REVIEW OF HORACE JUDSON, THE GREAT BETRAYAL

Elihu M. Gerson
Tremont Research I nstitute
458 Twenty-Ninth Street
San Francisco, CA 94131
emg@tremontresearch.org

Horace Freeland Judson, *The Great Betrayal: Fraud in Science*. Orlando: Harcourt, 2004. $28.

The Great Betrayal is a broad-brush discussion of scientific fraud over the last quarter-century. It is a tale told by a very skilled and engaging writer, aimed the educated general public, and deeply concerned with the viability of a precious institution. Judson's exposition of the subject is certainly the best of breed. But the book is also useful to specialist researchers because it is so rich with interesting points worth following up. The book is really about a much broader range of issues than scientific misconduct. Because the nature of misconduct is entwined with all the other aspects of scientific work organization, Judson's topic is an opportunity for him to discuss many of the issues that science currently faces.

The organization of the book

Judson opens his discussion with a survey of major business frauds that appeared on front pages in the late 1990's and early 2000's. He also mentions the sexual scandals of the Roman Catholic church, as part of a concern with the problem of understanding organizational responses to misconduct. He doesn't mention any comparable failures of accountability on the part of government; perhaps there were none. Chapter 2 introduces a simple classification of scientific frauds borrowed from Charles Babbage's *Reflections on the Decline of Science* (1830): hoaxing, forging, trimming, and cooking. Hoaxing and forging consist of making things up out of whole cloth. Trimming means editing or "cleaning" the data to make it conform better to expectations. Cooking covers a variety of practices, such as "cherry-picking" the best data points, or adjusting the constants in data-smoothing formulas. Together with plagiarism, these are equivalent to the categories used in the official rules of American administrative law today: fabrication (hoaxing, forgery) and falsification (trimming, cooking). Judson fleshes out this chapter with a review of controversies which have grown up around a series of famous scientists accused of fraud-- Mendel, Darwin, Haeckel, Pasteur, Millikan, Freud, Cyril Burt.

In Chapter 3, Judson continues with his description of scientific frauds which have made news over the last quarter-century, focusing on what he calls "patterns of complicity". Fraud requires enabling circumstances if it is to be successful. Here Judson makes a critical point: fraud is not a matter of individual pathology, independent of circumstances in which it occurs. Rather, it takes place in the context of institutional arrangements which allow the forging, trimming and cooking to occur and be successful, at least for a while. One of the most interesting (some would say, disturbing) characteristics of discovered scientific frauds is that they are often part of an extensive pattern of misbehavior-- not single episodes of forging or trimming, but many, published again and again, sometimes dozens of times. This concern with institutional circumstances (and the concomitant refusal to blame misconduct on the psychopathology of individual offenders) is an important strength in Judson's approach.

Judson notes several "syndromes" which seem to accompany fraud. Many of the miscreants have been prodigies, for example, producing exceptionally large amounts of work in a short amount of time at a young age, sometimes appearing almost too good to be true. Another pattern, which Judson refers to again and again throughout the book, is failure of the mentoring relationship: inadequate or lax supervision, misplaced trust, a willingness to be (as Judson puts it) seduced by a miscreant protégé. He also notes what he calls the arrogance of power, by which he means not merely that important people think that they can do no wrong, but also that they think that no wrong can be done to them-- at least, not by one of their own folks. Of course, this is the traditional stuff of myth and literary tragedy. It's important, when considering tragedies, to keep a firm grip on the institutional analysis.

Chapter 4 recounts the history of the difficulties encountered in defining misconduct in the 1980s and 1990s. Judson summarizes the efforts to deal with two problems. The first is estimating the prevalence of fraud, and the concomitant problem of deciding what counts as an instance of fraud for the purpose of counting. The second is the story of formally defining scientific misconduct for administrative purposes.

This discussion is followed in Chapter 5 by a summary of the most famous single investigation of a possible fraud, which Judson calls the "the Baltimore affair" after the most prominent scientist involved with it. The accused in this case was immunologist Thereza Imanishi-Kari, who was in process of moving from the Whitehead Institute to Tufts University at the time the case blew up in May of 1986. Ten years later, Imanishi-Kari was cleared of the charge of fraud, although her reputation was left in tatters. Indeed, virtually all the people and organizations involved suffered damage to their reputations. The case is an outstanding example of the difficulties that can emerge when a new kind of situation engages institutional arrangements that are unprepared to deal with it.

