the imperfect validity due to the limitations of domain coverage is counteracted by a material increase in the net attenuated validity of tapping the "higher cognitive functions" that everybody talks about. Furthermore, I am unaware of any evidence or armchair arguments to show that, in large undergraduate classes on the introductory level of a subject matter, there *are* any higher cognitive functions typically taught or acquired, and this is important because it's in the large classes that the use of the essay exam is so financially expensive.

There is a final point here that one does not expect anybody but psychologists to be familiar with, but it is important in practice. A large pool of multiple choice items can be constructed (by sampling randomly from lectures and textbooks) and given a kind of "validation" by statistical analysis of each item's performance in relation to total score. If the material to be learned is highly heterogeneous, this approach can be improved by first classifying the item content by subdomains (say, in a course in medieval history we get a subdomain about the history of the papacy, or the development of water transportation), and it doesn't take very many items in the subdomain to serve as a basis for studying the statistical behavior of single items. There is a well-known story about Ohio State University's psychology department that I have been unable to verify, but (since it is consistent with what I know of the professor about whom it is told) I am inclined to think it is probably substantially accurate. Professor Herbert A. Toops, one of the "greats" in the early development of psychometrics, constructed a pool of several thousand multiple choice items for an undergraduate psychology course. These items had all been carefully built, revised, and internally validated in the usual way. Furthermore, the content was stratified so the number of items representing a subdomain of the course material was roughly proportional to how much lecture time or textbook pages it occupied. Every time the course was taught, a hundred-and-fifty-item test was constructed by randomly sampling from this huge pool. The entire pool of three or four thousand items was on file in the psychology department and in the university library, and students were told that they could go and study these items if they wanted to prepare for the exam, specifically that they could do that instead of attending the lectures. This horrified some people, but Toops's position was simple; namely, that if the student learned the content of that item

pool, so that she did well when presented with a sample of it on the final exam, she had acquired the content of beginning psychology.

I had a similar experience when I took the neuroanatomy course in the medical school, competing with second-year medical students thoroughly familiar with studying that kind of more-or-less meaningless rote-memory material. A medical student friend warned me that the practical part of the final exam, which involved identifying various parts of the central nervous system on slides or dissected preparations, would be difficult for me because I was competing with medical students who had a master set of flash cards (based upon previous examinations) in the medical fraternity house. He kindly lent me his set, which I copied and memorized (in studying the brain, at least in 1942, there was nothing to do but memorize, since none of it made any sense to speak of). I managed to get the A grade expected of the clinical psychology students taking that medical school course. A couple of years later, I asked the young neurologist, who had been studying for a PhD and was the laboratory assistant in that class, whether he was aware of the existence of this fraternity item set. He pointed out that of course he was, since he had taken his M.D. at Minnesota and had been in that fraternity. He said it didn't bother him or Professor Rasmussen because the final exam consisted of a randomly chosen set of some fifty slides or preparations from a larger set of five times that many, and if a medical student learned to identify the whole batch, he had mastered the content of a neuroanatomy course. I do not think there is any good criticism of that point of view.

A final comment which bears on the character of the academy. Almost every argument—I think literally all of the major arguments about this issue—were thrashed out in the 1920s and presented forcibly in a book by psychologist G. M. Ruch, *The Objective or New-Type Examination*, in 1929. Even so, seven decades later, one constantly hears the same old arguments presented, not only by professors of history or philosophy, but by psychologists. One is impelled to quote André Gide to the effect that you have to keep saying certain things over and over again, because nobody listens.

4. Rating Teachers

The problem of assessing the efficacy of lecture format classroom instruction at the college level is still not clearly solved despite quantitative research on teacher rating scales going back to the middle 1920s. I conducted some research on this as a graduate student and have retained an interest in it, as well as periodically having myself rated by the senior college class I taught in introductory clinical psychology. However, I have not kept up on the recent research literature and can claim no technical expertise in the subject, so what I have to say here is based partly on armchair considerations and largely on Marsh's (1987) definitive summary of the research literature. I am indebted to my colleague P. W. Fox for alerting me to useful recent references and for helpful conversations about what conclusions can be drawn from them. On the basis of these sources I am

inclined to be somewhat more optimistic than he is. This is not the place to argue that (mild) disagreement between us, especially since he possesses the expertise and I do not.

There are some fairly consistent research findings which I believe are essentially not in dispute. I note that by and large they confirm what was believed in 1941, when I wrote a thesis on the topic for the BA degree *summa cum laude*. (Even in the social sciences, *some* empirical regularities are robust and not subject to marked social change.) They suffice to correct certain widespread misconceptions held by professors who are, for one reason or another, strongly opposed to student ratings of college teachers. First, the inter-rater pairwise reliabilities are in most studies found to be lower than we get in other human judgments of other people's traits (e.g., intelligence, shyness, dominance, or conscientiousness). Pairwise reliability coefficients tend to run in the low .20s, a discouraging value attributable partly to "unreliability" in a narrow sense (e.g., as in judging lifted weights), but chiefly to students' differing trait *meanings* and different *baseline anchorings*, in turn reflecting different values and expectations. Fortunately, student ratings, like most other kinds in the clinic, industry, and the military, have been shown to follow, not exactly but well enough for practical purposes, the Spearman-Brown Prophecy Formula, which predicts the reliability of N pooled judgments as a function of N. For one of the best-built scales, Marsh's SEEQ [Students' Evaluation of Educational Quality], the pooled rating reliabilities of its five factors are .23 with one student, .60 for a class of five students, .74 for 10, .90 for 25, and .95 for 50 students. As Remmers concluded over a half-century ago, "given a sufficient number of students, the reliability of class-average student ratings compares favorably with that of the best objective tests" (Marsh, 1987, p. 275). Of course, 'reliability' as the psychologist uses the term is not the same as validity, but in rating situations it at least tells us that *there is something about the professor* that students are discriminating better than mere chance.

It remains to be seen from other relations within the data what that "something" that they are reliably discriminating *is*. Opponents of student ratings often say things like, "Oh, well, it all depends on how good a ham actor you are, whether you make jokes, or what kind of clothes you wear." Data going back to the late 1920s, including my own research in the early 1940s, suffice to refute this sort of general statement. Many teacher rating scales include a rating on sense of humor, and many include a rating on "personal appearance" or something similarly labeled. The statistical fact is that student overall "global" rating on teaching effectiveness (or a composite of all of the variables weighted or unweighted in terms of the statistical structure of the system) does not show high correlations with either of these "irrelevant" traits. When one examines the pattern of relationships produced by student ratings, students are able to differentiate aspects of the professor's behavior and appearance and, by and large, the traits that come out with high correlations with an overall global rating of effectiveness (or a composite score, or a first big statistical factor) are the ones we would hope students are interested in when they make such an overall judgment. They are items like: "How well does she explain the subject matter?" "How intellectually stimulating is his presentation?" "How well

prepared does she seem for class?" "How well does he seem to know the subject matter?" A teacher who is rated as a clear, well-prepared, intellectually exciting expositor will receive good overall ratings and a good composite rating even if he has a low rating on sense of humor or dresses sloppily.

So the usual armchair complaints are properly answered by stating the statistical facts of what correlates with what. Another sort of objection is not a matter of factual inaccuracy but of internal inconsistency, namely, various ways of saying that one cannot trust the judgments of students about a matter of this kind, that they simply don't know enough, or they don't have the right values, or don't make the ratings conscientiously enough to be taken seriously. That students do not respond in a completely whimsical or random way is, of course, refuted by the reliability coefficients, and some of the "subjective" elements (e.g., individual differences in what students care about, or such things as whether they are taking a given class under duress or don't like classes that start early in the morning) are, of course, partly the source of imperfect pairwise reliability coefficients, about which nothing can be done, except as we can rely upon boosting them via the Spearman-Brown Formula.

The inconsistency component can be teased out by asking objectors to student ratings whether we should be paying any attention to the impact that professors have on students in a class. Is an objector prepared to give teaching ability *literally zero weight* in hiring, promoting, tenuring, or meeting a retention case? Despite having heard many faculty say they don't believe in student ratings, don't trust them, and are opposed to any requirement of them, I never find anyone who is willing to say, when pressed, literally, "I don't think we should pay any attention whatsoever to classroom-lecture performance, *however assessed*." When discussing the value of a faculty member to the department, in practice one of the topics that is almost invariably brought up is student reactions. When one asks an advocate or opponent of promoting Professor Jones where he gets his notion of student impact, he regales one with anecdotes. When pressed further, it comes down to what one or two students said to him about Professor Jones. Now, this is a gross inconsistency for anyone, and egregious in a social scientist who is supposed to know about random and systematic sampling errors. It makes no sense to say, "I think Jones must be a pretty good teacher because a student once told me that at the drinking fountain or in the coffee room," and then to say, "Don't tell me what the pooled judgments of the whole class are on this matter because I don't have faith in students' judgments." Such a position amounts to, "I only listen to students' judgments when they are anecdotal and based upon a small sample—and one quite possibly biased as to representativeness—but I don't care about getting an opinion from everybody in the class giving anonymous ratings." It is hard for me as a psychologist to conceive that the average of three students (selected, who knows how, by their motivation) telling me about another professor, not anonymously, should be more trustworthy than finding out what all of the class members have to say about the same trait. Furthermore, anybody who objects to an overall judgment as being unduly infected with specific traits that he thinks students over-attend

to (such as appearance or sense of humor or personal warmth, instead of, say, scholarly competence) should in consistency attach a *greater* importance to collecting *total class judgments* on a *set of specific traits* rather than anecdotal evidence on a global rating by a small number of students.

If one agrees that classroom teaching is something that we ought to pay some attention to, the next question is how it is to be assessed. Faculty visits is one way and student ratings is another. I have no basis for comparing the "validity" of these two kinds of judgments, and perhaps a department should use both, as is done in our law school at Minnesota. Unfortunately, the reliabilities of single faculty visitor ratings are as poor as single students or worse, and we can't arrange to get 25 or 50 visitors so we can apply the Spearman-Brown Formula. Agreement is very poor between student, faculty, and self-evaluations.

An important question concerning validity is the correlation between traits as rated and student achievement as measured by what is learned. A comprehensive, scrupulously fair, and methodologically sophisticated research review is that of Marsh (1987). Many studies are so gravely flawed as to be useless, yet are widely cited by opponents of teacher effectiveness ratings as demonstrating invalidity. For example, there is no point in correlating *individual* students' global ratings of the same instructor with their grade expectations for that class. Another clear finding is that the kinds of "bias" commonly alleged are only feebly discernible in the research data, or not at all (see Marsh, 1987, p. 328). Yet, somewhat surprisingly, a national (USA) survey of college faculty reveals that 75% of them favor using student ratings as evidence in tenure decisions. I cannot do better by way of summary of this vast literature, going back to the classic Purdue studies of the 1920s, than does Marsh (1987):

> Research described in this article demonstrates that student ratings are clearly multidimensional, quite reliable, reasonably valid, relatively uncontaminated by many variables often seen as sources of potential bias, and are seen to be useful by students, faculty, and administrators. However, the same findings also demonstrate that student ratings may have some halo effect, have at least some unreliability, have only modest agreement with some criteria of effective teaching, are probably affected by some potential sources of bias, and are viewed with some skepticism by faculty as a basis for personnel decisions. It should be noted that this level of uncertainty probably also exists in every area of applied psychology and for all personnel evaluation systems. Nevertheless, the reported results clearly demonstrate that a considerable amount of useful information can be obtained from student ratings; useful for feedback to faculty, useful for personnel decisions, useful to students in the selection of courses, and useful for the study of teaching. Probably, students' evaluations of teaching effectiveness are the most thoroughly studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research. (p. 369)

Studies of the relation between student ratings and average class achievement scores vary widely, partly because of vitiating methodological defects. An adequate study of this

important correlation must compare sections of the same course material, employ reliable examinations, and so forth. P. A. Cohen's (1981) meta-analysis of all known multisection validity studies across 68 multisection courses showed average student achievement correlations from .43 to .50 with five major rated variables, averaging .47. Omitting the rating on "Overall Instructor Ability" as largely redundant, suppose we have four predictors each correlating .45 with the achievement criterion and (bad case, highly unlikely) all 6 pairwise correlations equally large at $r = .45$. Assuming (safely) the 4-trait composite to be approximately normally distributed, the multiple correlation $R \approx .59$ would mean that 98% of students in the low-rated teacher's class fall below the median of those in the high-rated teacher's class. Approximately 83% of students taught by a bottom decile instructor will do worse than the median of all students. These are very respectable values, superior to most predictive validities relied on in clinical, counseling, industrial, and military psychology. Readers who find them hard to believe should also read Gessner (1973, somehow not cited in Marsh's 215 references), where medical student achievement was measured by a high quality standardized national examination and sophisticated statistics were used, yielding rating/achievement correlations of .77 and .69. Another carefully conducted study (not cited by Marsh) showing good student achievement correlations and qualitatively illuminating internal relations is Sullivan and Skanes (1974; see also Sullivan, 1985). I should warn readers that in this kind of "messy" research domain, I tend to have more faith in positive outcomes (absent methodological defects) than in negative, a mental set that some competent but more skeptical colleagues consider unsound.

Uses of Teacher Rating Scales

Given the sizeable body of mixed but (when properly analyzed) broadly encouraging research evidence, if one is not adamantly prejudiced against collecting *any* information about student appraisal of college teachers, one can see reasonably defensible uses of such information. Proceeding from the least threatening, *a first plausible use is by instructors to improve their teaching performance*, a "privately available" procedure that can be mandated by the department apart from whether the ratings are utilized administratively for such consequences to the individual as merit raises, tenure, promotion, or class assignment. I find that some faculty who are wary of student ratings but not adamantly against the whole idea are favorably disposed to using them in this voluntary self-improvement way. The research evidence as to this use is somewhat variable and its interpretation therefore a trifle subjective, showing that the efficacy of student feedback is statistically significant, although rather smaller in size than one would hope. P. A. Cohen's (1980) meta-analysis of teaching effectiveness reveals typical improvements of around $\frac{1}{3}$ of a standard deviation in total rating ("overall" and "average important trait") and shows larger changes possible for the more important variables (e.g., instructor skill). Informational feedback augmented by consultation produces better results for instructors than merely making the ratings available. Some ratings are more susceptible to change

than others, and some teachers move more than others. These differences are hardly surprising and would be expectable by a reasonable person after a few minutes' thought, even if the empirical data were not available to us. Suppose, for example, that an instructor is far below average in sheer verbal fluency, then he is not likely to be rated well in explanatory skill, even if he is trying hard to explain material in class. Some unfavorable ratings a person might find distressing and be strongly motivated to change, whereas others would not trouble him.

I can provide some interesting anecdotal matter as to this. Every time I administered the student rating scale in my clinical psychology class I attained an overall global rating in the top 2% in the college and on almost all of the other traits in the top 10%. But there were two traits in which I got poor scores, falling below the median of the college norms. One of these disturbed me because it conflicted with my self concept as being a reasonable, rational, basically fair-minded thinker, namely, "How tolerant is he of viewpoints different from his own?" I couldn't understand this, so I inquired of three very bright students who took the course and who also were personal friends of mine, two of whom I had known since high school days; I knew that they would speak frankly to me. All three of them independently said this bad rating was because when I analyzed a grossly defective research study or a clinical practice that was not scientifically defensible, even though I didn't distort the evidence and my reasoning was sound, I often tended to say *in conclusion* something derogatory about the investigator or the clinician, such as "this is a stupid way to proceed." They pointed out that, for many students, you aren't supposed to say that anyone is stupid or incompetent and that they were mixing up condemnatory phrasing with intellectual dishonesty. This confusion is probably especially tempting for the student in a "soft" area like clinical psychology, because there are quite a few procedures, diagnostic and therapeutic, that are not validated or, even worse, for which there is strong evidence against validity, but which are defended by their devotees via shoddy research studies. If a student has a strong Freudian or Adlerian orientation or is fond of projective methods, a critical scientific skepticism on the part of the lecturer can amount to taking away the student's security blanket. But having had it explained to me by these three frank friends (each of whom pointed out that he had not given me an unfavorable rating on the "tolerance of difference" trait because he *did* understand the distinction), I made a heroic effort the following year to avoid any editorial comments on the stupidity or carelessness of an investigator or practitioner, while still demolishing shabby studies as such on the merits. It was gratifying, and increased my marginal faith in the student ratings, that the result of this effort to avoid personalities while still insisting upon scientific rigor was that my "tolerance" rating increased very substantially, moving me from something like the twenty-fifth percentile of the college norms to the seventy-fifth percentile.

The other unfavorable rating was on the trait, "How well does he adapt his explanations to the level of the class understanding?" I must confess I didn't work as hard on this one, but I did make some systematic efforts, such as warning the class at the

beginning that I sometimes use technical terms from other fields, such as philosophy of science, that they may be unfamiliar with, and that this is mere forgetfulness on my part and nobody should be ashamed to raise a hand and ask what a word means, because the other people in the class probably don't know it either. I also tried to make it a point when offering a definition or explication to speak very slowly and in some cases to say it twice. And thirdly, I sometimes asked for a show of hands saying, "How many of you know what thus-and-so means?" Despite these efforts (which admittedly were a bit half-hearted because I bore easily) the change next year on this rating, while it would have been statistically significant in terms of the standard error of measurement, was only slight in size and still left me below the median of the college on this trait. I concluded that I was not able to do much about this characteristic and furthermore, on discussion with some of the top students, I concluded that I wasn't sure I wanted to change it. It is a fact known to everybody that has taken courses in social, biological, and physical sciences, that the intellectual demands made in the typical social science course are quite minimal, and a great deal of what is said in classroom lectures could be described as obvious and trivial. I knew that when I was a student I liked those professors who said lots of things I hadn't thought of and who strained my powers of comprehension, and I was irked by professors who said mostly things that had been obvious to me since high school or that were in the textbook. So I concluded that it was my pedagogical fate to speak to the kind of student that I and my first-class friends were as students and the rest could get whatever they could get from it. I realize that this is an elitist kind of remark but I have already said elsewhere in the book that I am basically an elitist about the life of the mind. Every year, over many years, I have received fan letters from previous students who took that course or my graduate seminar in methodology anywhere from ten to forty years ago, saying such things as, "Yours was the most intellectually exciting class I ever took at the University." Since I believe that not everybody is equally good at everything in the academy, my conclusion is that it's all right for people of my sort to lecture as if everybody in the class was as smart and informed as the top ten percent.

The relevance of these two stories to our present problem is that they help to flesh out what the statistical research shows about student rating of college teachers. To improve the useful application of this first usage it would be desirable to have some faculty and student observers participate in the consultation process, some of whom have been antecedently informed about which ratings had come out rather badly. In interpreting mean values of change from meta-analysis, it would be desirable to know something about the motivation of the various instructors. That some just don't care very much, or don't see much impact on their career from being rated poorly on, say, "rapport with students" (I know such persons), pulls down the mean shift in the meta-analysis. But it would be foolish to allow the existence of such persons in the sample to cancel out the improvement scores of those who are motivated to work at improving those ratings that are subject to change. Sometimes even a fairly simple and easily cured "mechanical" error may be involved. *Example*: I have seen two otherwise able instructors (mathematics,

statistics) to stand directly in front of the portion of the blackboard on which they are writing an equation so the class can't see it, and then erase what was written almost immediately upon continuing with the lecture. A trait that was not among those rated in my class, but could have been covered by a rating on "distracting mannerisms," was my tendency, especially on topics where I became intellectually excited, to pace back and forth, which some students found amusing but others found distracting. It was only by looking at the open-ended comments at the end of the questionnaire (there should always be that option) that I became aware of both the fact and the bad effect, at least for some, of this mannerism and regularly reminded myself to inhibit it.

A second and somewhat more controversial use of teacher ratings is with respect to teaching assignments. Ratings could be undertaken at the initiative of the informed teacher but also used administratively. It is a matter of common knowledge that there are individuals who are gifted at one type of instruction, say, a highly technical seminar for advanced doctoral candidates specializing in the teacher's field of research, but not good at general instruction with a class of 250 freshmen, many of whom are taking a required course and are basically not academically inclined in the first place. I know of no evidence showing a negative correlation between rated abilities and these levels, but I have seen no data showing them to be strongly positive, either. Relying on anecdotal evidence, I should think this would be a helpful use. There is considerable anecdotal evidence about well-known scientists who were apparently oblivious of their own defects of exposition. The great Niels Bohr once expressed amazement upon hearing that someone had said of him that he was a brilliant mind, but a rather poor lecturer—a combination of characteristics well known to physicists at all levels of competence and sophistication. There are some excellent departments that have a policy that every major professor must take turns instructing in the elementary sections, but I would be astonished if that was a boon to the students in every case.

A still more controversial use is to assist in administrative decision-making at the level of nontenured faculty. I would be strongly opposed to employing ratings *throughout the range* in connection with matters like merit raises; but I incline to think it legitimate at least at a state university, which, let us say, cannot, by policy or resources, enjoy the luxury of research competent faculty who are unable to do a good teaching job at any undergraduate courses. If, despite consultation and faculty visits, such persons are either unwilling or unable to rise from the depths, they should not be continued. In saying this, I am relying on the statistics presented above, concerning teachers in the bottom 10% on global student ratings or suitable multiple trait composites. If an academic administrator agrees, however grudgingly, that one should pay at least some attention, however slight, to how students react, and to the demonstrated probability of that reaction being appreciably correlated with what students learn, then one could make a case, even though it troubles some people, that there should be a floor of teaching ability for achieving tenure rank.

Finally, a use which most of my colleagues disapprove, and about which I myself have grave doubts, is at the very high end [*of administrative decision-making*]. Here would be some cases in which a professor's research productivity is marginal, and in which what appears to be outstanding instructional talent would warrant us in granting tenure, promoting to full professor, or meeting a retention case challenge. Challenging Sacred Cows I and IV, proposing the radical reforms I do with regard to "research institutions," saying that $^{4}/_{5}$ of the faculty should not be required to publish or perish, and should teach more hours, and yet to say that in those departments which do have a research unit attached… should even superb classroom teaching receive no weight in doubtful cases? I don't know what to say, but like most of my colleagues, I find it distasteful if not literally inconsistent.

I am not prepared to say anything very strong about teacher rating scales, but I will insist that a professor who rejects them out of hand on the grounds that he doesn't trust the discernment and conscientiousness of which students make the ratings, should in consistency admit that he doesn't think we should pay any attention to classroom lecturing skill, since he is not warranted in substituting two or three nonrandom anecdotes for carefully constructed scales. If student ratings are employed as part of the total evidence that goes into granting tenure, giving raises, meeting retention offers, and the like, I believe careful attention should be paid to whatever systematic trends exist between junior and senior college classes, between large and small classes, between classes that involve mathematics and classes involving only words, between classes that involve memorizing large amounts of material (for example, a beginning biology course learning the scientific taxonomies), and certainly classes that are required of many students versus electives. I do not know how big these differences are, and my own prediction is that with the possible exception of mathematically difficult material, they would show only small systematic trends. But if some large differences between classes categorized in various ways are observed, then it would be only fair to establish separate norms for interpretation.

We all have heard horror stories about concrete episodes that exemplify poor ratings on certain traits. It is unfortunately a fact that there are a few professors, very few but not zero in number, who are sarcastic, arrogant, or rejective to students and who view the classroom as a place where they can get by with the expression of malignant motives that would not be tolerated by their equals. Since I do not operate on the absurd premise that everyone has some sort of an inherent and indefeasible right to be a college professor any more than I have a right to be a brain surgeon if I'm afraid of blood or am all thumbs, I do not think it inappropriate to deny tenure to someone where repeated administration of student rating scales in a suitably normed fashion reveals extremely bad student impact. And taking the optimistic view of the correlations with student achievement, I think this is warranted even if one were to say that student impact as such, aside from its objective correlates in what is learned, is of no importance to us. That some injustice or disappointment will occur from time to time, even when hedged carefully with precautions and

the availability of grievance procedures, convinces some opponents that teachers' ratings should not be required. I understand this position and I am sympathetic to the concerns. I can only say that the statistical inevitability of a certain amount of injustice or psychic pain on the part of faculty who get bad ratings does not seem to be dispositive of the issue, except for one who believes that the comfort of professors is always to be weighted more than the psychic comfort of thirty students, a view which I think neither the students nor their parents nor the taxpayers would accept. I think it comes down basically to whether one believes it is or is not of any importance to know how students react. But if it is, then simple psychological and statistical considerations suffice to show that it should be done with formal rating scales, if carefully built and normed, rather than by anecdotes.

Additional References for Teacher Rating, but not discussed in text:

Blackburn, R. T., & Clark, M. J. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. *Sociology of Education*, *48*, 242-256. [Covers studies of teaching & research—cites many, *r*'s negligible but (+). Self-evaluations agree poorly with colleagues, deans, students. But latter agree fairly well. Is cited in Marsh (1987).]

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal*, *14*, 17-24.

Gillmore, G. A., & Greenwald, A. G. (1999). Using statistical adjustment to reduce biases in student ratings. *American Psychologist*, *54*, 518-519. [Student ratings, statistical corrections for nuisance factors (e.g., leniency).]

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. (1964). Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, *55*, 344-351. [Good except they call 1st Factor "Skill" [I agree] but then text says it's "halo"???]

Marsh, H. W., & Roche, L. A. (1999). Rely upon SET [student evaluations of teaching] research. *American Psychologist*, *54*, 517-518. [Reply to critics]

5. Meetings

It is a cliché in academia that faculty meetings and committees are an irksome necessity taking precious time from one's scholarly work. Like most clichés voiced by members of a trade or profession, this one expresses an element of truth, but it is only a half-truth; and the half that is false is important as a money waster. No quantitative research study is needed to conclude that professors differ a great deal in the extent to which this aversion to meetings is genuine. There are persons who really do detest taking the time for them, who would not be present if they did not have a combination of social pressure and inner conscience requiring attendance, and who relish legitimate excuses, even a minor illness, not to show up. (I enjoyed "emergency" appointments with psycho-

therapy patients that sometimes provided an excuse to miss a meeting.) At the opposite end of this attitudinal dimension are people who quite obviously enjoy meetings and would feel deprivation if they were no longer part of the required job functions. I became convinced many years ago that for some persons in the academy, having meetings (whether a faculty meeting, a committee meeting, or sitting on a PhD oral, a team meeting in psychiatry, or whatever) is a significant portion—for some people I would think the biggest portion—of their social and "intellectual" life. It is evident, again relying on strong anecdotal impressions, which is all I have to go on, that the number of hours spent in faculty meetings can be reduced or increased depending upon the strength of the motivation to do one or the other. In my university there are departments that have mandatory faculty meetings every week, each lasting from a couple of hours to an entire afternoon. When I point this out to colleagues in psychology, they are horrified and question what on earth the people in these other departments have to talk about when they meet so often. That it is not necessary to do so is clearly demonstrated by other departments which have more different activities going on, larger student bodies, more graduate students, more external grant support, more teaching and research relations with other departments, and yet have far fewer meetings. (I know of one distinguished department—some consider it the most eminent among all colleges of my university—which has fewer meetings of its full faculty than are required by the university regulations.) The psychology department has a mandatory one meeting per quarter and adds to that only as agenda accumulate or something urgent, such as a retention case, arises. Several times in recent years regularly scheduled faculty meetings have been cancelled because of too sparse agenda. If a large department, with many faculty and students and numerous scholarly relations with other units, can manage with two or three meetings per quarter, a smaller and less diversified department could obviously do so if its members wanted to. One can hardly avoid a conclusion that if they have three to five times as many meetings per quarter, it's because they prefer to spend that much time that way.

I do not mean to challenge the necessity of faculty meetings in general. I am only saying they can be held to a minimum if the professors sincerely want that. Parkinson's Law applies in the academy as it does in government and to some extent in private business when a company becomes large enough to be, in effect, bureaucratized. I look upon the faculty meeting chiefly as a guarantee of democratic control; and that is quite different from viewing the group process as a high quality, *positive* setting for complex problem solving. One of the widely held illusions, not confined to the academy, is that thirty or forty heads are better than one for problem solving. Nobody has ever proved this, and industrial psychologists who investigated "brainstorming" twenty or thirty years ago found pretty good evidence to the contrary. In carefully counterbalanced designs in private industry, it was shown that the quality of ideas produced by a group in a meeting is inferior to what the individuals produce thinking solo (Dunnette, Campbell, & Jastad, 1963). I would have been astounded had it come out otherwise. But what did surprise me, and surprises most people when they hear it, is that even the quantity of new ideas

produced (per person) declines to a statistically significant extent. My colleague, P. W. Fox, has a derisive formula which I took as an insightful joke, but on pressing found that he meant it seriously. His formula is that if you want to know the effective functioning IQ of a group as displayed in discussion, you take the mean IQ of the individuals in the group and from it you subtract twice the number of individuals: $(IQ)_G = (\overline{IQ}) - 2N$. This seems extreme, but he sticks to his guns when challenged. It is probably safe to estimate the mean IQ of the faculty in a first-class department like Minnesota psychology at around 150, so if 25 of the voting faculty attend a meeting, Fox's "group IQ" is around 100. I think this is somewhat too low, but I have had to admit that on some occasions that seems about the intellectual level of what is produced.

There is a considerable research literature on group versus individual efficiency in various kinds of tasks, and the interaction of factors seems complex, hard to generalize, and not robust enough to warrant strong generalizations (Dalkey, 1969; Dalkey & Helmer, 1963; Diehl & Stroebe, 1987; Dunnette, Campbell, & Jaastad, 1963; Einhorn, Hogarth, & Klempner, 1977; Goodrich, Henry, & Goodrich, 1954; Guzzo & Shea, 1992; Hill, 1982; Janis, 1972; Miner, 1984; Varca & Levy, 1984). As I read the record, there is a "process loss" due to the group dynamics that often counteracts the heightened probability of good ideas flowing from sheer statistical aggregation. I here adapt a personal communication (1997) from a disinterested, high competence colleague eminent in Industrial/Organizational psychology who shared his summary conclusions from the research evidence:

> If (1) the quality of a problem solution is the relevant outcome; (2) the group members know each other; (3) there are *no* significant vested interests in how the individual vs. group comparison turns out (i.e., scoring is not biased); (4) the problem is difficult and ill-structured; and (5) the members each possess significant, but perhaps varying, amounts of expertise: *then*, there will *always* be a shortfall when the group's output is compared to the solution(s) that would be possible by asking a knowledgeable individual to develop a synthesized solution from the *independent* efforts of the group members.
>
> If the utility of spending time in teams is to be positive rather than negative, it must come from the value of other outcomes, such as:
>
> - More communication of relevant information among team members;
> - Satisfying social needs;
> - Just getting to know each other better;
> - Making everyone responsible for certain management control functions (e.g., monitoring each other's work, providing feedback, coordinating multiple efforts);
> - Achieving public commitment to critical goals.

A powerful argument is found in massive coherent evidence from history of science, by far the most successful cognitive enterprise minds-in-society have ever practiced. Historians have pointed out that major theoretical breakthroughs are almost always achieved

by single persons of unusual talent, drive, and luck, working largely alone. Even pairs of such first-rate contributors (e.g., Cayley and Hamilton in mathematics) working together are rare, although independent "multiples" of discovery occur with frequencies up to a half-dozen close in time. Today's burgeoning of multiple authorship is due to the multi-disciplinary demands of empirical research (e.g., medicine, nuclear physics) plus the social pathology of our absurd publish-or-perish disease. So far as I know from extensive reading in biography and history of science, *not one major theoretical contribution was the work of a committee*. This should tell us something.

The kinds of tasks ordinarily faced in a full faculty meeting, a special committee, a standing committee, etc., are so variable and the group composition so idiosyncratic that one hardly knows how this research should be applied to the academic setting. Being forced to rely partly on my half-century anecdotal impressions, theoretical considerations, and a whiff of "common sense" (e.g., the folklore sarcasm about the camel being an animal designed by a committee), I am courageous enough to opine, until clearer research evidence, meta-analyzed, says otherwise, that considerable professional time is utilized sub-optimally in group meetings.

Distinguishing among kinds of tasks and the sort of group process, size, and previous acquaintance of members involved, a typical academic faculty meeting is more similar to the industrial psychologist's "brainstorming" than to laboratory experimental pseudo-tasks (e.g., estimating the population of Washington, DC, or solving the cannibal-missionary crossing the river problem). The evidence is convincing that brainstorming—a much touted fad in the business world 40 years ago—is not an efficient procedure, the work-product (*good*, *new* ideas produced per person) suffering in both quality and, surprisingly, quantity.

Why is this? Being neither a social or industrial psychologist, I have no technical expertise in group process, and as a clinical practitioner I have never engaged in group psychotherapy. But I do have some conjectures which I offer as plausible, and in principle researchable, although it would take quite a bit of work. The descriptive "Murray needs" of personality theory are useful here. The optimal mental set for effective problem solving would be almost exclusively motivated by Murray's *n Cognizance*,[28] with an assist from *n Order* and—even as in creative scientific and artistic cerebration—*n Play*. When you get a group of humans into a room, a considerable set of other needs are mobilized and these frequently operate in opposition to the effective three I have just listed. We are dominant, affiliative, aggressive, erotic, social primates, and there is hardly anything about *n Dominance*, *n Aggression*, *n Affiliation*, *n Sex*, *n Nurturance*, *n Deference*, *n Succorance*, and the like that tends to improve intellectual functioning. Everybody who has attended faculty meetings knows about this. Jones likes the sound of

---

[28] [*Read as "need Cognizance"; a need postulated by Henry Murray (1938, p. 744), characterized by an "Inquiring attitude…. To explore (moving and touching). To ask questions. To satisfy curiosity. To look, listen, inspect. To read and seek knowledge." Other needs mentioned here should be self evident.—LJY*]

his own voice even more than most professors. Smith hates Robinson for being brighter than he is. Fisbee wants to be sure that others recognize that he is a wise, fair-minded, charitable person. Hockheimer is intellectually arrogant and a dogmatist. And so it goes. One speaks of "hidden agenda" as a matter of course, but it is important to realize that in addition to hidden agenda there are all sorts of agenda that are hardly hidden at all.

I can illustrate this problem of irrelevant needs dominating discussions and getting in the way of rational problem solving by my experience serving on the Arts College dean's promotion and tenure committee. The year I served, some fifty faculty were being evaluated, around half for promotion to tenure and the other half promotion from associate to full professor. If a candidate was marginal (neither a shoo-in nor grossly incompetent), so that one would hesitate to lay a bet on how the committee would decide his fate, the ideal situation for such a gray region candidate would be to come up on the alphabetical list following another marginal candidate who was rejected for tenure or promotion after somewhat heated dispute. Even safer was to follow two such rejects. The situation would be that the "winning side" was feeling guilty about having forced a negative decision about the previous poor devil's fate, and also somewhat embarrassed to have overcome the opposition, who had wanted to promote or tenure, and therefore they would wish now to ingratiate the losers (who meanwhile would now be thirsty to "win one"). I concluded that if you were a marginal candidate, this was a sure-fire guarantee that you would make it. I am sure that a content analysis of protocols of group discussion would unearth many other sequential patterns of this sort when the objective merits would yield a borderline case to be decided.

When I chaired the psychology department in the 1950s, we were in a transitional period in that we were rapidly expanding our faculty and moving into new areas of teaching and research, and there was some degree of stress and friction because of the changes that were taking place. I wrote a thoroughly democratic department constitution (at a time when constitutions were not required by the university's rules) and I abided strictly by it, as everybody knew. But strict, honest obedience to the "political" forms (which included such rules as at least one week advanced notice of agenda, mandatory ballot vote on certain topics or when even a single person requested it, and the like) did not mean that I had any great faith in the group process as a problem solving machine. If a complicated issue had arisen, I usually spoke with one of the most rational and high-competence faculty members to get a suggestion, and then tried it out on somebody else equally rational and informed, choosing that second person partly on the basis of his being "on the other side" from the position of the first one. Sometimes the proposal that emerged in my mind after these informal conversations (and no effort was made to conceal that they took place) would be proposed in a written agenda. Other times I would refrain from a proposal until the meeting itself but only put the *problem* on the agenda.