Chapter Six is devoted to the problems of peer review. Judson thinks that peer review, both of grant proposals and of manuscripts submitted to journals, is an institution in crisis. It consumes a lot of time which might be devoted to research, it isn't at all clear that it's superior to the judgment of individual editors and program managers, it's subject to a variety of biases, it enables some forms of plagiarism, and it doesn't detect fraud. Judson is an effective exponent for the idea that peer review needs extensive reform, but he connects this complex of problems only loosely to the problem of fraud.

Chapter Seven is devoted to problems of authorship, credit, and intellectual property. Here, Judson deals with the problem of plagiarism, and with the broader problem of allocating and mis-allocating credit for work performed. This is a complex family of problems. Plagiarism and other credit misdeeds are a different sort of offense from fabrication and falsification. As David Hull (1988) puts it, plagiarism injures the individual victim of the offense, but it doesn't injure the scientific enterprise as a whole. The community still has the intellectual value of a work, no matter whose name is on it. On the other hand, falsification and fabrication injure the community because they introduce bad results into the body of materials with which the community works. Judson doesn't make much of this, because the now-official categories of misconduct are good enough for his purposes.

Plagiarism is not the only kind of misconduct involving credit allocation. Sometimes, people in authority may insist that their names appear on papers to which they have made no direct contribution. Sometimes, people in weak positions are excluded from authorship when they have a reasonable claim. Sometimes, co-authorship is given as a means of cultivating relationships or favors. All of these suspect usages (and there are many others) are difficult to detect and to evaluate. They bear little resemblance to the things covered in lawyers' theories of intellectual (or other) property rights. Yet they are an important part of the research world, where careers depend on credit for contributions to publications. Judson doesn't discuss the nature of the relationship between fraud and plagiarism. Misappropriation of credit is a much larger issue than plagiarism, and it deserves greater attention and analysis than it has received.

Chapter Eight is about the challenges and opportunities provided to scientific publishing by the emergence of the World Wide Web. The Web was invented as a means to improve distribution of research papers, and Judson's chapter is a good presentation of the possibilities that the Web has opened. The Web enables a vastly expanded and refined peer review. It also makes detection of plagiarism much easier and more convenient. On the other hand, the Web's connection to problems of fraud is less clear, although important. Judson's message is that openness of access, which the Web makes practical, provides one of the major hopes for ameliorating some of the difficulties which he has discussed.

Chapter Nine, perhaps the best part of the book, is about the relationship between the law and science, as revealed by the developing administrative procedures for dealing with misconduct. Much of the chapter is devoted to the equally fascinating story of the law's problems in making use of science, in the form of expert testimony. This is a large subject in its own right, even though it is only vaguely connected to the problem of misconduct; Golan (2004) introduces the history of it.

When Judson turns to the administration of misconduct problems, his treatment is sensitive, revealing, and effective. How is misconduct to be defined? How are people accused of misconduct to be treated? What rights do they have? Who is injured, in what manner, and to what degree? How is this to be decided? How are corrections to be determined? What are reasonable penalties for error? What are to be the arrangements for restitution? For retribution? What models from other institutions should be applied, and how should they be modified? Judson shows clearly that we do not yet have good institutional arrangements for dealing with accusations of fraud, although they are certainly better than they were twenty-five years ago. His description of the alternative approaches for recognizing and dealing with offenses advocated by different participants illustrates most powerfully and effectively the underlying problem once again: we have a hard time recognizing when we need significant institutional reform and innovation. New ways of doing things are made of the broken careers and macerated reputations of those caught up in the process of collectively learning how to solve a class of problems which no one has dealt with before.

In an epilog, Judson meditates on another issue altogether, the problems faced by science in a world where it is no longer growing exponentially.

Judson's work is rich, and raises many interesting problems. Before I discuss some of them, it's necessary to make some points about rhetoric. First, Judson frequently mentions what he calls "Constructionism", which seems to be a view whose tenets are so wildly specious as to not merit serious or careful discussion. "Constructivism" or "social construction" has become a catch-all term used to gather up a variety of scholars and critics who have not accepted the official view of science prevalent at the time of the Eisenhower administration. The category includes a variety of sociologists (who are inclined to disagree with one another), some rhetoricians, and a mixed bag of others, a few critical of science, some post-modern in approach. As a descriptive category, the notion is useless. Either specific positions should be addressed (for example, reasonably fair and technically competent criticism from an opposing philosophical viewpoint has recently been provided by John Zammito (2004), or the whole controversy should be dropped. There's nothing to be gained from perpetuating the utterly pointless and irresponsible "science wars" of the 1990s. Since the ideas of the “constructionists” play no acknowledged role in Judson's discussion, there's no point in his mentioning them.