College professors are prima donnas who like the sound of their own voices, so the first thing you have to do is give them a chance to talk. And, like a good psychotherapist, I didn't argue with them. I didn't argue with people even if I disagreed strongly with

what was being suggested. If it was, in my view, a sufficiently foolish solution, I was certain that somebody else—without having been primed by me in advance—would point out what was wrong with it. In a complicated controversial matter, what usually happens is that people begin to run out of gas and start looking for leadership by the alpha baboon, which is the chairperson. When it was pretty obvious that they had satisfied their need to talk, and that some of the hidden agenda had been expressed, I would summarize the discussion and then propose my preferred solution. The result of this mode of operation was that in six years, in the course of which quite a few controversial matters came up, I only "lost" a single vote. (That was a resolution expressing our support of another department, which my colleagues knew I was doing as a favor to a friend and I really didn't care about it any more than they did.) The sociobiological fact is that humans, while they do not want to be tyrannized over, usually wish to be led. This is why it is important for a department chair to have a high order of intellect and to be esteemed as a scholar, despite the wastage involved because ninety percent of what the chair has to do in routine matters could be done by somebody of thirty less IQ points. But since in the academy "brains is the name of the game," professors want to be able to truly inwardly esteem the scholarly qualities of the leader.

If one could get by with it without arousing resentment, my ideal approach as an administrator would be to get one of the smartest people I could convince to take some time to write a position paper on "the issue," which I would then have critiqued by somebody else equally able, and then make up my mind what should be done. That occupies the time of three people instead of thirty or forty and, in my opinion, would on the average yield a higher quality of problem solution. I realize that this viewpoint can be characterized as "elitist," for which I offer no apology because I am an elitist in matters of this kind. I notice that people who claim to dislike elitism, the people I am inclined to call "frenzied egalitarians," suddenly become quite elitist when their own vital interests are involved. A frenzied egalitarian whose son gets busted for dope doesn't pick a lawyer randomly out of the Yellow Pages; he starts calling up friends who are knowledgeable and asks who is "the best defense lawyer in town for a drug case." When people are going to have a brain tumor removed, they like to know that they've got one of the best brain surgeons in the business. And so it goes. It is absurd to pretend that all people are equally competent, and having lots of meetings based on that erroneous idea is a waste of the taxpayers' money.

I have the impression that on our campus there is a negative correlation between the national scholarly status of departments and how much time they spend in meetings, and that would be quite consistent with Parkinson's Law. In academia the official emphasis on scholarly research is applied pretty much over the board, despite some differences in teaching loads among departments that take their quality into some account. And that means that in departments of low scholarly production—and I am not here contradicting what I say about the sacred cow of useless research; I'm just talking about the administrative setup—people literally have time on their hands. If you aren't in the laboratory

or the library and don't dare to spend your in-the-office time listening to the baseball game, what are you going to do with your time in order not to feel guilty? Admittedly there are some freeloaders who make little or no attempt even to look as though they were pulling their weight, but it is my impression that the majority of college professors are not comfortable being loafers, and therefore they need to be spending their time on something that can count as a legitimate use of it. Having a meeting is for such people a relatively painless way of satisfying the old superego that you are earning your keep.

It goes without saying that if one is a compulsive ingratiator these insights won't do any good, because you will have difficulty saying "no" to people who want you to do something. My practice, for over forty years, has been to ask four questions about a committee before agreeing to serve on it. First, is the committee's task of any real scientific or social significance, or is it simply Parkinson's Law make-work by some busybodies, or something set up to ward off criticism? Secondly, does it seem at all likely that I have anything unique or special in terms of knowledge, analytical skills, social potency, or the amount of time I have spent thinking about the issue in question that at least one other member of the committee doesn't have? "Nobody is indispensible" is true in large groups, but in a five-person committee facing a hard task, Professor Jones's refusal to serve can sometimes make the difference between failure and success, if she, *among those available*, has sufficiently rare knowledge, skills, social potency, etc., to determine the quality of a work-product. Thirdly, are the members of the committee smart enough, informed, and rational enough so that if I present my special and valuable ideas they will be capable of grasping and assimilating them? Fourthly, assuming that the committee produces a report that has some content (not the usual set of compromise clichés and academician's jargon), is the political and economic situation such that the report has a reasonable likelihood of being implemented? It is my experience that each of these four conditions has a good deal less than a fifty-fifty chance of being empirically correct about any given committee job, so that their joint probability over a scholarly lifetime is surely less than ten percent of all committees on which one is asked to serve. I think that would be about my batting average of acceptance; and I claim that, using these criteria, I have had significant impact via committee work on the rare occasions where I have consented to accept it.

6. Student Selection

Several years ago, I watched a TV program produced by the selections committee of a small high-quality private college in which the process of screening applicants for admission to the freshman class was set forth in considerable detail. Unfortunately, I cannot recall the name of the school and have therefore not been able to correspond with them to check my recollection, but I am confident that the numbers are roughly accurate, because I distinctly recall doing some monetary calculations at the time. My recollection is that there was an administrative officer, say, an associate dean of admissions for whom

this activity constituted his full-time job, an assistant dean who spent half his time on admissions, and four or five faculty who varied in how much of their work load consisted in admissions decisions but averaging around twenty-five percent. If the dean of admissions had a salary of $100,000, and the half-time cost of the assistant dean was $40,000, and an additional four faculty's time contribution averaged $60,000, we get a total time cost of approximately $200,000 spent on this task. First, everybody had to spend many hours reading student applications, letters of recommendation, grade transcripts, students' statements of why they wanted to attend this particular school, etc., and making some preliminary ratings or judgments. Then there would be a meeting in which discussion case by case took place, with a final vote on each. One such discussion was televised, and it had the usual properties of poor focus, free association, repetition, anecdotage, irrelevancy, jocularity, non-operative language of agreement or qualification—the "group process" at its feckless worst. This procedure meant, of course, that some students were "obviously" suitable, although even there the usual press of speech by academics would take place despite the absence of any significant disagreement. Others would require ten or fifteen minutes' talk apiece. And then some marginal cases in which a certain adverse fact (e.g., a couple of poor grades in high school) would be set in opposition to a very favorable fact (e.g., highly successful athletic and other extracurricular achievements, or social popularity as evidenced by election to class offices), sometimes corroborated by letters from high school teachers, sometimes not—presenting an insoluble puzzle for the selection task. A group of six academics can spend a lot of time on such puzzles, and of course there is no "correct answer" available. *The important point is that this process is almost certainly a complete waste of time*.

A simple equation or actuarial table based upon grades or test scores, some weighted composite of "nonacademic" achievement, ratings on personal charm, etc., would certainly do as well as this time-consuming, unreliable group decision process, and stands a good chance of doing significantly better. There is now a vast research literature dealing with the relative efficiencies of impressionistic, informal, "clinical" judgment in predicting behavior versus explicit, formal, "actuarial" procedures of combining the same data. Let me emphasize, *this is not a question of using different kinds of data*. I am not talking about whether, say, an intelligence test is a better predictor of academic performance at college than high school rank, or whether a high school teacher's ratings on personal charm should be eliminated. The point is, given any specified data set (which could include letters of recommendation), the question is *how to combine them for predictive purposes*. There are now 136 published studies comparing the formal with the informal (the "actuarial" versus the "clinical") methods of prediction from a specified data set, ranging over a wide variety of behavior domains to be predicted (e.g., college grades, survival in air crew training, response to electroshock treatment in psychiatric patients, winning high school football games, making business firm decisions, criminal recidivism; see Grove & Meehl, 1996). A fair-minded summary of these studies would be that it is difficult to find more than 5% of such comparisons among the 136 in which informal

human judgment does significantly better than a regression equation or an actuarial table. Of course, when we deal with as many investigations as this, the statisticians tell us that you should expect a small proportion to come out in the "incorrect" direction solely by random sampling error. Hence, *it is not clearly established that there is even a single predictive task, ranging over this wide variety of settings and kinds of judges, in which the subjective, impressionistic method is reliably better than even a crude actuarial one*.

Theoretical considerations and research on cognitive processes show us why this could be expected, even though it goes against common prejudices and practices. Even if there are a few special contexts in which informal judgment performs better than the mathematically objective way of data combination, it is likely—I would say almost certain—that intensive study of the subset of cases in which informal judgments do better would identify factors which the equation had omitted to consider, so that a special factor could be then included into a revised equation. Most people have an inordinate difficulty in accepting these results, and I cannot think of another controversy in the social sciences in which the number and variety of research studies comes out so clearly in a given direction (Meehl, 1954, 1986a). The chief psychological block against accepting these facts seems to come from focusing on the "special case" in which the human judge notices something different that she believes is so important that it should countervail the deliverances of the equation or statistical table. This is a valid *qualitative* concern, but it does not suffice to show what people take it to show. The question is whether, when the human judge *thinks* that something can be discerned in the material that is not represented in the formal prediction system, that human judge is, in fact, correct in so believing. If the judge and the equation agree, the percentage of "hits" (successful forecasts) is a fixed quantity (since they're both saying the same thing); if the judge tends to be right more often than wrong in countervailing the equation because of a special fact or impression, it follows from high school algebra that *the overall hit rate of the judge must exceed that of the equation*. But since the studies show that this does not happen, we must conclude that the judge's accuracy in identifying countervailing facts is poor. There is no way to get out of that simple arithmetic.

A large body of research by industrial, military, and educational psychologists, going back to the 1920s, shows that letters of recommendation are about as close to worthless as any datum can be. My teacher, Donald G. Paterson, used to say that if several independent letters of recommendation are strongly adverse, you should probably reject the applicant. Or if you have calibrated a certain letter writer and know he doesn't regularly describe everybody as a genius, you could usefully attend to that. Given neither of those two conditions, the best thing to do with letters of recommendation is to throw them in the waste basket. A substitute for letters of recommendation which is commonly used these days in selection procedures is a set of quantified rating scales on certain traits. If the anchoring adjectives are carefully located, this does sometimes convey valid information and, of course, that format leads to easy representation in an actuarial prediction system. If the small, elite college on that TV program eliminated those feckless meetings

and saved $200,000, it could hire three or four assistant professors or reduce student fees by a significant amount.

7. Research Productivity

Evaluating the quality of scholarly research (even in the sciences, where it may be a little easier than in languages, arts, and humanities, or so many believe) is a difficult problem, and I do not wish to minimize the difficulties. However, when academics object to various statistical approaches to this evaluation problem as inadequate, or impossible, or not "fair," rather than merely saying they are fallible and subject to error (as are all human judgments), they are in pretty much the same position as faculty who object to the evaluation of teaching performance; namely, that *they themselves*, when they function as participants in the decision-making process regarding personnel (whether to hire, tenure, raise salary, meet a competing offer in a retention case), invariably rely on the same list of *qualitative* criteria suggested by advocates of (admittedly imperfect) quantification of research performance.[29] It's like the professor who says, "I don't trust student judgments and therefore am opposed to formal teacher rating scales on twenty traits," and then, when asked in a faculty discussion about promoting Joe Glotz to tenure rank, tells an anecdote about a student who told him that Glotz once said something sarcastic in his office.

How do faculty go about evaluating someone's research? Well, they can read it themselves if they're competent in the area; they can ask other competent faculty for their opinions; they can simply count number of books or articles or even, as I have seriously seen advanced pro or con in a tenure committee meeting, number of pages (!!); they can look to see whether the individual receives various prizes, awards, invitations to deliver colloquia at other institutions; they can look at the number and size of research grants from public and private fund-granting agencies; or they can ascertain how often the person's published research is cited in the research literature. Barring direct inspiration of the Holy Ghost or a claim of personal infallibility, this six-fold list pretty much exhausts the available data in personnel assessment. So the only question is whether one prefers to do this judgmental job in a helter-skelter, haphazard, and probably biased anecdotal manner, or whether these various *indicators* (I avoid saying "criteria" here), each of which is susceptible to some kind of measurement or counting, should be numerified. Given the pressing economic and morale problems of contemporary academia, I believe a large scale cooperative research project is in order in which these various indicators of quality and quantity are collected in a systematic manner and subjected to statistical analysis (e.g., factor analysis) to find out how well they agree with one another and whether there exists, as I predict there will be, one large common statistical factor that we could, without arbitrariness, label "research productivity."

---

[29] [*See* "Miscounting by Pretending Not to Count" *in the Introduction.—LJY*]

When I was chair of psychology, I began such a project but dropped it when I quit the chairmanship. First, I had the faculty assign rough subjective weights to the value or importance they attached to different *kinds* of publications. I found, for example, that there was impressive agreement among professors, almost perfect agreement as to rank order, and substantial agreement as to the numerical weights they assigned. Averaging the weights in a crude way, it turns out that if a scientific paper accepted by a refereed journal is arbitrarily given unit weight, then a book review gets a weight of one-half, a longish monograph or undergraduate text a weight of three, and a technical scientific treatise a weight of six. It would be silly to pretend that these have some absolute metaphysical significance, but it is equally silly to reject such numerical findings on the grounds that the composite score that a professor would receive is not a perfect infallible measure of his lifetime scholarly productivity in the eyes of Omniscent Jones. Either we are going to pay *some* attention to writing a book review, or we are not. It turns out that everybody on the faculty was willing to pay some attention to it. If we pay some attention to it, are we going to pay more attention to it than we would to writing a scientific paper on an experiment or theory? It turns out that everybody thinks the latter is more valuable. If everybody agrees the latter is more valuable, are we going to say it is twice as valuable, or thirty times as valuable? Since I am not a purist about psychometrics, I am quite contented to discover that there is a reasonable amount of consensus among college professors in psychology about the relative value or merit attached to these several kinds of scholarly publication. Such an index of "sheer quantity" of output could be correlated with citations of one's work in the *Science Citation Index*. Another approach would be to write inquiring letters to a sample of department chairpersons over the country and to carefully chosen "experts" in specific domains asking about one's own department: "Write down, in the order that their names occur to you, who are on the psychology faculty at Minnesota; check those that are in your own field of research." Finally, one can tally cumulative references to people's work in the *Annual Review of Psychology*. These various measures can be intercorrelated and factored, and weights assigned to the indexes on the basis of their loadings on the first big factor, or perhaps some composite of the first one or two factors, if they can be meaningfully interpreted.

This is nothing but formalizing what everybody does impressionistically—informally, subjectively, and anecdotally—when they participate in personnel decisions. If you're going to "count pages," then *count* them, rather than guesstimating them. If you're concerned with visibility in the field, the way to find that out is citations, and not anecdotes told by somebody you met on an airplane. As for the *Science Citation Index*, there are two nuisance variables that might be producing considerable bias, especially when we are evaluating young persons whose work has not yet begun to have the impact that it may subsequently receive. First, I would make a statistical correction of each citation on the basis of the total population density of scientists publishing in that field.

Thus, for example, the world's authority on the olfactory sense is Carl Pfaffmann,[30] who doesn't get nearly as many citations as would a rather minor figure in the field of clinical psychology, simply because there aren't that many psychologists working in the olfactory sensory domain. Secondly, for people who get cited at all, it is almost arithmetically necessary that there should be an increasing accumulation of cites over time, because if I've written fifty papers with a certain citation probability and you are twenty years younger and have written twenty papers with the same citation probability, obviously I will get a higher score. Thus, corrections for *age* and for *domain density* will almost certainly be desirable. How much such corrections matter and whether there are similar ones that might be offered for the other indicators, is itself a mixture of common sense and empirical statistical study.

In the social sciences, a special problem arises because of the existence of so many ephemeral "fads" of theory, data kinds, and instruments. I don't have any cure for this one, and I have not heard anybody else come up with one. A department wanting to guard against faddism can often protect itself by critical, rational policies of personnel selection, and the Minnesota department over the years has done that rather well. For example, the status of psychoanalysis, both as a method of healing and as a psychological theory, has markedly declined since my tenure as chair in the early 1950s. Since I myself was engaged in psychoanalytic therapy and intensely interested in it, I prevailed upon the faculty to put into its five-year plan a special slot for somebody with training at a psychoanalytic institute and more knowledgeable than I was. We hired a first-rate person for his repute in this domain to teach our course in psychoanalytic theory (and he meanwhile embarked upon a control case while he was at Minnesota). But, this individual, despite his psychoanalytic know-how and fairly strong Freudian theoretical identification, was not a dogmatic Freudian, being fully aware of the terrible methodological problems of testing Freud's ideas. He also possessed recognized expertise in several other areas of psychology, such as psychometric assessment of values, quantification of stories in the Thematic Apperception Test, behavior genetics (he had done some well-known early studies on the hereditability of traits like dominance and anxiety in the mouse), considerable expertise in the field of social psychology, and in the general field of personality theory apart from Freud. So if psychoanalysis had gone totally "out of fashion," we would still have had a first-rate psychologist on our faculty, not a Freudian fifth wheel.

8. Reverse Discrimination

As has been pointed out, discussion of the economics of higher education would ideally have to include acceptable measures of the "product," the quality and quantity of educational changes produced in the students, whether verbal knowledge (of facts,

30 [*Carl Pfaffman* (*1913–1994*)*, American physiological psychologist.—LJY*]

generalizations, and definitions), problem-solving skills, beliefs, attitudes, or values. The effectiveness of how an institution conducts the instructional process would in such an ideal knowledge situation be quantified; but in addressing the problem of how to pay for whatever we get (which I attempt with the calculations presented in Chapter VIII) there is no plausible way to include some of these important considerations. That is surely the case in the matter of reverse discrimination. The most that I am prepared to argue is that lower intellectual caliber in the faculty or lower intellectual caliber in the students' peer group (e.g., those admitted for doctoral training in clinical psychology) can plausibly be expected to have a deleterious effect on the educational outcomes. I am convinced from my anecdotal evidence to date that this armchair expectation is factually born out, but for obvious reasons I avoid reporting the anecdotes. Almost every first-rate intellect I have asked about this agrees with me, but many are chary of saying it to others or writing it down.