Second, Judson is often at pains to identify the political commitments of selected scientists who have become embroiled in one controversy or another. Richard Lewontin, for example, is noted as a Marxist (p. 93) while Leon Kamin is called a "man of the left" (p. 94). Strangely, nobody is ever labeled a Fascist, a reactionary, or a "man of the right". This practice is asymmetrical, and if some of us have learned anything from the "Strong Program" wing of "Constructionism", it's that asymmetries of this kind are to be avoided, because they are biasing and usually invalid. Moreover, the use of such political labels, unattached to any analysis of why they are worth noticing, is a kind of improper *ad hominem* argument, the equivalent of "you don't know where it's been". These rhetorical moves do not serve either the scholarly community or public policy.

Any work like this needs some sort of analytic approach if it is to be more than a catalog of specimens or an exercise in table-pounding. The basis of Judson's approach isn't clear. He relies on the standard administrative categories of misconduct (falsification, fabrication, and plagiarism) to limit his subject matter. That aside, Judson has little to say explicitly in the way of an analytical approach. He mentions Merton's classification of the norms of the scientific community, first postulated in 1942 and then made familiar to sociologists over the years in various editions of Merton's collected papers. These norms, ("universalism", "disinterestedness", "organized skepticism", and "communism") have achieved what historian David Hollinger calls a codification of mythological writings about science (Hollinger 1983: 15). Merton's classification has never been an effective guide to understanding research; Judson himself rejects it, but does not provide an alternative analytical approach. In all fairness, Judson is not attempting a formal general analysis. Still, a coherent and explicit view would be helpful.

Detecting fraud: replication and robustness

One of the enduring tropes in science apologetics is that science is self-policing, so that errors and misconduct are both reliably detected. Two institutions are held to accomplish this feat: replication of results and peer review. Judson rejects the idea that scientists police their own work. He points out, quite rightly, that most work done in science is never replicated. In part, this is because much work (perhaps most) is simply ignored. Moreover, there is little payoff to replication. No one gets any credit for replicating someone else's work, nor are grant proposals to do so likely to get funded. Indeed, as Judson points out (p. 39), reports of attempts to replicate someone else's work are likely to be a sign of trouble, since the effort is undertaken only when inconsistencies have been encountered, and are reported only when the replication seems to fail. The notion that science polices itself via replication of studies is thus an oversimplification.

There are, of course, many technical difficulties with the notion of replication, and Judson notes them. Chief among these is the problem of ensuring that the conditions of the work have been precisely duplicated in a different setting: all the reagents similarly compounded and purified, all the equipment calibrated to the same standards, all the steps performed in exactly the same way and exactly the same order. This is almost impossible to accomplish. The most one can hope for in replication, is to eliminate a spurious result at very high cost in time and reputation. These issues were discussed in some detail in Harry Collins' *Changing Order* (1985), a "constructionist" classic that Judson does not mention, and in Neuliep (1991). Economists will also enjoy the parallel discussion by Mirowski (2004)

Replication narrowly construed consists of doing the *same* thing again and again, and seeing if the same results appear. Replication is good for recognizing the presence of accidents and uncontrolled variables. When we add up a column of figures twice, for example, we assume there's been an error of some sort if the results don't match. Replication in the context of a study of fraud means something different: performing the same procedure again, as nearly as possible, but in a different laboratory with a different staff. The fact that the results seem to be indifferent to who produces them is considered a good thing, and counts as "replication". But we need to distinguish between exact replication and reproduction of similar results under varying circumstances.

Strictly speaking, this is an example of robustness rather than replication. Where replication means doing the same thing again, robustness means doing *different* things, looking at *different* lines of evidence, *different* circumstances, in order to see if the same conclusion holds.

That is, independent lines of evidence and/or argument should lead to the same conclusion, and when they don't, something is wrong. This is a fairly complex and subtle subject, but certainly the general notion is correct; there would be little room for faith in science if the world were not self-consistent. Judson draws upon this notion again and again throughout his discussion (he calls it “triangulation” or “consilience”, p. 36), and repeatedly points to others doing the same: there are inconsistencies in the data, there must be something wrong; some characteristics of the situation seem unlikely or inconsistent, further checking is needed. It's very clear that scientists rely heavily on robustness as a criterion of adequacy.

Performing an experiment in different laboratories then, is a kind of robustness test. But the focus of concern in research is usually not on the scientists but on the procedure-- the collection and analysis of data, the consistency and relevance of models, the degree to which models and data are in agreement. It is robustness among different data sets, different procedures, and different models that constitutes the real pay-off.

In contrast to replication of the same study, robustness comes relatively cheaply. New results are noticed by scientists in the same or closely related specialties. Their concern is not to replicate so much as it is to make use of the result in their own research. Typically, they respond to claimed discoveries in one or both of two ways. First, they build on the result by conducting additional studies which extend and refine it (e.g., does the effect reported appear under different experimental conditions? In comparable places or organisms? What are its limits? What's a plausible model for the effect?). Second, they assume the validity of the results, and make use of that assumption in complementary studies. So, for example, Kelvin assumed the ideal gas law relating pressure, volume, and temperature (pv = kt) in his work defining absolute temperature; he also assumed the Carnot cycle relating heat and work and used gas thermometers to test his results (Chang 2004: 192 ff.). But these assumptions depend upon one another; their use is supportable only if we interpret validity as a matter of successive approximation via a series of different experiments. Another example is Boyle's studies of pressure, in which the conceptualization of the "spring of the air" and the development of the vacuum pump went hand in hand. This work gave rise to controversies over the nature of trust and reliability in science. These have been described in great detail by Steven Shapin (Shapin and Schaffer 1985, Shapin 1994), in additional major works of "Constructionism" unremarked by Judson.

Follow-on studies that assume, extend, and refine a result act, in effect, to test the quality of that result and assess its robustness. When the additional research meets expectations, it confirms, not only the immediate idea which motivated the research, but the assumptions and previous research which support it as well. If a result is spurious for some reason, the attempt to build upon and extend it will fail; expected results in the next round of work will not occur. Trouble-shooting the failure (which may involve replication in several ways) is a routine practice. It will usually reveal the source of the difficulty, which may or may not be some sort of misconduct.

The accumulated body of knowledge formed by many studies whose results support one another directly and indirectly provides a second way that robustness-checking guards efficiently against fraud. As knowledge increases, the increasing density of conceptual connections among things reduces the number of ways in which undiscovered phenomena might appear plausible. For example, if someone claimed to discover a flying horse, it would be of great interest because no one has seen a six-limbed mammal, because the descent of such a species would be a complete mystery, because the physiology needed to support respiration, digestion, and motion in a horse-sized flying animal are hard to imagine and because the ecological relationships of such an animal would certainly give cause to wonder. In short, such a thing is made very implausible by the network of established, self-consistent knowledge we already have. It follows that a would-be forger must work very hard indeed to come up with a plausible new result in a well-studied area. Conversely, forgery is relatively easy on the frontiers where there are relatively few experts and little is known.

The marginal cost of robustness testing is relatively low because scientists are doing the work as a routine part of the next steps in their program, in contrast to the high cost of replication. It is simply irrational to undertake replication studies unless and until there is a failure of robustness. And this is just what scientists do in practice; they follow the very reasonable rule of thumb, "if it ain't broke, don't fix it".

How reliable are the various kinds of robustness checking? What proportion of the problems of research (including fraud) do they catch in a reasonable amount of time? We don't know. Scientific research proceeds on the assumption that the world is consistent with itself, and that nature does not play dirty tricks on researchers. Unfortunately, sometimes researchers do play dirty tricks on each other. There has been very little research aimed at understanding the implications of this, and so the question of reliability of robustness-checking (and by implication, the need for other approaches to the problem of fraud) is open. The pertinent questions then are: what are the institutional mechanisms for detecting that things have gone wrong? And how reliable and effective are they?

One major problem for consideration is that much of the misconduct detection has come from "whistle-blowers", i.e., technically knowledgeable people in the same organization (typically, in the same laboratory) as the accused. The whistle-blowers typically notice that something is amiss, and report their concerns to organizational officials. This implies that the robustness-checking mechanisms we have seem to work to some degree as fraud detectors at close organization range and small organizational scales. But we are also faced with the problem of fraud that escapes the notice (or courage) of local whistle-blowers, and escapes into the wild. How often is it recognized? There's no doubt that many results prove anomalous and are rejected-- but this simply means that they are unused by the community; not incorporated into the expanding body of knowledge discussed in review articles, textbooks, seminars, referees reports, and so on. We have no way of knowing how much of these anomalous results are due to fraud, and how much to other causes. And of course, we also don’t know how many fraudulent results are not detected by robustness checking.