No one who has participated in discussion sections of an introductory freshman or sophomore class, then in senior college classes small enough for some discussion, and then in graduate seminars confined to PhD candidates, could doubt the difference in the quality of one's experience when evaluated solely in terms of intellectual content. (I leave aside whatever other advantages accrue from listening to the ideas of less able intellects; I know some value these highly, although I myself do not.) My first experience of this peer group effect took place in my adolescence because my junior high school (grades 7–9) used a "tracking" system and there were six sections assigned by IQ. It was quite obvious how it must have been done, although we didn't know what the cutting scores were. I learned years later, in a conversation with the science teacher of that school, that the cutting score for the "alpha" group was an IQ of 130, two standard deviations above the mean, so these teenagers were in the top 2–3 percent of the general population. In senior high school (grades 10–12) the principal was a frenzied egalitarian who strongly disapproved of such tracking, and intellectual levels were mixed in all classes. It was blindingly apparent—whether or not one put a big value on it—that the quality of classroom discussion underwent a marked decline, although this depended partly on the extent to which the teacher called on people or relied on volunteers. If she relied on volunteers it was the kids from the former alpha section who did almost all of the volunteering, and that was fun. Whereas if she called nonselectively on students, she got the whole range of talent, and that was usually pretty boring for those of us from the alpha group. There would be no reason for the restriction of class size in seminars and the requirement of numerous prerequisites, even "permission of instructor," or a pretest on knowledge, if students and faculty had any real doubts about this matter.

I use the term 'reverse discrimination' to refer to affirmative action as it is currently practiced on many campuses, including my own (although apparently somewhat less so here than other places), not with inflammatory intent but as an objective operational description of the process. My semantics is this: If there is a policy of admitting students or a policy of hiring faculty which either explicitly, or as inferred from the statistics,

prefers Applicant A over Applicant B when there is a clearly discernible difference in intellectual quality of B over A, but A belongs to a minority group or is of female gender, this I call reverse discrimination. There is no valid reason for not calling it that, because that's what it is.

While my objection to it in the present context is a social engineering objection in terms of reduced quality of research and teaching as a probable consequence, it is unavoidable for me to say, however briefly, something about my moral and legal objections to it. To begin with, as regards faculty positions, it should be obvious to any thoughtful person that nobody has an intrinsic indefeasible "right" to be a college professor. It is remarkable how many defenders of reverse discrimination accept the idea of such a right as an implicit premise in their thinking. Nobody possesses a *right* to engage in any specified profession that affects other persons and costs the taxpayer money. *Analogy*: Suppose that I am attracted by the idea of being a brain surgeon. In such a career, one has the great satisfaction of knowing that there are not hundreds, but, over a lifetime, several thousand patients given an extension of life or an improvement in functioning because of one's professional ministrations in the operating room. It is, according to various surveys, one of the very top prestige occupations one can have; it pays extremely well; great deference is accorded to you by everybody, including other physicians; it's interesting work, and while you do have to get up pretty early in the morning, and some operations last a long time and are quite fatiguing and stressful, nevertheless, there is a sense in which it appeals to an important human need, namely, autonomy: "I am the boss, and I don't take orders from other people." This is a very attractive set of attributes in a job, both altruistic and egocentric. So, if I have an ambition to become a brain surgeon, it is certainly understandable that to be prevented from achieving that goal would be frustrating to me. Assume that I have an adequate IQ to get through medical school and good spatial intelligence and steady nerves and am not upset by the sight of blood. There is, however, one little problem; namely, that due to an unfortunate set of genes or anomalies in some portions of my brain, I am all thumbs. I often have trouble using a can opener or tying my shoes. Nobody in his right mind thinks that I have some sort of inherent, intrinsic, indefeasible right to become a brain surgeon because I very much want to, despite the fact of my inferior motor performance. It's not my "fault," after all, that I am at the fourth percentile of the Johnson–O'Connor Finger and Tweezer Dexterity Test! Furthermore, the question how it comes to pass that I have this grave deficiency in one sector of my abilities is literally irrelevant. It could be a genetic trait due to the accident of which sperm cell got there first. It could be that I come from a gene pool in which most available genes are of that sort. It could be that I am suffering the aftereffects a mild undiagnosed case of mumps encephalitis. Or maybe I had a tyrannical father or cruel shop teacher who traumatized me emotionally, so that anytime I have to do anything delicate or complicated with my hands under observation and time pressure, I suffer undue anxiety. While a moral philosopher or God in his infinite justice might be concerned about the historical events that gave rise to this counter-surgical

ineptitude, which weighted composite of them is responsible for it is completely irrelevant to whether society is somehow "obligated" to give me a license to do brain surgery simply because I want to very much and did not "choose" the set of causal factors that impairs my performance. Nobody chooses his genes, nobody chooses the kind of parents she has, nobody chooses the sort of neighborhood he is raised in, or any of the other factors that may operate causally to one's disadvantage. The same, of course, is true at the high end of the ability distributions. When I took a class in group aptitude testing from the educational psychologist W. S. Miller (who built an excellent test for discriminating among high IQ students, the Miller Analogies Test) he pointed out that a person who is proud of his IQ is like a cow at the state fair who gets a prize for a high butter fat yield in her milk, which she had nothing to do with, it being a consequence of breeding.

When we are selecting people to attend dental school or medical school or law school or to get a PhD in clinical psychology, there are three distinguishable groups who have a valid interest in the selection procedure. The student applicants obviously have a personal interest in who is accepted and a kind of "right"—which, however, must be carefully defined in a realistic and moral way—to be considered on their merits; and further, those who are accepted have, as indicated above, a pretty strong interest in being surrounded by an intellectually stimulating peer group. Secondly, the taxpayer has a legitimate interest in the caliber of the students admitted for training, especially in areas where student fees contribute only a small fraction of the total cost of instruction, as in medicine and technology. (I have seen figures that ninety-five percent of the cost of training a physician today comes from the taxpayer, with the student paying only about five percent.) Thirdly, in many areas and certainly in the "helping professions" such as law, medicine, and psychology, the future clientele—whether they are patients, clients, business firms, governments requiring the services of the civil engineer, or whatever—have a strong interest in professional competence.

It is not necessary to present any statistics to say that, in many fields of technical education, it is literally impossible, economically and logistically, to admit all interested applicants on the theory that the competent will survive and the rest will drop out. In our PhD program in clinical psychology in Minnesota, what the industrial psychologist calls the *selection ratio* (the proportion of applicants admitted) runs between five and ten percent, so that there are ten to twenty times as many students applying as we could handle and retain our accreditation. The idea of admitting them all (say, 150!)—filling up the classes, advising them, providing study cubicles—is so outlandish that I have not heard anybody seriously advocate it. The most obvious way to meet the interests of the three groups fairly (applicants, taxpayers, future clientele) is to select applicants on the basis of their track record, test scores, and whatever else has been shown empirically to be evidence of success in surviving the program. If, in practicing reverse discrimination, one admits a person of lower ability, thereby necessarily excluding a competitor applicant of higher ability, one is going against the legitimate interests of all three of the above groups.

My objections to reverse discrimination are four-fold: First, I view it as a violation of Title VII of the Civil Rights Act, when it is conducted by a state university, which is an arm of the government and falls under that section. If refusing to read the plain language of the statute (which some federal judges seem incapable of doing) one turns to the legislative history for clarification of the meaning, it is a fact that in that legislative history when the opponents of the bill objected on the grounds that there was a danger of reverse discrimination and the setting up of quotas, they were assured by its supporters that the language would be written in such a way that this could not possibly occur. It is now occurring at the hands of federal judges, federal bureaucrats, and college administrators. If one misconstrues the language of the statute and the legislative history so as to warrant reverse discrimination, then I hold, as do many law professors and at least some judges, that the act as thus interpreted is in violation of the due process and equal protection clauses of the Fourteenth Amendment.

More important to me than these legal matters, however, is that I view reverse discrimination as fundamentally unethical. What it does is attempt to achieve commutative justice to a class of persons, not identified other than by their ethnicity or their gender, by committing a violation of distributive justice with respect to the *individual* person before us. I believe that this is fundamentally wrong and I am not aware of any other area of socio-political life in which such a strange procedure has been seriously defended. There was a time when police officers in my section of Minneapolis (an "academic" neighborhood) apprehended marginal speeders who were driving foreign cars more than they did those driving American cars. I understand that a stop has been put to this unfair practice. But suppose the police chief had said to the officers, "Look, we have been doing a wrong thing in tagging people who were going thirty-three miles an hour driving a Volvo and not tagging people driving thirty-five miles an hour driving a Ford. In order to make up for these wrongs to these two classes of people, foreign and domestic car drivers, for the next six months you should let foreign car speeders go and issue tags to drivers of American cars." I cannot conceive of anybody in their right mind advocating this as a morally defensible position. But in reverse discrimination that is exactly what we do. We say that Mr. White applied to dental school ten years ago at West Overshoe College and he was admitted at the expense of applicant Mr. Black, who had better grades and a higher IQ and superior manual dexterity. The basis of this injustice was Mr. White's skin color being like his name, and Mr. Black's like his. As a result of this, we have a problem of commutative justice, that is, undoing a wrong. The favored Mr. White and the wronged Mr. Black are no longer available to us, but today Mr. Jones, who is black, wants to go to law school and Mr. Robinson, who is white, also wants to be a lawyer. Mr. Robinson has better grades and a higher law school aptitude score, but we reject him in favor of Mr. Jones because Mr. Jones has black skin color. So what we have done is to commit a second injustice, based upon a non-relevant attribute of skin color, except that we're now doing it in the reverse direction. I have not seen any persuasive moral argument for the legitimacy of such a proceeding and I do not expect to find one anywhere.

Having presented briefly and I fear dogmatically my legal and moral objections, I come now to the one that's relevant to our present problem, that is, the problem of educational quality. You do not need a PhD in mathematical statistics to follow this simple actuarial argument. We have a class of persons W and another class of persons B who would like to get a PhD or—let's take this case—want to be professors at our college. Assume that there is no appreciable difference, or one too small to be of any practical significance, in the distribution of ability and achievement between the Ws and the Bs. In that case, if we hire on the merits—partly track record, and perhaps other evidence forecasting future performance—there will, after a period of time, be no difference in the percentage of Ws and Bs except whatever is reflected in the statistics of job seeking. (If someone objects to the "after a period of time," on the grounds we have to repair past group-based injustices right away, I refer him to the moral and legal arguments above. *They're not the same people*!)

Suppose, on the other hand, that for whatever reasons, perhaps the causes we cannot satisfactorily untangle, or if we can unscramble them conceptually we have no presently available main means of manipulating them, the quality of the Bs is lower than that of the Ws. If this is an empirical fact, then it behooves rational social engineering not to pretend otherwise. The whole point of reverse discrimination is that, recognizing an appreciable statistical difference between the two groups, it consciously or surreptitiously adopts a policy of applying different standards. An undergraduate statistics course is not necessary to see that if that is what is done—and that is what reverse discrimination *means*, it means taking one person of lower competence in preference to another, on the basis of race or gender—it follows as the night the day that the quality of the department must decline. Either there are differences in track record, tested ability, motivation, etc., between two groups or there are not. If there are no such differences, reverse discrimination is unnecessary, as the proportional representation will reflect proportional interest. The problem takes care of itself statistically if we select on the basis of talent and achievement. If we select *against* talent and achievement on the basis of the nonrelevant attribute, we are doing so because we have seen from our statistics that there is a *difference* in talent and achievement, which we intend to countervail by reversing the discrimination. It is an arithmetical truism that this is sure to result in a lower quality of personnel. Consequently, it's bad social engineering from the standpoint of the quality of the faculty, assuming that has any effect on the educational product. It does not seem to me to be effective social engineering from the standpoint of providing role models, since it will be quite apparent to students and colleagues (of both groups W and B) that some people are here not because of their merits but because of their race or gender. You simply can't have it both ways. If the groups don't differ in relevant attributes of intellect, work habits, previous education, and the like, there is no necessity for reverse discrimination to attain a "fair" distribution even in terms of a quota concept. If there is a difference, which is what tempts us to engage in reverse discrimination, it follows mathematically that we will be hiring inferior academic products.

9. Tenure

Prof bashing books often emphasize the undesirable effect of academic tenure on professorial work habits, but I am unaware of any quantitative data bearing on this emotionally charged topic. Anecdotally, I am thoroughly convinced that in first-rate departments the granting of tenure does not result in a slackening of scholarly effort. I doubt there is a single member of the Minnesota psychology department who puts in a bare "forty-hour week." However, I have some pretty solid anecdotal evidence about loafers in mediocre departments, and I know of individual instances that could properly be described as scandalous. Years ago, I knew about a department chair in one of our undistinguished units who was a problem drinker (if not a diagnosable alcoholic) who would show up at the office in the middle or late morning, spend an hour reading the student newspaper and signing a few letters, go to lunch at the faculty club, take a nap or play some pool in the faculty club billiard room, then come back to his office and put in another hour, or, at the most, two, seeing a student or two, and engaging in general gossip—not scholarly conversation, which for an academic is legitimate *work*—and then go home between three and four o'clock to have his first martini of the day. All of the faculty and many students were aware of this arrangement, and I presume that the dean had at least an inkling of it, but nothing was done. I know of a department—not at Minnesota but from a trustable source—in which one of the clinical psychologists, who is officially full-time, comes in one day a week to check his mail (and in case there are any students who want to talk to him) and spends the rest of his time in his private practice at home. This is at a state university in financial trouble, and he is defrauding the taxpayers.

The problem here is not that professors *as a group* are lazy or dishonest when compared with businessmen, army officers, or government officials, but that the extreme freedom given to us in the academy to set our own agenda, schedule our classes pretty much when we please, and so on, makes it possible for unethical persons to abuse their autonomy. Sacred Cows I and IV (about research and teaching) are largely responsible for this. The basic idea is that in order to be a productive researcher, you must be allowed to do what, when, and where you please for such things as going to the library, talking to students engaged in your research projects, working at home (that's where I keep my library), running your rats, interviewing your schizophrenic patients, or visiting a faculty member in another department who has expertise related to your work. The freedom from having to punch the clock at nine a.m. and be constantly under supervision of an administrative superior is undoubtedly one of the main attractions of the academic life. It was appealing to me and that's true for most of the academics that I know well. But if somebody is in fact not producing scholarly work, and Sacred Cows I and IV are worshipped at his institution, such a denizen of a mediocre, low-production department gets the benefit of those cows. Although his teaching load is likely to be somewhat higher than it would be if he were a high research producer in a distinguished department, the umbrella of Sacred Cows I and IV still leaves plenty of time not spent preparing for classroom

teaching. *Nobody in the academy who gets around the campus can deny the prof-basher allegation that there are quite a few people who don't put in an honest day's work doing anything that matters much*.

With regard to other wasteful practices discussed in this book, I have expressed definite opinions as to directions for improvement or even definitive solutions. Unfortunately, on the matter of academic tenure, I have not been able to come up with any such. I concede this is partly due to a personal bias. As one who was a teenager during the Great Depression, I have always been conscious of the job insecurity problem. I would list job security along with the intellectual fun and the autonomy as to time and tasks as the big three features of the academic life that led me to decide on being a college professor at age 15 (before I had met one or decided what to profess). I consider job security a desirable feature in a society, not only in the academy, but elsewhere as well. I believe, despite not being a knee-jerk liberal, that employees should, in general, be protected from capricious or whimsical actions of discharge, whether they are in the academy or anywhere else. But every thoughtful person knows that we pay a price for this, of which the difficulty of firing a grossly incompetent civil servant these days is perhaps the most egregious example. The abuse of tenure would surely be somewhat lessened if Sacred Cows I and IV were effectively challenged so that larger amounts of time spent in instruction, preparation for instruction, and office hour conversation with students would be required of those persons who are not affiliated with attached research units. That would cover, in my proposal, at least eighty percent of the faculty in a liberal arts college at a state university.

Some non-academics fail to appreciate the importance of tenure with respect to academic freedom, viewing the latter as merely a rationalization for what is desired on purely egocentric grounds. While I personally value tenure on egocentric grounds, I must emphasize that the importance of tenure in preserving free inquiry is not just an excuse but has social reality. Anyone familiar with the history of the American academy knows that even into the present century some persons were not acceptable, despite their high quality scholarship, if they were known to hold deviant views, especially on the subjects of politics and religion. In the nineteenth century there were universities in which it was impossible to become a college president if you were not an ordained clergyman, and even to be a faculty member, you could not let it be known that you were a socialist or an atheist. In some prestigious "Protestant" schools one could not be a Roman Catholic or a Unitarian! Nor was this kind of thing confined to politics and religion, although those are the highly visible areas that can produce rage and fear in the community. A famous non-ideological example (which kept one Midwest university on the "black list" of the American Association of University Professors for many years) was that of a biochemist, who in his doctoral dissertation research, showed that oleomargarine was as nutritious as butter. Since the state in question was a "dairy state," he was fired. This has nothing to do with communism, or atheism, or pacifism, or homosexuality, or any of the more obviously charged topics, it's just a question of biochemistry with an economic impact on a

strong vested interest which had control of a state institution via the legislature's power of the purse. During the Vietnam War, one of the most brilliant classroom teachers at the University of Minnesota, political scientist Mulford Q. Sibley, came under attack for his opposition to the war, and some state and city officials said publicly that he should be fired for "treason."

These days, the political correctness syndrome has taken over so that it is unacceptable to express traditional or "conservative" opinions rather than liberal or left-wing ones. I, for instance, was publicly attacked in campus speeches and bulletin board posters as a Nazi, a fascist, an advocate of Vietnam genocide (I was in fact strongly opposed to the war in Vietnam, but fanatics don't bother to collect facts before they make accusations) because I signed, along with distinguished scientists in several fields, including Nobel laureates, a statement to the effect that Professor Arthur Jensen of Berkeley ought not to be persecuted and denied research support or access to subjects because of his views on the genetics of intelligence. I am personally aware of several cases of psychologists who were persecuted, and in some cases denied tenure, because they held views on this matter which were disharmonious with the current political correctness of the frenzied egalitarians.

One of the functions of "the clerk,"[31] a category which includes college professors as well as authors, newspaper editors, and other intellectuals, is to examine received views, whether of a majority or minority, and to offer criticism of how society conducts its affairs. If that social critic function is to be performed, immunity from being fired, denied promotion, salary discrimination, or other more subtle forms of unfair treatment for criticizing whatever happens to be the going orthodoxies must be guaranteed. I repeat that I favor people in all walks of life having job security, and for those kinds of occupations in which that is intrinsically unachievable for some economic or political reason, one can discuss the questions of what kind of social insurance should be available and who should pay for it. All I am saying here is that although professors, like everybody else, like their jobs to be secure, it is not a mere rationalization to point out the importance of tenure in performing the social function of the academic. It is well-known that in totalitarian regimes, whether of the right or of the left, one of the first things tyrants do, after controlling the press, is to start controlling the professorate, because tyrants know that it is "these damned intellectuals" who are always causing trouble for the regime.

On the other hand, in supporting continuance of the idea of academic tenure, I find myself unhappy with a current development, which is that non-tenured faculty have a kind of unilateral right to be granted tenure and that a heavy burden of proof is now put on the institution if this is not done. It seems to me that such a thing as academic tenure is entitled to put the burden of proof on the assistant professor, or that at least it should be possible in the formulation of a contract of employment to distinguish between institu-

[31] [Reference to Julien Benda's, *La trahison des clercs* (1927). Meehl read Aldington's (1928) English translation, *Betrayal of the Intellectuals*.—*LJY*]

tions that accept the burden of proof for not tenuring a person and those which are reluctant to assume such a burden. *It is harder to get rid of an unsatisfactory tenured professor than it is to change your occupation, your religion, or your spouse*. When we tenure somebody, we're stuck with that person, absent the most gross abuse on his or her part. Since I do not accept the concept that there is some sort of a unilateral, inherent, indefeasible right to become a college professor, I do not like the current trend which operates administratively and in courts of law as if there were such a right.

10. The Two Worlds

In discussing some proposed curriculum changes in our undergraduate offerings, my colleague P. W. Fox once said, "The world of the professor and the world of the student are two different worlds." The context made clear that by "student" he meant "typical undergraduate student." One besetting sin of intellectuals is the idea that everybody else should be persuaded to lead the life of the mind, to be occupied with concepts, theories, arguments, to prefer "deep" intellectual discussion (at dinner parties it is usually pseudo-deep) rather than talk of sports, the weather, business, politics, pets, or babies. Now, leading the life of the mind—being preoccupied with concepts, theories, arguments, abstract ideas—is in itself rather an extreme statistical aberration in all cultures we know anything about. It may be that it is more of an aberration in American society than in other advanced industrial nations, an observation that students and professors from such places as Germany, Sweden, and Holland have often made to me. It is also anecdotally evident, although I have seen no statistics on this, that the average American academic is more narrowly informed, with less general culture, than the typical European counterpart. Be that as it may, a pedagogical stance regarding course offerings based upon the idealization that the typical undergraduate is passionately interested in ideas and that in his post BA life he will tend to read Plato or Spengler or Freud for the fun of it, is a delusional notion which could be refuted either by a few simple statistics on what most people do in their leisure time or even by anecdotal evidence. Of course, those students who ultimately end up as college professors or research scientists in government or industry will tend, by and large, to have had above average cognitive passions even as undergraduates. When I was a college freshman, I would be arguing, while walking across the campus with a friend, about whether A. J. Ayer had ever delivered on his promise to "demonstrate" the verifiability criterion of empirical meaning in his bombshell book *Language, Truth, and Logic* (1936), so that one could say that Meehl, at age nineteen, was already a "budding academic" who chose his friends on that basis.

But Fox's remark is not about people who ultimately go on to become scientists, professors, poets, or composers, but about the "typical" undergraduate student. There are different reasons for going to college, but everyone knows that the main reason for most students is that it helps you to get a better job when you get through. College serves a "super-kindergarten" function for many students, a transitional, marking-time period

between adolescence and adulthood, in which many developmental changes occur that have little to do with the intellectual content of a formal educational process. It is a form of snobbery (as well as a manifestation of the universal insane impulse that humans seem to have to make other people be like them) to not understand that the great majority of undergraduates have three or four simple reasons for being in college, none of which includes being passionate about ideas or leading the life of the mind. I have made it a practice to eavesdrop on what students are talking about either on our campus or on the campus of the University of St. Thomas (which is the turnaround for my daily five-mile walks). They are usually talking not about ideas, theories, concepts, the life of the mind, but rather about something concrete about a course (e.g., whether the teacher grades on a curve), or their part-time campus employment, or an argument with their parents about their major, and so on. Otherwise, the males are talking about sports and cars and sometimes politics; the females are talking about clothes or personal human relationships. Despite the fact that they may be going from one class to another, in which ideas were being/will be discussed, ideas are simply not something that they value highly or care to spend time about.[32]

The point here is that it is irrational for academics (many of whom were hyper-cathected on intellectual activity already as children, and almost all by the time they got to college) to project this value orientation on everybody who happens to have an IQ high enough to get into college and survive. You can't lead the life of the mind if you have a low IQ, but the great majority of high IQ persons are not desirous of leading it, and I have seen no persuasive argument that they should somehow be induced to do so. What this means for undergraduate teaching is that, although provision should be made for the small subset of "intellectuals" that one runs across in a typical undergraduate class, it is foolish to build a curriculum as if the majority of undergraduate students are going to become professional intellectuals. For example, the notion that every undergraduate student should be compelled to have a number of one-to-one student-faculty tutorials is silly—most of them don't want it and won't get anything out of it, and it is time-consuming for the professor. Even among the students in our undergraduate honors psychology major, only a small minority elect to take the available "directed readings" courses that have this instructional format. To someone who bemoans this fact, my standard answer is, "What's so terrible about that? Why *should* they?" Some say it is "elitist" of me to take this position. While I am not terrified by the buzzword 'elitist' (being avowedly elitist in some respects and anti-elitist in others), I claim the term is

---

[32] The library at the University of St. Thomas has an excellent collection on philosophy of science, and I have taken to looking at their recent acquisitions at that area, rather than at the University of Minnesota library. The student body at St. Thomas consists largely of upwardly mobile middle- or lower-upperclass Irish, Polish, and (in recent years) Italian students majoring in business and many of whom with high IQs intend to go on to get the MBA degree offered there—but they are learning economics in order to become successful managers and not as part of the "intellectual life." The result is that books on philosophy of science are almost always available on the shelf, which is handy for me.

undeserved. It is not elitist, in any wicked sense, to face psychological and social facts about individual differences in talent, drive, values, and interests. I would insist that one of the most objectionable forms of elitism is "knowing" what's good for others who are unlike oneself and forcing them to "*be more like me*." I am not suggesting that we should eliminate such intensive one-on-one instruction at the senior college level if we can free up faculty time by effectively challenging the seven sacred cows. The saving would be sufficiently great so that the subset of faculty who are interested and competent at this type of tutorial (many academics are not suited for it) would easily be able to meet the needs of the small minority of juniors and seniors majoring in a department who desire this kind of thing. It is foolish to conduct oneself administratively and pedagogically on the basis of assumptions concerning people's interests and wants which are factually erroneous, and failure to recognize Fox's two worlds involves this sort of mistake.

11. Administrative Bloat

A colleague, on reading a draft of this chapter, complained that I had failed to discuss what he considers to be one of the most wasteful features of academia, namely, what he calls "administrative bloat," meaning the proliferation of deans, associate deans, assistant deans, provosts, vice presidents, and so on. I had not intended to discuss this (despite having, of course, thought of it as a source of economic waste), because I am somewhat suspicious of the negative generalizations often made by professors concerning administrators. I hope the book I have written, including this chapter, is scrupulously fair, even when strong positions are adopted against the grain. Having chaired a department for six years in one of the largest departments in the college during a period of transition with many changes and conflicts, and also having served on the executive committees controlling two other departments, I am sympathetic to the problems faced by competent and honorable administrators from posts of assistant chair on up. Ex-chairpersons are likely to display somewhat more understanding, caution, and sympathy in discussing the hard choices made by them and higher administrators. As I am sure is true of executive posts in the military, government, private business sector, churches, or whatever, to have a full appreciation of these matters it is almost a necessity to have "sat in that chair yourself." Nevertheless, it occurs to me that my reluctance to discuss administrative bloat could reflect some bias or timidity on my part (although, being retired, I can't imagine what bad consequences a spiteful chairperson or dean could work upon me). It is an easily checkable fact, by simply counting titles, that the number of faculty in administrative posts has expanded greatly from the period immediately after World War II, when I joined the faculty as a fresh-baked PhD at the rank of assistant professor. Whether the number of persons with administrative titles (such as dean, assistant or associate dean) has increased proportionately more than the total number of faculty, I do not know, but it seems likely, because it has gone up five- or six-fold and the faculty has not increased by that amount. Not having functioned at that administrative level I have no trustworthy

way of judging, even anecdotally, concerning the increased number of unavoidable administrative tasks. So, I content myself here with only two kinds of "evidence" for administrative bloat.

First, it is not in dispute among economists and political scientists that it is in the nature of all bureaucracies to expand and to fill up available time (e.g., special issue of *Academe*, 1991: Strohm, p. 11; Bergmann, p. 12-16; Halfond, pp. 17-19; Anderson, pp. 20-24; Downs, 1967). Parkinson's Law is not a joke, but a fact of social reality, and there is no reason to think that academia is different from the military or the post office (or, for that matter, General Motors) in this respect. I conclude with some confidence that it is likely true of academic administration also. Arguably, it may be even more tempting in that context because academia does not, as the Republicans used to say, have to meet a profit-and-loss accounting. We are minimally responsible to the "consumer" (student, taxpayer), so that there is no adequate quality control over efficiency.

Secondly, I can look back at my own administrative functions as chair, and also when I was director of our clinical psychology training program, and reflect on how much time it took me per week to carry out these functions (which I conducted in a manner satisfactory to all parties, so far as I know). When I was chair of the psychology department in the early 1950s, I had *only one or two brief telephone conversations annually* with the Arts College dean, and *one physical meeting* (at budget time), which lasted between five and ten minutes. Today, our department chair and associate chair, as well as directors of intradepartmental training programs, spend several hours monthly meeting with such higher administrators on a variety of topics. I do not know how many of these time consuming conversations are necessary, and if not, why it is that they are required to take place; I merely record an extremely clear-cut, dramatic increase in time spent. When I was director and then co-director of the clinical psychology training program, when the number of PhD candidates was at least 80% as large as it is currently (so that can't explain the change), I had, to the best of my recollection, a single conversation with a higher administrator during a six-year period! (He wanted some special treatment accorded to an unsuccessful applicant for the clinical program whose father was a member of the State Legislature. I told him in no uncertain terms that no such special consideration would be given and that I trusted he would not ever call me again about such a matter, and he never did.) When another professor and I jointly ran the program, we had access to a fraction of the two departmental secretaries' time; no special secretary was assigned to the clinical training program. He and I met for one-half day in the late summer to riffle through the 3×5 course cards on which the lock-step curriculum was arranged. We made a few telephone calls to those other departments which offered courses in the required clinical curriculum, verifying that those courses would still be offered the coming year and at the usual times. If any times had been changed, we would rearrange our lock-step program. If that turned out to be impossible, we would allot to one another the telephone calls to the half-dozen faculty in the clinical program to see whether they would accept the best suggestion we could come up with (e.g., substituting

one course, striking it from the offerings, moving it to a different year). I am asserting, literally, that "running the clinical program" consisted of a half-day's meeting by the two of us, plus one annual meeting of the seven or eight main and adjunct clinical faculty. Some years, even that meeting did not take place, because there was no agenda to talk about. But I should mention that none of the members of the clinical faculty were fond of meetings and they all had plenty of other things to do, such as treating patients, writing articles, and supervising students. A recent director of our current clinical program informs me of how many hours per week, tabulated from his calendar, he spent administering the program. As I predicted, only a small fraction of his time was spent conversing with students; almost all of it was conversing with other clinical faculty or with administrators from the assistant chair on up to the dean. It is hard for me to believe that the number or complexity of administrative decisions about a matter like this could conceivably have increased to such a great extent. So I am forced to conjecture with my colleague that administrative bloat is partly responsible. So far as I am aware, no bureaucracy has ever cured itself of this malignant disease, and I have no useful suggestions to make in how to fix it. I suspect that the only way to do the job would be for the legislature or the Regents to adopt the unpopular "meat-axe" approach, simply saying that X percent of funds will be slashed throughout, including administrative salary money. Short of that, since the allocation of funds is siphoned through the decision process of higher administrators, there is no reason to anticipate that they will liquidate their own positions, even when economic pressures are severe. Of course, official statements and shadow-boxing motions will be made along those lines, but one can be quite confident that the net effect on administrative bloat will be slight compared to the reduction in teaching personnel, teaching assistants, secretarial help, and the like.

Additional References for Administrative Bloat/Bureaucracy, not discussed in text:

Albrow, M. (1970). *Bureaucracy*. New York: Praeger.

Krislov, S., & Rosenbloom, D. (1981). *Representative democracy and the American political system*. New York: Praeger.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Chicago, IL: Aldine-Atherton.

Niskanen, W. A. (1973). *Bureaucracy—servant or master*? London, Eng.: Institute for Economic Affairs.

Chapter X

Making the transition

Challenging these Seven Sacred Cows would doubtless produce administrative and morale problems. As with every social change, one has to ask what has been lost, and what new problems will have arisen in the effort to solve existing problems. But given the current academic situation, it seems to me that administrative courage is appropriate in challenging these Seven Sacred Cows to see what would happen. Agreed, it is hard to imagine situations that would compel a faculty and administration to challenge these Sacred Cows in earnest. However, I do recall that during the Great Depression there were colleges that paid professors in scrip, and even with farm produce. Some institutions of higher learning have gone bankrupt and closed their doors in recent years. One can conceive of political and economic factors that would lead some state legislatures to adopt a meat-axe approach to cost cutting, and "cold wars" between legislatures and state universities are not unknown in our own day.

Assuming such compelling circumstances, how might an academic administration proceed to challenge the Sacred Cows, while avoiding unacceptable levels of conflict, morale damage, or injustice? Unless a solid majority (say, 80–90% consensus) of faculty were convinced that catastrophe impends, it would be hopeless. Given a solid majority, however reluctant, a clear-thinking and determined administration would, I believe, have a fighting chance. As a first step, on the cognitive level, an educational job about the Sacred Cows is imperative, with plenty of opportunity for free discussion. Faculty should come to sense that, while the dollar crunch has pressured them to *examine* the Cows (which most of them have never done), the *result* of such critical examination can stand on its intellectual merits. The ideal educational outcome would be something like, "Well, if I *had* examined these Cows before, in a purely detached spirit, I would have realized their doubtful status. But I had no occasion to do so until my job was threatened. The dollar crunch merely provided the motivation to take a close look. Having looked, I now believe they are largely incorrect. Given our desperate economic situation, it seems rational to act accordingly."

To make it simple, assume the legislature has said loud and clear that they will appropriate only inflation-correction increments for the foreseeable future. They are apprised of our Sacred Cow challenge and sympathetic to it, so they have "promised" *not to react to faculty reductions by correspondingly lowered appropriations*. Obviously this unusual condition requires that they really accept the innovative approach, sincerely desiring to maintain a good university by gradually raising salaries while gradually also reducing tax dollars and student fees. We have offered to do something that they had assumed was undoable. While they must put *some* reliance on our honesty as to scholarly

integrity, they can rely mainly on the fact that we recognize the hard economic and political facts and want to keep our jobs. Given this solid legislative backing, how do the deans get it rolling?

Administrator determination plus high faculty consensus are necessary but not sufficient conditions for movement, as anyone knows who has studied the growth and decline of large social institutions. "A good policy" usually achieves nothing against inertia, so *to make anything happen* persons at the grass roots level must have a definite and strong incentive to initiate concrete actions. The incentive is a sizeable raise in pay (doubling!) in the near future; the concrete actions have been specified in the chapters above. The planning takes place at department level. The dean says (not as fuzzy "policy," "goals," or "guidelines," but as a specific budgetary *dollars-and-cents promise*), "Our college will receive somewhat reduced total real dollars annually from now on. If you innovate by a suitable mix of Sacred Cows challenges, thereby reducing your faculty, a portion of the saved dollars will be your department's, to divvy up as you see fit. Only the grossest abuses of instructional propriety (e.g., a psychology department offering no courses in learning theory or perception; eliminating all live lecture or discussion classes at the undergraduate level) will be countermanded by this office. Now it's up to you." Given the conditions as described, I think this would work. *But persons at all levels must mean what they say*.

Selecting departments to have research units

The imperfection of *any* indicator or any composite of multiple indicators of significant contribution to knowledge (including the subjective judgment of a department chair or a dean!), does not bother me much for a simple, obvious reason: *We are always dealing with this in one way or another, whether we like it or not*. The only question to ask, if one approaches the subject rationally, is: How should we do it? Since my proposal is to use any such evaluations of scholarly productivity and impact as a means to classify departments into those that have research units attached and those that are pure teaching units, the element of unreliability that may be disturbingly large if an index is blindly and mechanically applied to a particular individual (e.g., the Pfaffman case) will tend to get washed out when we are sorting not individuals, but departments. If we look at multiple criteria of visibility, with emphasis but not sole reliance on citation counts, there is no appreciable probability of somehow misclassifying a first-rate department by giving research unit preference to a mediocre one. It won't matter how you conduct the evaluation if you employ criteria having any rational or empirical justification. That within a first-rate department there are still marked individual differences among faculty as to their productivity and impact, goes without saying. But that shouldn't bother anybody unless one is greatly distressed by the ubiquitous current situation. I would be prepared to argue on the total body of evidence—including some *SCI* tallies within departments that I prefer, for reasons of ethics and comity, not to detail—that the least productive and least

impactful members of Minnesota's psychology department would certainly be rated higher than the most productive and impactful members of a department nationally rated as mediocre.

If all scientists published the same number of papers and each was on the average equally impactful (as measured by citation rate combined with other criteria of evidence), then cessation of publication by, say, 90% of academics who now publish at least one paper, whatever its impact, would get rid of a large amount of wasted time and wood pulp used for paper. However, whatever small fraction of papers are impactful would suffer that 90% loss as well, which of course, would be bad for the advancement of scientific knowledge. Fortunately, nothing remotely like this is the statistical situation. Both the distribution of number of publications and the distribution of impact among publications over scientists are extremely skew, and quality and quantity are highly correlated. This means that one could even truncate the distribution of scientists by publication quantity and not suffer much loss; but obviously truncating the distribution in terms of impact would be preferable. Statistics available to derive loss in impactful papers from the mathematics of the two distributions involved are not, so far as I know, available, and in any case, the mathematics here is beyond my competence to handle any such derivation.