Detecting fraud: peer review and checkability

The second major institution for preventing and detecting misconduct is peer review. But peer review has not detected the persistent frauds which some miscreants have perpetrated over many years. It seems obvious, at first glance, that the cumulative skill and wisdom of a specialist community should be able to detect that something is seriously wrong over a sequence of dozens of fraudulent papers-- and yet, it doesn't, or doesn't reliably. Why not? Peer review is not effective at detecting fraud because it isn't intended to detect fraud: "The system is not designed to catch fraud, and could be expected to do so only rarely" (Judson, p. 247).

Scientists' work is typically reviewed by other scientists at three stages. In the early stage, it is reviewed prospectively by colleagues who assess the merit of proposals for funding, and award grants on the basis of technical merit and administrative feasibility. After the work is completed, draft reports submitted to journals are refereed to ensure that publications adhere to minimum standards of quality. And after a report is published, it (and its authors) are reviewed informally by fellow specialists, whose opinions form the basis of a scientist's reputation. This latter stage, completely informal and distributed over years, is probably the most important for the long run, since it cumulates the community's sense of how important the work is. This community opinion can be detected in many ways: in citation rates for papers, in the comments of book and literature reviews, in the invitations to give guest lectures, participate in closed conferences, run for office in learned societies, and in the accumulation of various honors and awards which the scientific community bestows. As a practical matter, it is reflected in requests for a scientist's opinion in many circumstances, formal and informal. Informal comments by respected scientists on a research project or program at meetings probably have as much practical weight as formal reviews in shaping the opinions of the relevant community.

Reviewing and refereeing are aimed at assessing and establishing the plausibility of reported results. They aren't aimed at auditing work performed, but rather *begin* with the working assumption that reports are honest (if sometimes mistaken). Sometimes, the report has flaws, and sometimes those flaws are of a nature to arouse suspicion. But there's no reason to assume that well-crafted frauds will be detected in the run of ordinary peer review, formal or informal.

The notions of peer review and robustness together suggest an improved procedure for detecting fraud, one which fits well with conventions of research practice already in place. This might be called *checkability* of results; the idea that results (i.e., claims about the character of nature) and the procedures used to obtain them should be easily (i.e., cheaply and conveniently) reviewed by peers and juxtaposed with other lines of research. The traditional mechanisms for doing this are bibliographic citations (which allow scholars to check claims about others' work) and publication of the data, models, and computer programs upon which conclusions are based. Such publication allows re-analysis and re-interpretation, and thus supports an important (if limited) form of robustness testing.

To do this scientists need ready access to the publications and to the supplementary materials upon which publication are based. Access to publications has itself become an important problem for the research community in recent years. On the one hand, the cost of books and subscriptions to journals has been rising steadily for many years. It has long since become routine for the data of a study to be omitted from published reports, so that a separate step is required to obtain them. On the other hand, the new Internet-based technologies make distribution of publication far faster, cheaper, and more convenient than ever before. The lowered costs of distribution mean that it is now feasible to provide supporting data, programs, and other documentation very easily, as the restrictions of print have been removed. The institutional arrangements to pay for and provide these services are still being hammered out, but there can hardly be any doubt that some such arrangements will become increasingly common in the next few years.

But merely providing ready access to supporting materials in a traditional way is not enough, and there are practical alternatives. Publishing tables summarizing the research is good;

it is far better to publish them in a form which supports additional analysis and comparison with analogous tables in other publications. The critical point is that checkability is greatly expanded if the detail work of checking can be partially automated. The growth of the Internet and associated technologies gives some hope (only hope, thus far) can this work can be greatly reduced in effort and cost. Several possibilities seem worth investigating, although I'm sure there are others.

First, materials should be presented in formats that support further analysis. For example, data should be presented as spreadsheets or database tables rather than text tables, and models should be presented in a form suitable for import into mathematical analysis programs. Moreover, we need to develop descriptive formats ("metadata") which support search, retrieval and comparison by machine. This will certainly be a difficult task, and there is no guarantee it can be done at all. But the value of doing so is more than detecting fraud. It also would support discovery of relationships and re-formulation of problems.