There is a study that goes to impact directly rather than relying on such intra-statistical probability inferences. This is the study by Canadian psychologist Myers (1970) which, in addition to the variety of relationships examined with scrupulous care, has the advantage in that it deals with psychology, a behavioral science that lies midway between advanced physical sciences (like physics and chemistry) and non-sciences as regards citation rates, cumulative character of knowledge, clear criteria of advanced problem solving, and proof of conjectures. (See Sacred Cow IV, p. 57, for discussion of the intrinsic gain of research, which differs typically in a *qualitative* way between sciences and other scholarly disciplines.) For a busy reader who is skeptical as to the value and especially the "fairness" of citation counts as the *best available single indicator* of one's scientific contributions, I recommend a reading of Myers' classic paper. If you don't find that paper convincing, you can skip any of the related references, because they will surely not persuade you either. Myers employs multiple non-citation criteria of evidence and impact (e.g., reception of various prizes and awards, opinion polls as to contribution), and he also examines carefully and fair-mindedly some of the sources of error that might apply to individual cases. In discussing these, he makes the same assumption that I make in advocating paying heavy attention to citation counts, that the faculty group or individual administrator is not feeble-minded or devoid of the desire to do justice. For example, psychologist Carl Pfaffman was the world's authority on one little-studied sensory modality olfaction. Noncitation criteria appropriately classify him as an eminent contributor, but he does not fall among the top 1% (30 names) or the top 2% (61 names) in number of articles in a 5-year span of major journals that cite him. Everybody knows that it is because only a small number of psychologists and physiologists are interested in smell as a sensory modality. If there are very few scientists writing about that subject,

they can't generate many cites to Pfaffman, despite the fact that nobody who does write about it would *ever* fail to cite his definitive world-renowned work.

The agreement of the multiple criteria with citation rates is extremely impressive and came as somewhat of a surprise even to me, one already favorably disposed to the idea. The most important single statistic that emerges from Myers' study is the proportion of the whole group of cited authors that can be classified as "visible" by Myers' highly tolerant criterion of visibility. Any such criterion is, of course, somewhat arbitrary but is not capricious. The main point for our purposes is that Myers has surely leaned over backward in his stipulation. He decided that a psychologist could count as "visible" if his or her work was cited at least six times in a six-year period, that is, an average of one citation per year. In the collection of 14 mainstream journals in both soft and hard areas of psychology, there were 143,260 citations and they referred to a grand total of 48,903 different authors. The majority of these many authors were cited only once, and many more were cited only a few times. Having set the threshold of visibility of six citations in the six-year period, he finds that there are 300,056 authors at or above this citation frequency, representing the top 6% of the total population of authors cited (p. 1042). So, despite the mathematical complexities involved in relating proportion of articles to proportion of impactful articles (which my colleagues, on the average, judge to be 1 in 20), it turns out that same proportion of cited authors are impactful by this tolerant criterion. (I made a "safe lower bound" use of this 6% impactful proportion in the grand equation for faculty size reduction.)

Unwarily falling victim to the "cutting-butter-with-a-razor" fallacy can generate an obsessive rumination about the terrible danger of making a "mistake" in selecting the departments in the college as research departments. I discern four errors that jointly contribute to this unwarranted decision anxiety. The mistaken notions are: (1) There are no clear cases of appropriateness for having a research unit attached to a particular department, that most of the "research/non-research" decisions will be difficult ones and open to challenge by informed reasonable persons; (2) close calls between departments marginally appropriate for a "research" designation could constitute, objectively speaking, *a mistake*, an undefensible judgment call; (3) if such a mistake (in the eyes of Omniscient Jones) were made it would have dire consequences for the advancement of knowledge; and (4) in a close judgment call, a non-selected department will have suffered some sort of injustice. All four of these erroneous notions are easily refuted.

There Are Clear Cases

Suppose the administration has calculated that approximately a half-dozen departments in the college are to be designated as research departments having research units attached, light teaching loads, doctoral programs, and the usual requirement of publication for tenure, promotions, and raises. In the liberal arts college of my university, there would be absolutely no question to any rational mind as to three departments that *must* be

selected. Nobody in an unselected department could make even a feeble case to the contrary. Similarly, the great majority of departments in the college would not be considered candidates for selection, and most members of their faculties, whatever they might be prepared to say publicly, would not seriously entertain the idea that they are in the running.

I find that most academics who have not studied the research on communication networks and literature citation rates[33] grossly underestimate the size of the differences between first-class departments and those that would be rated average, mediocre, or substandard. For example, I asked some colleagues in psychology what they would guess to be the ratios of literature citations of a top producer in an "excellent" social science department, compared with a top producer in a "strong" one, and in turn with one in a "mediocre" department. People typically say, "Oh, maybe twice as many citations for the excellent over the others." (A former department chair guessed "around 10 : 1" but added, "That's only because you once told me that the differences are surprisingly large; otherwise I would have said more like 2 : 1.") I computed citation rates (*Social Science Citations*, *Science Citation*, *Humanities Citations*) for a cumulative 5-year period[34] for Minnesota's psychology department (nationally rated "excellent"), for a social science department nationally rated "strong," and for a mediocre department. Correcting for the total number of academics and averaging the three most cited members in each department, the "stars" in that nationally-rated "excellent" psychology department are cited twice as often as those in a "strong" department, and 12 times as often as those in a mediocre department. A social science department ranked around 50th nationally will show 24 citations annually of its most visible faculty members, whereas in Minnesota's psychology department there are two Regents' Professors and one former chair who are currently cited 295 times annually. The distribution of citation rates for a random sample of faculty in the Minnesota liberal arts college (90% tenured) is extremely skew, as is to be expected from studies of research productivity and impact. The graph resembles that of a Poisson distribution, suggesting that the preferable "locator" (representative, typical) value is the median [= 14] rather than the mean [= 58]. The modal number of citations is *zero* over the college faculty. Stellar members of excellent or strong departments have citation counts ten times the average, being from 2.9 to 6 standard deviations out. The only tenured faculty in an excellent department with counts as low as the college median were a couple hired for reasons other than excellence. The unrealism of Sacred Cow IV in the light of statistics is strikingly illustrated by a letter to the editor of a Twin Cities newspaper from a University of Minnesota professor (I think a chemist?), claiming that all tenured professors must have scholarly eminence. It is hard for me to think that the

---

[33] Research on individual productivity (measured by publications) and impact (measured by citations) are discussed in the chapter on Sacred Cow IV. Here we are evaluating whole departments, and that presupposes the evaluation of individuals.

[34] [*Perhaps late 1980s, but more probably early 1990s.—LJY*]

writer sincerely believes such a preposterous statement; it is plainly false without contemplating statistics.

In 1995 the Minnesota psychology department was ranked by the National Research Council seventh in the country for scholarly quality and fourth in quality of teaching PhD candidates, the peer departments being Stanford, Michigan, Yale, UCLA, Illinois, and Harvard. Numerous other rankings by various government and private agencies have consistently shown the same thing for two generations. Considering subdomains, in the three fields of clinical, counseling, and industrial psychology, we are tied for first place nationally. If such subfield ratings existed for the fields of behavior genetics and social psychology we would unquestionably be in the top three or four. Several years ago, someone tallied the total number of scholarly papers published by all faculty in the college, and the psychology department value was "off the curve," something like 4.5 standard deviations out. Considering all research money that comes from outside sources (not from the Minnesota Legislature but from private industry, the federal government, and private research-subsidizing institutions, e.g., Carnegie), psychology ranks first in the college, an astounding 52% of all such outside money comes to psychology. Not to indulge in self-gratuatory overkill, I do need to drive home the point that it hardly matters what criteria—objective or subjective, measured or judgmental, inside or outside review committees—psychology comes out at or near the top of the heap. There are other features that do not emerge as statistics, but which, if statisticized, would be equally dramatic. A minor example: the intradepartmental monthly newsletter invariably records from two to a half-dozen lectures delivered at other institutions by members of the psychology faculty. In mediocre departments, a whole year can pass without a single faculty member being invited to give a talk elsewhere or serve on a panel. If the administration were selecting a half-dozen departments in the college to have research units attached, it would be inconceivable for psychology not to be in the list, and I could name two or three other departments at Minnesota about which an equally strong statement would be justified. By contrast, 70% or 80% of departments in the college could not conceivably be chosen on the basis of their sheer productivity, their impact, or their national reputation.

What It Means to Make a Mistake

After the clearly appropriate selections have been made, decisions among additional departments to have research units attached might be so close that one might just as well flip a coin. I do not understand in such cases how one would defend the idea of a *mistake*. It would be a mistake to exclude a nationally-ranked excellent department, as it would be a mistake to include a social science department ranked 45th in the country and with low research productivity and impact (citations), not attracting first rate PhD candidates from around the world, etc. Suppose departments A and B are so close that by one set of national ratings A is ahead of B; but in terms of average number of publications per faculty member, they are equal; and in terms of literature citations B is a little bit ahead of A. What would be the meaning of "making a mistake" as to which we select?

The reason one could be mistaken about the clear cases, but not "mistaken" (in any explainable way) with the knife-edge cases, is that imposing dichotomous decisions on quantitative dimensions, or on a subjectively weighted composite of several quantitative dimensions, is not "arbitrary" at the extremes, but *is* "arbitrary" when two are very close together. Karl Marx was bearded, Herbert Hoover was not bearded, and it would be a clear unarguable mistake to characterize them in reverse. This remains true as to Marx and Hoover even though a bum, after a week-long binge, presents a marginal case, so that one would not properly describe him as clean-shaven. To quote Edmund Burke, "Despite the existence of twilight, high noon and midnight are nevertheless tolerably distinguishable."

On the unlikely assumption that no matter what considerations, quantitative and semi-quantitative, beyond the first-preference indicators of productivity (impact and rated national scholarly prestige), when taken together with these, shows a preponderance in favor of Department X over Department Y—a balance I consider to be highly unlikely given the multiple available criteria—the decision-makers couldn't flip a coin or enter a random number table. (That straightforward procedure, when the evidence is on the knife edge, would shock many people, mainly because they are unaware of how many allegedly rational choices have negligible reliability and amount to flipping a coin.)

The assumption that a pair of such departments would be equally eager to have research units is unwarranted. If one department had approximately an even split as to preference and the other one showed a very strong preponderance in one direction or another, that would now become, in the absence of other compelling evidence, a rational basis for selection.

In the period of transition from the old system to the new one, it would be desirable to offer powerful economic incentives for departments to adopt more efficient instructional methods as one main component of my savings equation, but also to opt for being a non-research department, which would allow tenured faculty to increase their teaching loads while coming out from under the publication pressure. Since almost two-thirds of college professors have stated in opinion polls that they wish they were not under such pressure, it is by no means obvious that a department somewhat marginal in its production and in its impact would still insist upon having a research unit attached.

Finally, since the bargaining position of the university's administration *vis-à-vis* the legislature would be extremely powerful because some fraction of the 75% savings is going to be allotted to reduction of tax support, it would be worthwhile to ask the legislature for a little more money during the transition period (although that actually means a little less rapid savings) so as to classify both as research departments, thereby attenuating the component of the grand savings equation attributable to increased teaching loads. I realize that approaching the problem along these lines presupposes a high order of rationality among professors, administrators, and legislators, and that this is an excessively optimistic assumption to make. I include it because under sufficiently stringent

economic circumstances, people can sometimes be so driven to the wall that they can depart from ideology in the interest of money.

Close-Call "Mistakes" Are Not Catastrophes

Even if we were able to give some empirical meaning to the notion of a "mistake" in choosing between two very close departments—say, some such concept as what their net scholarly impact over the next 25 years will turn out to be—that mistake does not constitute a social catastrophe.

Potential Loss of Teacher Supply

Inasmuch as non-research departments would presumably not be accredited to turn out PhDs, if my reformation were conducted on a large scale covering most state universities, would this result in a serious shortage in college faculty after a few years? Although it might result in some such temporary shortage if not foreseen by most schools—an unlikely blindness, given the constant recourse to forecasting statistics in academia—we need not fear that it will be big enough or hard enough to cure or to require rejecting my proposals. Assume perfect fungibility and no large-scale selectivity with respect to the social value of the competing disciplines. If we accept a figure of around 25-30% as the residual faculty sizes (including both non-research and research departments nationally) that would mean that a reduction of 70–75% in teacher production is tolerable. In my numerical illustration (in Chapter VIII) of the total savings formula, I assumed that roughly 80% of faculty would be in non-research departments, but of course, not in the same fields, and they would be at different schools. So we would have a mere 5% undersupply, which one can safely rely on the laws of the market to adjust for in the long run. Given the current opinion that there is an excess of PhDs having a hard time finding academic jobs these days (e.g. Begley, 1991; Hartle & Galloway, 1996), I am not inclined to take this worry seriously. Furthermore, the several savings not included in my master formula (listed in Chapter IX) can easily take up a slack of 5%.

There is a more fundamental question that deserves raising here, although it involves another sacred cow, which is even holier than some of those I have been criticizing. I shall not develop it at length, but merely mention it for the benefit of readers who are not too devout about the academy. There is no strong theoretical reason, and there is no empirical evidence, to suggest that the irksome chore of producing a PhD thesis is necessary for one to function as a college teacher at the undergraduate level. For many years, an M.A. degree was acceptable, and in many small colleges it still is. The basic factual issue here is whether conducting the sort of research most candidates do for the PhD can be expected to yield what psychologists call "transfer of training" when it comes to *critically reading research* and *expounding it to undergraduates* in an interesting and competent way. Nobody has ever proved that it does, and I, after many years of observa-

tion, don't believe it does. (Unfortunately, when research is non-existent, one is forced to rely on anecdotal impressions—plus, one hopes, a dash of common sense.) I have argued in another connection that intensive seminar work in *surveying and evaluating research on a variety of substantive topics* is a superior preparation for someone to be a competent "research consumer" than is spending a year or two on a narrowly defined topic (Meehl, 1971). Most undergraduate instruction is not aimed at producing high-level research sophistication. Instructors should know how to avoid misleading students as to what the current research on controversial topics says, and why it is worth knowing about if one is not a professional. Probably 80% or 90% of what is taught in undergraduate psychology courses is hardly controversial in any interesting scholarly sense. Finally, a non-research department, most of whose faculty lack the PhD (or hold a doctorate that did not require a major individual research project), could hire a traditional research-qualified PhD for the specific purpose of covering senior seminars. We have one such seminar in our department for the honors students, in which current research at the "cutting edge" (in the social sciences, it's usually rather a blunt edge, but pass that) could be entrusted to the possessor of the sacred doctorate. I confess that I find the whole mythology here so unplausible that it's hard for me to spend time discussing it.

Potential Loss of Growth in Knowledge

It is plausible to conjecture that my conservative selection of 20% of faculty to be in departments having research units attached (where publication is required for tenure, promotion, and raises, with the other 80% being free of such obligation and not encouraged—although I emphasize, not forbidden—to write books and papers) will have a consequent reduction in scholarly publication. It would be foolish of me to try to argue from the armchair that *no* such reduction would occur, but fortunately I need not make such an argument to defend my position. Keeping in mind our "don't cut butter with a razor" principle, we cannot insist that administrative decisions and institutional long-range policies be predicated on absolutely precise numerical forecasts, because no such thing can be found, and no college administrator, in making decisions under the present system, requires such illusory precision. Starting with the overarching premise that we are going to save money by some such combination of policies and institutional arrangements as I am setting forth, we want to set reasonable bounds on possible undesired consequences of the proposed reformation. That's the best that could be had, and if the bounds are set with care, it is quite sufficient for our purposes. The rational question we face is not whether it is probable that there will be some reduction in publication, but instead, what is a safe upper bound on the size of such a reduction; and, speaking qualitatively, what is the serious import of such a predicted extreme reduction on the generation of valuable knowledge? Nobody with the slightest acquaintance with the academy would be prepared to equate a crude "*p*% reduction in number of articles" with "*p*% loss in increment to valuable knowledge." I am taking "valuable knowledge" throughout this book as being either an *intellectually prized improvement in our theore-*

tical comprehension of the world, or a *significant step in the alleviation of some important social problem*—whatever this last may be, as long as our society cares about it.

In setting bounds on a plausibly expectable loss in knowledge growth, I am going to lean over backwards to a degree I consider absurd and show that—even on preposterous, surely false numerical assumptions—the expectable valuable knowledge loss is very small. Starting with data on publication in the physical sciences, I am going to pretend that (1) the citation rate for social sciences, humanities, literature, arts, etc., is the same as that in physics and chemistry; (2) that on the average (to turn it into proportions, I will identify this with the median) an article that is cited at all in these other fields makes a valuable knowledge contribution; (3) that none of the faculty in departments lacking a research unit ever publish anything. What do the numbers look like on these hyper-"pessimistic" assumptions, all unfavorable to my position?

Since the most productive 20% in productivity are responsible for over 75% of the articles published (Price, 1986, p. 41), if the bottom 80% of a university faculty were publishing nothing instead of what they are now publishing, we would suffer a 25% loss in publications. If half of these are cited at least once, we have a 12% loss in cited publications, which in turn translates into a 6% loss in valuable publications. So this small decrement, given the above absurdly unrealistic statistical assumptions, is the worst we have to worry about. Further, even this 6% is an inflated estimate of loss in knowledge growth, because it is computed as if all universities and colleges have chosen the same subset of departments to have research units. The reasoning is that if the psychology department at Minnesota had no research unit attached, and as a result a professor did not publish a paper discovering fact X or formula Y or inventing theory Z (although he would have if he had been in a research department), then a professor at Michigan, which does have a research psychology department, would never do so either. The data we have on discovery "multiples of the sciences" show that is plainly false. Furthermore, the more important kinds of discoveries are, of course, on the average much more likely to be made independently by different people in different places, although it turns out frequently at almost the same time.[35] Besides that, the important contributions are of course less likely to be made by faculty in lower prestige departments, since it has been known for many years, going back to the early work of Lehman, that there is a definite correlation—contrary to the usual cliché—between the quantity of work that a professor publishes and its quality. Given the unrealism of the above assumptions and their neglect of this important point, I do not hesitate to express my personal hunch that what I have called a safe upper bound on the loss could, if we consider expectable values rather than the "worst" scenario, be 1% or less. [Relevant references for this section include Begley, 1991; Burton & Kebler, 1960; Hamilton, 1990; Myers, 1970.]

---

[35] [*See e.g., Merton* (*1961), or online* https://en.wikipedia.org/wiki/List_of_multiple_discoveries —*LJY*]

Injustices and Morale

As to injustice, I have already discussed the requirement that, in a phase of transition to my reformed institution, presently tenured faculty are protected under the explicit or implicit terms of their employment, so that we are not doubling their teaching loads when we classify their department as a non-research unit. Obviously no injustice is done to new hires, since they take the job with the clear understanding of how this academy works with regard to these matters.

A special problem arises here as to the knowledge generating units. I think the most serious problem would be the identification of academic personnel who are appropriate for each of the two tracks: the "new-knowledge producer" and the "knowledge transmitter." It would be difficult at big universities, whose faculties would include a subset (although a small minority) who are nationally-eminent knowledge producers, to avoid invidious comparisons and feelings of second class citizenship by the others. That problem would have to be solved somehow, and I do not have any clever solution to offer. Suppose the dean selects a few departments (around one in 5 or 10 would be reasonable) that are not only *allowed* but *required* to continue as research producing units. This does not preclude such a research department from hiring a small number of pure knowledge-transmitters, if the understanding with these instructors is clear at all levels at the time of hiring. We may safely assume that the ethos of these selected departments will be such as to work against any temptation to abuse this "exceptions" privilege. If anything, these high prestige research units will tend to resist keeping even a few such special non-publishing faculty. All departments now have a strong incentive to innovate instruction and start reducing faculty size. Keep in mind that almost $^{2}/_{3}$ of college professors express a preference not to be under the pressure to publish, so we here do less violence to an inwardly *accepted* tradition than might have been supposed. The combination of dropping courses, changing lecture format, and shifting to more efficient instructional modes results in gradual reduction in faculty. This saving holds for both research and non-research departments.

The next big step is the increased teaching load for non-research faculty. If we achieve sizeable salary increases in this way, over all departments, there is an appearance of injustice to the non-research faculty, because a portion of dollars they save by doing more course work goes to augment salaries of researchers, but the latter now teach no more hours than before. This will generate resentment and morale problems unless it is clearly understood to be a plausible misformulation, which of course it is. Nobody is now doing *more work*. If they were all working a 40-hour week before, they will still be putting in their 40 hours. The only change is in the *kind* of work the pure teaching faculty now do, whereas the research-productive faculty are distributing their hours as before. Research and PhD supervision are more labor-intensive than undergraduate instruction, especially large class and non-lecture instruction. Hence it is rational to allocate a fraction

of dollars saved by pure teaching units to the old style double purpose units. Nobody is getting gypped or being a free-loader when we look at it rationally.

What would be the administrative strategy in the transition period? (*Tactics* I do not discuss, being neither a cost accountant or an educational economist, and because departments will vary widely in constraints imposed by age and rank distributions, class sizes, current ratio of graduate to undergraduate courses, and other idiosyncratic factors.) The general principle is simple and obvious: *We lower cost by reducing faculty*. Improved efficiency of instruction (e.g., Keller Method) saves no dollars unless Professor X's freed up hours are shifted to other instructional tasks formerly done by Professor Y who is no longer on the payroll. I use *faculty attrition* to cover death, retirement, resignation, reduction to part-time status, and termination of non-tenured faculty. The big point of the whole business is that faculty attrition saves money, and the money saved can be allotted to lowering student fees, reducing tax support, and raising faculty salaries.

The transition strategy must provide incentives to change, and avoid injustice or appearance thereof. (As usual, I ignore the small percentage of litigious paranoids.) Ideally, the reformed department will have voted to be classified as a pure teaching unit, assuming it was considered borderline as to research productivity and given the option. But even if its classification as a non-research unit department was involuntary, I would expect reasonable cooperation, given the incentives of alleviated publish-or-perish pressure and of substantial salary increases. We may confidently anticipate that faculty strongly opposed to the reform on self-concept and ideological grounds will contribute heavily to the attrition, whether tenured or not. For example, an eminent tenured scholar in a mediocre department (I have known several such) will let it be known that he is "available" and will begin to get attractive job offers elsewhere, one of which he will accept because we do not attempt to compete in such retention cases.

Flexibility and a rational time perspective are imperative. The group effort in reform benefits everyone financially *in the long run*, and in the short run we reward merit, now defined in terms of more efficient instruction *and* increased instructional hours. Exactly how this is managed is a delicate matter, and some short-term inequities are almost inevitable despite the best intentions and high rationality. All I can say is that the same is true of the conventional "research institution" arrangements. To think that the weights informally assigned to classroom teaching hours, number of PhD advisees, amount of Federal grant money, committee work, community service, quantity versus quality of publication, citation rates, national visibility, etc., are presently free of unfairness, one must be blind, deaf, and dumb.

Having spoken in such general terms, let me concretize by an invented example. Assistant Professor X is terminated, or goes to another university because she is ideologically opposed to the reforms and prefers the publish-or-perish pressure to a heavier teaching load. We do not hire a replacement, so we save $K,000. What happens

to the four 3-credit courses she taught? (This is the official teaching load guideline in Minnesota's College of Liberal Arts, rather feebly enforced.) Course #1 was concocted to fit her special interests—a common practice—and we simply drop it from our offerings. Course #2 is taken over by Professor Y, who has shifted to the Keller Method in one of his courses and offers another only alternate years, so his total load undergoes only a slight net increase from picking up #2. He is rewarded with a merit increase accordingly. Course #3 is picked up by a part-time practitioner, at around 40% of what it cost when X did it, so *part* of the 60% saved there goes to Professor Y's slight merit increase. Assistant Professor Z, a zealous advocate of the reform, takes over #4, a sizable load increment, receiving a merit increase bigger than that given to Professor Y. Principle: Allot more of the freed-up dollars to those who contribute more to the pot—this is what "merit" means in a pure teaching department. Whatever portion of the $K,000 remains unassigned reverts to the dean (i.e., taxpayer and students).

The imagined sequence of reductions assumes only small, negotiated, short-term teaching load increases. My own view as to a full-scale head-on challenge of Sacred Cows I and IV jointly would make the process simpler and easier. When a professor who has been considered both a teacher and researcher is replaced by a pure knowledge-transmitter, the latter's teaching load can be set two or more times as high as was the former's. Systematic adoption of the knowledge transmitter/producer dichotomy could, over a few years' time, greatly reduce faculty head-count with considerably less adjustment in course offerings than I have outlined.

For the transition, there is no good reason why special "deals" cannot be made about salary with appropriate individuals. It is against the tradition, but there is nothing immoral or illegal about it, and it's strange that it isn't done more in academia. The chairperson approaches the two assistant professors most qualified to pick up an uncovered but needed three-credit class and says, in effect, "Why don't you two decide which of you would like to give this course, or perhaps take turns giving it in alternate years, because it's a class we can't properly offer without class meetings, and it's one that should be available every year. Nobody else is either interested or qualified, but it would mean three more weekly hours, one quarter of the year, every other year, for each of you if you shared it. However, the executive committee has agreed to recommend to the department that your salary increase should take this additional teaching time into account, over and above across-the-board increases or merit for scholarly publication. It is the dean's firm policy during our faculty retrenchment to stand behind such understandings, which will be committed to writing, if you wish to go along with this suggestion." I don't see anything except a silly hypocrisy and conventionalism preventing this kind of special tailor-made agreement. Chairs and colleagues do take this sort of extra work into account when it comes to merit raises after the fact, but it is considered tasteless or wicked to make such specific "deals" explicit before the fact. This is an unrealistic mental set in the academy and in these times should be abandoned in favor of a more rational, explicit contractual obligation approach that prevails in the business

world outside of the academy. If the Seven Sacred Cows are seriously challenged by a rational faculty that understands the requirement that *if they want their salaries to go up, something has to be done of a fairly drastic nature*, I find it hard to imagine many faculty members thinking that such temporary individual arrangements are inherently objectionable. Of course the pressure would be greater on junior staff who are, on the average, more pressed for additional money. But even some full professors would not be averse to sharing the instructional burden in a course if they knew it would have a specifiable impact on their salary during whatever period in which they carry an increased load. The administrative model already exists where department chairs receive an increment that is taken away when they cease being chair, a practice that did not prevail even in the 1950s when I chaired the psychology department. As in so many areas of life, "it's a matter of what you have gotten used to."

A reader of the book draft says it looks as if we are getting something for nothing, increasing salaries with negligible loss in output and fixed weekly hours. But we're not getting something for nothing. The instructional innovations have increased efficiency, as does any labor-saving invention (think of the cotton gin or the word processor). Of course the total work product has decreased, namely, by non-appearance of those reams of dismal publications. Using economist lingo, one might view this latter not as decreased production but as eliminating an externality, accepting Imre Lakatos' characterizing much of it as "intellectual pollution" (Lakatos 1970).

Afterword

I confess that I have written partly for catharsis, and partly as a "clerk" (Julien Benda, 1928) feeling an obligation to comment critically on the conventional wisdom, especially when it appears to be causing trouble for the social group. I have been asked why I bother to expound views that, even if correct, have negligible chance of being accepted, let alone implemented. My reply is that fulfilling the clerk's intellectual criticism role need not require optimism as to one's influence. As I said in the Introduction to this book, I do not claim to *prove* that these Seven Sacred Cows are all false. I do claim that they are not self-evident truths, and that there is pretty persuasive argument and empirical evidence against them.

REFERENCES

Academe (1991). *Administrative bloat*. Vol. *77*, Special section, pp. 11-24.

Albrow, M. (1970). *Bureaucracy*. New York: Praeger.

American Psychological Association [APA] (1963 v. 1; 1965, v. 2; 1969, v. 3). *Reports of the American Psychological Association's Project on Scientific Information Exchange in Psychology*. Washington, DC: APA. [1963 has findings: 50% of papers published in APA journals are read by fewer than 200 people, & ⅔ of the articles are never cited by another article.]

Andersen, K. E. (1991). Anatomizing bloat. *Academe*, 77, 20-24.

Anderson, M. (1992). *Impostors in the temple: American intellectuals are destroying our universities and cheating our students of their future*. New York: Simon & Schuster.

Andreski, S. (1972). *Social sciences as sorcery*. London: Deutsch.

Ayer, A. J. (1936). *Language, truth, and logic*. London: Victor Gollancz.

Bane, C. L. (1925). The lecture vs. the class-discussion method of college teaching. *School and Society*, *21*, 300-302. [Old, maybe 1st, comparison of teaching methods, 1925, Minnesota (lecture vs. discussion). Well done, no difference found. Cite as earliest, in Chap II, for "history"?]

Barzun, J. (1945). *Teacher in America*. Indianapolis, IN: Liberty.

Begley, S. (1991, Jan 14). Gridlock in the labs: Does the country really need all those scientists? *Newsweek*, p. 44. [Cites Inst. Scientific Info., 45% of papers published between 1981–1985 not cited 5 years post.]

Benda, C. (1994). *Going up: A tuition primer*. University of Minnesota Alumni Association, *Minnesota* (November-December, pp. 28-32).

Benda, J. (1928). *Betrayal of the intellectuals* (Trans. R. Aldington). New York: W. Morrow. Original publication, *La trahison des clercs* (1927).

Bergmann, B. R. (1991). Bloated administration, blighted campuses. *Academe*, 77, 12-16.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, *57*, 195-203.

Blackburn, R. T., & Clark, M. J. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. *Sociology of Education*, *48*, 242-256. [Cover studies of teaching & research—cites many, r's negligible but (+). Self-evaluations agree poorly with colleagues, deans, students. But latter agree fairly well. Is cited in Marsh #4042. Ref section Teacher Rating but not to discuss in text.]

Boyer, E. L. (1987). *The undergraduate experience in America*. New York: Harper & Row.

Braun, J. R. (Ed.). (1966) *Clinical psychology in transition* (Rev. ed.). Cleveland, OH: World.

REFERENCES

Breland, H. M. (1977). A study of college English placement and the Test of Standard Written English. College Entrance Examination Board Research and Development Reports RDR-76-77, No. 4; Project Report PR-77-1. Princeton, NJ: Educational Testing Service.

Breland, H. M., & Gaynor, J. L. (1979). A comparison of direct and indirect assessments of writing skill. *Journal of Educational Measurement*, *16*, 119-128.

Bresler, J. B. (1968). Teaching effectiveness and government awards. *Science*, *160*, 164-167. [Excellent study. In Sci, Soc Sci, Humanities faculty with more pubs and/or more external money are rated as better teachers by students]

Broad, W. J. (1981). The publishing game: Getting more for less. *Science*, *211*, 1137-1139.

Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, *37*, 34-36.

Burton, R. E., & Kebler, R.W. (1960). The "half-life" of some scientific and technical literatures. *American Documentation*, *11*, 18-22.

Calkins, M. E., Iacono, W. G., & Ones, D. S. (2008). Eye movement dysfunction in first-degree relatives of patients with schizophrenia: A meta-analytic evaluation of candidate endophenotypes. *Brain and Cognition*, *68*, 436-461.

Campbell, D. P. (1965). The vocational interests of American Psychological Association presidents. *American Psychologist*, *20*, 636-644.

Campbell, D. P. (1971). *Handbook for the Strong Vocational Interest Blank*. Stanford, CA: Stanford University Press.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378–399.

Casse, D., & Manno, B. V. (1998). The cost and price of college and the value of higher education. *Academic Questions*, *11*, 38-54.

Castell, A. (1935). *A college logic*. New York: Macmillan.

Cattell, R. B. (1950). *Personality: A systematic theoretical and factual study*. New York: McGraw Hill.

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal*, *14*, 17-24.

Chance, P. (1984). The revolutionary gentleman: *PT* conversation with Fred Keller. *Psychology Today*, *18*, 42-48.

Chase, S. (1938). *The tyranny of words*. London: Methuen.

Chu, G. C., & Schramm, W. (1967). Learning from television: What the research says. Washington, DC: National Association of Educational Broadcasters. [Classic summary, 207 studies (421 comparisons) show ETV, Traditional Lectures [p. 7 table]: 73% no diff, 15% TV better, 12% Traditional lecture better. Elementary through college level.]

REFERENCES

Clark, K. E. (1957). *America's psychologists: A survey of a growing profession*. Washington, DC: American Psychological Association. [cited by Myers, citation counts are best measure]

Coffman, W. E. (1969). Achievement tests. In R. L. Ebel (Ed.), *Encyclopedia of educational research* (4th ed.) (pp. 7-17). New York: Macmillan. [Can make selection test tap what supply-type test does (p. 10, col 1). Essay reliabilities range from .32 to .98 in studies, depending on things (p. 10, col 2). Readers react to spelling errors more than grammar or punctuation (p. 11, col 1).]

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 271-302). Washington, DC: American Council on Education.

Coffman, W. E., & Kurfman, D. (1968). A comparison of two methods of reading essay examinations. *American Educational Research Journal*, *5*, 99-107. [Holistic grading has larger error than analytical (p.103, 105) altho author doesn't stress that. "Analytic" is itself global, not a true school solution checklist.]

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

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, *13*, 321-341.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*, 281-309.

Cohen, P. A., Ebeling, B. J., & Kulik, J. A. (1981). A meta-analysis of outcome studies of visual-based instruction. *Educational Communication and Technology Journal* [*ECTJ*], *29*, 26-36. [Excellent. Visual "negligible difference" is shown. Box score shows significance in favor of ETV, but effect size is small. One of best sources I have.]

Cole, J. (1970). Patterns of intellectual influence in scientific research. *Sociology of Education*, *43*, 377-403.

Cole, J., & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the *Science Citation Index*. *American Sociologist*, *6*, 23-29.

Cole, J. R., & Cole, S. (1972). The Ortega hypothesis. *Science*, *178*, 368-374.

Cook, D. A. (1974). Personalized system of instruction: Potential problems. EPIE [Educational Products Information Exchange] Report No. 51, vol. *7*(April), 2-12.

Dalkey, N. C. (1969). The Delphi method: An experimental study of group opinion. *Futures*, *1*, 408-426.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*, 458-467.

Dennis, W. (1954). Productivity among American psychologists. *American Psychologist*, *9*, 191-194.

Dennis, W. (1955). Variations in productivity among creative workers. *Scientific Monthly*, *80*, 277-278.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*, 497-509.

REFERENCES

Downs, A. (1967). *Inside bureaucracy*. Boston, MA: Little Brown.

Dubin, R., & Hedley, R. A. (1969). *The medium may be related to the message: College instruction by TV*. Eugene, OR: University of Oregon Press.

Dubin, R., & Taveggia, T. C. (1968). *The teaching-learning paradox: A comparative analysis of college teaching methods*. Eugene: University of Oregon Press. [Pre-meta-analysis, but they both tally box score and pooled (individual) data. Lecture, discussion, mixed, supervised independent study, unsupervised independent study. Graphs and tables show clearly "nothing differs."]

Dunnette, M. D., Campbell, J., & Jaastad, K. (1963). The effect of group participation on brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology*, *47*, 30-37.

Dwyer, C. A. (1982). Achievement testing. In H. E. Mitzel (Ed.), *Encyclopedia of educational research* (5th ed.) (pp. 12-22). New York: Free Press. [Essay vs mult-choice: Table of pros and cons, each. p.14]

Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society* (*Series A*)*, 51*, 599-635.

Edgeworth, F. Y. (1890). The element of chance in competitive examinations (Part I). *Journal of the Royal Statistical Society* (*Series A*)*, 53*, 460-475; (Part II), *53*, 644-663.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, *84*, 158-172.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses. *Research in Higher Education*, *21*, 45-116. [Not meta-analysis but big batch of studies. Slight negative *r* between class size and student evaluation. Several U-shaped findings. Main theme, a negligible trend. Cite in ch. III, not quote]

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education*, *26*, 227-298. [McK says "best review of studies on research & teaching."]

Feynman, R. (1986). *Surely you're joking, Mr. Feynman!* New York: Bantam.

Fiske, D. W. (1974). The limits of the conventional science of personality. *Journal of Personality*, *24*, 1-11.

Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science*, *13*, 285-305.

Friedman, C. P., Hirschi, S., Parlett, M., & Taylor, E. F. (1976). The rise and fall of PSI in physics at MIT. *American Journal of Physics*, *44*, 204-211. [PSI did not "fail" at MIT, by usual standards. It was *perceived* to fail, plus a bad logistic error in first big use. A *social psychology*, sacred cow problem!!]

Gage, N. (1970). The Handbook of Research on Teaching. *Educational Researcher, 21*(8), 12-13. Retrieved 12/12/2018 from http://www.jstor.org/stable/44177809

Garfield, E. (1970). Citation indexing for studying science. *Nature*, *227*, 669-671. [Garfield introduced concept of science citation index 1955.]