Second, the use of footnotes should be revised and expanded in a form that supports retrieval for purposes of comparison and checking. One difficulty with this is that it's a lot of work to track down footnotes, and the payoff to doing so is not large. This would add significant labor to the preparation of scientific reports, especially at first. But the labor would be devoted to making the reports more accessible and more usable to the research community, even without any consideration of fraud detection.

A third possibility is to use modern information technology to cumulate public comments, thereby extending the peer review process. One way this might work is to attach a comment section to each published article or book. Comments would be signed with "handles" (i.e., pen names), but commentators' actual identities would be known to the publisher, just as referees are now. This would permit two important further steps. First, the comments themselves could be rated, and commentators could thus accumulate reputations for providing good (or bad) comments. This would be a strong incentive to keep comments fair and helpful. Benkler (2002) describes the workings of one such comment-rating system. Second, the community could make arrangements for refereeing and reviewing to receive credit at hiring and job evaluation times, so that the work becomes recognized as a valuable part of the research community's effort. As things stand now, there are few incentives to do it well (or even at all), and a significant part of the community's productive effort is thus disconnected from its system of rewards.

Why does fraud occur?

One of the many real strengths of Judson's book is his insistence that the problem of fraud must be understood as a one of defective institutions, not as a rash of individual mis- or malfeasance. This leads us to the question: How do the various institutions involved interact to generate higher and lower rates of misconduct? And how are institutional arrangements translated into actual misconduct? What are the mechanisms? Judson doesn't deal directly with these issues which are presumably the subject of future research.

For example, much of the fraud seems to have taken place in or near clinical situations--i.e., circumstances in which the practical payoffs are substantial. But this also raises the question of different evaluation standards at work. The pattern of institutional pressures that physicians face is different from those that constrain basic research scientists. Understanding these contingencies and their effects is an open research problem.

Since World War II, the federal government has become by far the largest funder of research. The growth of research funding has been large, but the rate of growth was as just high in the half-century before the expansion of the NIH and founding of NSF. Judson sees the consequence of this long-term growth as vastly increasing the number of scientists and specialties, which in turn tends to fragment communities which oversee the work.

But this assumes that "fragmented" communities are less capable or less interested in overseeing their own work. There's no reason to believe this in the absence of data to the contrary. The number of knowledgeable colleagues that a scientist must take into account is probably about the same now as it was before World War II. The number of specialties has increased dramatically, but the relevant communities tend to stay about the same size. What has changed is the chance that people *in the same organization* will be able to evaluate one another's work effectively. This points in turn to an important issue that is not explicitly discussed in Judson's book. It's the technical community that has the *skills* to oversee a line of research and recognize when something has gone wrong. But it's the host organization (e.g., an employing university) that has the administrative *responsibility* for preventing, investigating, and taking action against misconduct. This is so because it is the host organization that receives and spends the government grant funds. That's one reason that whistle-blowers tend to work near the fraud they reveal. Scholarly associations typically have no (or only very weak) apparatus for preventing, detecting, investigating, or punishing misconduct. Future studies will have to look more closely at the ways that alternative organizational arrangements encourage or block both fraud and its detection.

Dealing with fraud once it's been detected

The importance of government as sponsor is not simply the scale of work that results. Instead, the significance of government funding is that people take the expenditure of public funds very seriously, and they want to get good value for their tax money. This is hardly arguable in principle, but it makes for real problems in evaluating research. How are we to determine what constitutes "good value"? And who is to do the determining? This is difficult enough, but there is an additional complexity, which is that the assessment of propriety in the expenditure of public funds already has an elaborate set of standards in place, and those standards have been formulated by lawyers and accountants to anticipate injuries to the government's economic interest. So we have a situation in which the perspectives of different professions can come into conflict. I am not arguing here for more or less "accountability"; I am arguing that there are built-in misunderstandings and conflicts over just what accountability consists in. This results in the spectacle of scientists telling Congress that they are incapable of judging the technical issues involved in a charge of misconduct, which Congress hears as a claim that it is not qualified to make decisions about expenditures of public funds.