REFERENCES

Garvey, W. D. (1979). *Communication: The essence of science*. New York: Pergamon.

Garvey,W. D., & Griffith, B. C. (1971). Scientific communication: Its role in the conduct of research and creation of knowledge. *American Psychologist*, *26*, 349-362.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*, 309-320.

Gessner, P. K. (1973). Evaluation of instruction. *Science*, *180*, 566-569.

Gillmore, G. A., & Greenwald, A. G. (1999). Using statistical adjustment to reduce biases in student ratings. *American Psychologist*, *54*, 518-519.

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. K. (1982). *School class size: Research and policy*. Beverly Hills, CA: Sage.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goodrich, A.T., Henry, J., & Goodrich, D. W. (1954). Laughter in psychiatric staff conferences: A sociopsychiatric analysis. *American Journal of Orthopsychiatry*, *24*, 175-184.

Gronlund, N. E. (1988). *How to construct achievement tests*. Englewood Cliffs, NJ: Prentice-Hall.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 1-31.

Guzzo, R. A., & Shea, G. P. (1992). Group performance and intergroup relations in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, Vol. 3. (pp. 269-314). Palo Alto, CA: Consulting Psychologists Press.

Halfond, J. A. (1991). How to control administrative cost. *Academe*, 77, 17-19.

Hamilton, D. P. (1990). Publishing by—and for?—the numbers. *Science*, *250*, 1331-1332.

Hamilton, D. P. (1991). Research papers: Who's uncited now? *Science*, *251*, 25. [cited by Anderson 1992, *Impostors*]

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.) (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Hartle, T. W., & Galloway, F. J. (1996, Sep-Oct). Too many PhDs? Too many MDs? *Change*, pp. 27-33.

Haswell, R. H. (2008). Machine scoring of essays and computer analysis: A bibliography. [On-line, accessed 05/01/2017] http://comppile.org/profresources/machinescoringbib.htm

Hayakawa, S. I. (1947). *Language in action*. New York: Harcourt Brace.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.

Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin*, *91*, 517-539.

Hofstadter, R. (1963). *Anti-intellectualism in American life*. New York: Knopf.

REFERENCES

Hogan, R., Desoto, C. B., & Solano, C. (1977). Traits, texts, and personality research. *American Psychologist*, *32*, 255-264.

Hopkins, K. D., & Stanley, J. C. (1981). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall. [chapter 8, Constructing and using essay tests.]

Huber, R. M. (1992). *How professors play the cat guarding the cream: Why we're paying more and getting less in higher education*. Fairfax, VA: George Mason University Press.

Hudelson, E. (1928). *Class size at the college level*. Minneapolis, MN: University of Minnesota Press.

Hull, D. L. (1988). *Science as a process*. Chicago, IL: University of Chicago Press.

Iacono, W. G., & Grove, W. M. (1993). Schizophreia reviewed: Toward a integrative genetic model. *Psychological Science*, *4*, 273-276.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. (1964). Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, *55*, 344-351.

Jamison, D., Suppes, P., & Wells, S. (1974). The effectiveness of alternative instructional media: A survey. *Review of Educational Research*, *44*, 1-67. [Tables show ETV ≥ traditional lecture, all subject matters, all levels through college. Very clear, numerous studies. (Most tables from Chu & Schram, 1967)]

Janis, I. L. (1972). *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoe*s. Boston, MA: Houghton, Mifflin.

Jensen, A. R. (1998). The *q* factor. Westport, CT: Praeger.

Johnson, K. R. and Ruskin, R. S. (1977). *Behavioral instruction: An evaluative review*. Washington, DC: American Psychological Association.

Johnson, W. (1946). *People in quandaries: The semantics of personal adjustment*. New York: Institute for General Semantics.

Johnston, J. (1987). *Electronic learning: From audiotape to videodisc*. Hillsdale, NJ: Erlbaum.

Kaufman, A. L. (1998). *Cardozo*. Cambridge, MA: Harvard University Press.

Keller, F. S. (1968). Goodbye, teacher. *Journal of Applied Behavior Analysis*, *1*, 79-89.

Kerr, C. (1964). *The uses of the university*. Cambridge, MA: Harvard University Press.

Korzybski, A. (1933). *Science and sanity*. New York: Institute for General Semantics.

Krislov, S. & Rosenbloom, D. (1981). *Representative democracy and the American political system*. New York: Praeger.

Kuhn, T. S. (1962) *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–195). Cambridge: Cambridge University Press. Reprinted in J. Worrall & G. Currie (Eds.), *Imre Lakatos:*

REFERENCES

Philosophical papers. Vol. I: The methodology of scientific research programmes (pp. 8–101). New York: Cambridge University Press, 1978.

Logan, F. A. (1993, May). Multiple-choice tests. *APS Observer*, p.23.

Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, *2*, 363-376.

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology*. Vol. 1. *Matters of public interest*. Minneapolis, MN: University of Minnesota Press. Pp. 3-39.

Macrae, N. (1992). *John von Neumann*. New York: Pantheon. [In the 1093s, at American univerisities "the average teaching load was twelve or fourteen hours a week" [and] "research was an optional activity." (pp. 174-175)]

MacRoberts, M. H., & MacRoberts, R. R. (1984). The negational reference: or the art of dissembling. *Social Studies of Science*, *14*, 91-94. [Example bias source in content of published articles.]

Mahoney, M. J. (1985). Open exchange and epistemic progress. *American Psychologist*, *40*, 29-39.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*, 253-388. [215 refs. "Best survey"—PWFox]

Marsh, H. W., & Roche, L. A. (1999). Rely upon SET [student evaluations of teaching] research. *American Psychologist*, *54*, 517-518.

Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, *6*, 97-106. [Sizeable effects (.4 to .75 sigma) of handwriting neatness & compositional errors on grades, despite instructions to grade on content only. Big N. Clear. Also replicates earlier studies by authors.]

McGuire, W. J. (1973). The yin and yang of progress in social psychology. *Journal of Personality and Social Psychology*, *26*, 446-456.

Meehl, P. E. (1954). *Clinical versus statistical prediction*: *A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press. Reprinted with new Preface, 1996, Northvale, NJ: Jason Aronson. Reprinted, 2013, by Echo Point Books (www.echopointbooks.com/)

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, *13*, 102-128.

Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist*, *15*, 19-27.

Meehl, P. E. (1971). A scientific, scholarly, nonresearch doctorate for clinical practitioners: Arguments pro and con. In R. R. Holt (Ed.), *Anew horizon for psychotherapy: Autonomy as a profession* (pp. 37-81). New York: International Universities Press.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806-834.

Meehl, P. E. (1986a). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, *50*, 370-375.

REFERENCES

Meehl, P. E. (1986b). Psychology: Does our heterogeneous subject matter have any unity? *Minnesota Psychologist*, *35*(Summer), 3-9.

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, *1*, 108-141, 173-180.

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244. Also in R. E. Snow & D. Wiley (Eds.), *Improving Inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 13-59). Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

Menard, H. W. (1971). *Science: Growth and change*. Cambridge, MA: Harvard University Press.

Merton, R. K. (1961). Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society*, *105*, 470-486.

Miner, F. C. (1984). Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behavior and Human Performance*, *33*, 112-124.

Mischel, W. (1977). On the future of personality measurement. *American Psychologist*, *32*, 246-254.

Morrison, D. E., & Henkel, R. E. (Eds.). (1970). *The significance test controversy*. Chicago, IL: Aldine.

Murray, H. A. (1938). *Explorations in Personality*. New York: Oxford University Press.

Myers, C. R. (1970). Journal citations and scientific eminence in contemporary psychology. *American Psychologist*, *25*, 1041-1048, [cite in chap IV and for "loss of knowledge" and "selecting research depts"; This is overall a powerful justification for using citation counts. Ratings of *names* by faculty has poor interjudge agreement, for example.]

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Chicago, IL: Aldine-Atherton.

Niskanen, W. A. (1973). *Bureaucracy—servant or master?* London: Institute for Economic Affairs.

Novick, P. (1988). *That noble dream: The "objectivity question" and the American historical profession*. Cambridge, MA: Cambridge University Press.

Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, *47*, 238-243.

Page, E. B. (1985). Computer scoring of essays. In T. Husen & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (pp. 944-946). Oxford, England: Pergamon.

Page, E. B., Tillett, P. I., & Ajay, H. B. (1989). Computer measurement of subject-matter essay tests: Past research and future promise. In *Proceedings* of the first annual meeting of the American Psychological Society; Alexandria, VA: American Psychological Society.

Pendlebury, D. A. (2013) Research fronts: In search of the structure of science. Addendum to C. King & D. A. Pendlebury, *Research fronts 2013: 100 top-ranked specialties in the sciences and social*

REFERENCES

sciences (pp. 26-31). Retrieved from http://garfield.library.upenn.edu/papers/pendleburykingresearchfronts2013.pdf

Price, D. J. de S. (1965). Networks of scientific papers. *Science*, *149*, 510-515. [Reprinted in Price, 1986, pp. 82-118]

Price, D. J. de S. (1986). *Little science, big science—and beyond*. New York: Columbia University Press.

Rescher, N. (1978). *Scientific progress: A philosophical essay on the economics of research in natrual science*. Pittsburgh, PA: University of Pittsburgh Press.

Rescher, N. (1980). *Unpopular essays on technological progress*. Pittsburgh, PA: University of Pittsburgh Press.

Roche, G. (1994). *The fall of the ivory tower: Government funding, corruption, and the bankrupting of American higher education*. Washington, DC: Regnery.

Roe, A. (1953). *The making of a scientist*. New York: Dodd, Mead.

Ruch, G. M. (1929). *The objective or new-type examination*. New York: Scott, Foresman.

Schlenker, B. R. (1974). Social psychology and science. *Journal of Personality and Social Psychology*, *29*, 1-15.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173-1181.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115-129.

Schramm, W. (1977, Jul-Aug). The researcher and the producer in ETV. *PTR*, 11-21. [Schramm, the big ETV guy, on no helpful theory, time not ripe. Teachers' resistance. Edge favoring ETV is less, the better experiments are done. Important from him! "ETV can teach anything."]

Sechrest, L., Kihlstrom, J. F., & Bootzin, R. R. (1993, January). How to develop multiple-choice tests. *APS Observer* (pp. 10-11, 22-23).

Shaffer, L. F. (1953). Of whose reality I cannot doubt. *American Psychologist*, *8*, 608-623.

Sherman, J. G. (Ed.) (1974). *Personalized system of instruction: 41 germinal papers*. Menlo Park, CA: W.A. Benjamin.

Siegel, L., Adams, J. F., & Macomber, F. G. (1960). Retention of subject matter as a function of large group instructional procedures. *Journal of Educational Psychology*, *51*, 9-13. [Retention 1 or 2 years later, no difference between TV, large class, small classes. Nine subjects, nice spread (e.g., psychology, air science, geography, social studies, English).]

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24*, 86-97.

Skinner, B. F. (1984). The shame of American education. *American Psychologist*, *39*, 947-954.

Smith, M. B. (1973). Criticisms of a social science. *Science*, *180*, 610-612.

REFERENCES

Smith, P. (1990). *Killing the spirit: Higher education in America*. New York: Viking Penguin.

Sowell, T. (1993). *Inside American education*. New York: Free Press.

Standing, L., & McKelvie, S. (1986). Psychology journals: A case for treatment. *Bulletin of the British Psychological Society*, *39*, 445-450.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *The School Review*, *20*, 442-457.

Starch, D., & Elliott, E. C. (1913a). Reliability of grading work in History. *The School Review*, *21*, 676-681.

Starch, D., & Elliott, E. C. (1913b). Reliability of grading work in Mathematics. *The School Review*, *21*, 254-259.

Strohm, P. (1991). Editorial: The reforms of Diocletian. *Academe*, *77*, p. 11. [Editor for special section on Administrative Bloat.]

Strong, E. K. (1943). *Vocational interests of men and women*. Stanford, CA: Stanford University Press.

Sullivan, A. M. (1985). The role of two types of research on the evaluation and improvement of university teaching. In J. G. Donald & A. M. Sullivan (Eds), *Using research to improve teaching* (pp. 71-82). New Directions for Teaching and Learning, no. 23. San Francisco, CA: Jossey-Bass.

Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. *Journal of Educational Psychology*, *66*, 584-590.

Sykes, C. J. (1988). *ProfScam: Professors and the demise of higher education*. New York: St. Martin's.

Taveggia, T. C. (1976). Personalized instruction: A summary of comparative research, 1967–1974. *American Journal of Physics*, *44*, 1028-1033.

Thorndike, R. L. (1955). The structure of preferences for psychological activities among psychologists. *American Psychologist*, *10*, 205-207.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *29*, 437-484.

Tyler, R. W. (1934). *Constructing achievement tests*. Columbus, OH: Bureau of Educational Research, Ohio State University.

U. S. News and World Report. (1996, Dec. 30). America's best graduate schools.

Varca, P. E., & Levy, J. C. (1984). Individual differences in response to unfavorable group feedback. *Organizational Behavior and Human Performance*, *33*, 100-111.

Veysey, L. R. (1965). *The emergence of the American university*. Chicago, IL: University of Chicago Press. [Excellent history. From early to 1910, mainly 1870–1910.]

Wiggins, J. (1973). Despair and optimism in Minneapolis. *Contemporary Psychology*, *18*, 605-606.

Wispé, L. G. (1963). Traits of eminent American psychologists. *Science*, *141*, 1256-1261.

Appendix: Academic Novels

Books from Meehl's personal reading list for those who enjoy novels about academia. He added a subjective rating for most as to how much it is about academic life:

\+ All about ± Somewhat about – Academia is only the setting

Some items lack a rating merely because he died before getting this far with editing. The two marked with asterisks are non-fiction.

Alcorn, Alfred: *Murder in the Museum of Man*. Cambridge, MA: Zoland Books, 1997. (±)

Amis, Kingsley: *Lucky Jim*. New York: Doubleday, 1954.

Baker, Carlos Heard: *A Friend in Power*. New York: Charles Scribner's Sons, 1958. (+)

Bernays, Anne: *Professor Romeo*. New York: Weidenfeld & Nicholson, 1989.

Bradbury, Malcolm: *The History Man*. Boston, MA: Houghton Mifflin, 1975.

Bradbury, Malcolm: *Stepping Westward*. London, Eng.: Secker & Warburg, 1965. (+)

Cather, Willa Sibert: *The Professor's House*. New York: Alfred A. Knopf, 1925. (–)

Fuller, Henry Blake: *Bertram Cope's Year*. Chicago, IL: R. F. Seymour, 1919. (±)

Galbraith, John Kenneth: *A tenured professor*. Boston, MA: Houghton Mifflin, 1990. (±)

Grudin, Robert: *Book: A Novel*. New York: Random House, 1992.

Guérard, Albert Joseph: *The Hunted*. New York: Alfred A. Knopf, 1944.

Hassler, Jon: *The Dean's List*. New York: Ballantine Books, 1997.

Herrick, Robert: *Chimes*. New York: Macmillan, 1926. (±)

Howe, Helen Huntington: *We Happy Few*. New York: Simon & Schuster, 1946. (±)

Hull, Helen Rose: *The Asking Price*. New York: Coward-McCann, 1930. (±)

Imbs, Bravig: *The Professor's Wife*. New York: Dial, 1928. (–)

Jackson, Shirley: *Hangsaman*. New York: Farrar, Straus & Young, 1951. (–)

Jarrell, Randall: *Pictures from an Institution*. London, Eng.: Faber & Faber, 1954.

*Kluge, Paul Frederick: *Alma Mater: A College Homecoming*. New York: Addison-Wesley, 1993. [*Non-fiction, ⅓ time English professor, novelist, account of an academic year at Kenyon College. Good for readers (a) unacquainted with academia, (b) academics not familiar with small high-quality liberal arts college life. Pp. 121-124 a hilarious description of admissions committee free-association and fecklessness. Lots of quotes, the usual inefficient clinical prediction talk.*]

Larson, Martin Alfred: *Plaster Saint: A Novel of Heresy on the Campus*. New York: Exposition Press, 1953. (+)

Lehman, Benjamin Harrison: *Wild Marriage*. New York: Harper & Brothers, 1925. (–)

Lodge, David: *Changing Places*. New York: Penguin, 1988.

Lodge, David: *Nice Work*. New York: Viking, 1989

Appendix: Academic Novels

Lodge, David: *Small World*. London, Eng.: Secker & Warburg, 1984.

Lurie, Alison: *War Between the Tates*. New York: Random House, 1974.

*Lyons, John O.: *The College Novel in America*. Carbondale, IL: Southern Illinois University Press, 1962. [*Non-fiction; includes 250-item chronological "Bibliography of the College Novel in America" from 1883 to 1962.*]

Macauley, Robie: *The Disguises of Love*. New York: Random House, 1952. (–)

McCarthy, Mary Therese: *The Groves of Academe*. New York: Harcourt, Brace, 1952. (+)

Muske-Dukes, Carol M.: *Saving St. Germ*. New York: Viking, 1993.

Nabakov, Vladimir: *Pnin*. Garden City, NY: Doubleday, 1957. (±)

Neff, Wanda Fraiken: *Lone Voyagers*. Boston, MA: Houghton Mifflin, 1929. (+)

Nemerov, Howard: *The Homecoming Game*. New York: Simon & Schuster, 1957. (+)

Russo, Richard: *Straight Man*. New York: Random House, 1997. (+)

Sarton, May: *Faithful are the Wounds*. New York: Rinehart, 1955. (+)

Schmitt, Gladys [Mrs. Simon Goldfield]: *A Small Fire*. New York: Dial Press, 1956. (+)

Seton, Anya: *Foxfire*. Boston: Houghton Mifflin, 1950

Sherman, Susan: *Give Me Myself*. New York: World, 1961. (–)

Smiley, Jane: *Moo*. New York: Knopf, 1995. (+)

Snow, Charles Percy: *The Masters*. London, Eng.: Macmillan, 1951.

Stewart, George Rippey: *Doctor's Oral*. New York: Random House, 1939. (+)

Watkin, Lawrence Edward: *Geese in the Forum*. New York: Alfred A. Knopf, 1940. (+)

pdf by LJY December 2018