It's clear that the mere presence of government money changes the rules in complex ways. The role of government as sponsor is inextricably tied up with the role of the government as law enforcement agency and regulator. But the relative merits and demerits of the funder-as-cop aren't obvious. For example: if I collect data on human subjects without government funding, then I may protect confidentiality of respondents by destroying my field notes should threatening circumstances arise. But if my study is government funded, then the notes aren't mine; rather, they belong to the organization that was awarded the grant, and destroying them might be interpreted as an unethical (if not criminal) act. Many social scientists thus incur large risks that are difficult to estimate when they accept government funding.

More generally, different stakeholders in the system have different standards and different interests, and there is no meta-standard that tells us which values to apply under what circumstances. Much of the work of science studies in recent years has been devoted to considering these issues, which are extraordinarily difficult to articulate effectively. Instead, we need to look at the institutional arrangements for detecting and correcting misconduct.

Judson notices a frequent (but not universal) reaction among administrators told that fraud has been discovered in their organizations. There's a natural tendency to disbelieve at first, and the hope that somehow it's all a mistake. But of course, showing that it's a mistake implies a rapid, thorough, and careful investigation. Unfortunately, there's been a tendency to cover up-- deny something bad has occurred, investigate carelessly, load the investigation with people likely to be biased in favor of the accused, and so on. Similarly, there's been a tendency to blame down: to assign fault to the more junior people involved. This is a symptom, not of inherent meanness in senior administrators, but of a lack of adequate arrangements for handling the problem. Indeed, I am tempted to propose a rule of thumb: in order to find defective institutions, look for the wrecked careers of junior professionals.

Closely associated with this tendency has been a tendency to rely on assumptions of psychopathology as a convenient and automatic explanation. If something's gone wrong, it must be because the person who did it is mentally or morally defective; there's nothing wrong with the system of institutions in which the behavior took place, even if similar behavior occurs again and again. Judson is particularly scathing about this rhetorical stance (which, he points out, is never supported by actual data). His attack on the misuse of psychological pathology as an administrative defense strategy is itself a public service.

Over the last generation, we have begun to develop standard procedures which address the problem of recognizing misconduct, protecting the rights of the accused, accounting for the sponsor's money, protecting the integrity of the larger research process by withdrawing questionable results, and dealing with all the other contingencies that arise in cases of suspected misconduct. These arrangements are dominated by concern with appropriate expenditure of the Federal government's money. For this reason, they are fixed on regulating grant recipient organizations. The pressure to develop appropriate codes of ethics monitored by organizations of scientists themselves has been relatively slight, and few serious efforts along these lines seems to have been mounted. There doesn't seem to have been any attempt to develop other ways of dealing with better definitions of misconduct, ways of detecting it, and administering accusations of it. We are better off than we were, but only narrowly so, and the risks and costs of having the Secret Service second-guessing the administration of technical procedures (as happened in the Imanishi-Kari case) are yet to be determined.

We have a basic policy problem masquerading as a matter of definition. If the problem is to detect and prevent criminal activity and protect the public's tax money, then the way to prevent and detect fraud is to institute the kinds of checks and balances which accountants have invented over the years-- the equivalent of double-entry bookkeeping, and all the apparatus that goes with it. Because of the nature of the work, such systems would look rather more like the arrangements for protecting chain of evidence in criminal procedures than they would resemble a set of ledgers, although of course, a system of independent cross-referencing records would be at the heart of the process. This would help with certain kinds of fraud (not many). It might, for example, notice that a chart or table was being used in multiple publications, although this is, of course, not *per se* fraudulent. This approach to fraud is irrelevant to the strategy of using robustness-checking by the scientific community. Indeed, to the extent that such an approach leads to the defensive elimination of materials which might be "discoverable" in forensic proceedings, it is incompatible with robustness-checking and hence counter-productive

Robustness-checking relies on scientists noticing anomalies and the trouble-shooting them. It assumes that fraud is rare, and that most of its ill effects will be avoided by detecting, and then discounting the anomalous results. Auditing, by contrast, assumes that virtually every scientist and technician will be a miscreant if not discouraged and prevented, that fraud is relatively common, and that cross-checking of the same results is adequate to uncover offenses reliably. It also assumes that the important thing is to remove the offender and perhaps recover costs; setting the record straight is an afterthought. Hence, robustness-checking seeks to understand the order of the world, and considers human defalcation from this process as another source of error. Accounting procedure, on the other hand, seeks to control human defalcation, and is indifferent to the order of the world.

Systematic audit of research organized on the forensic model would cost a great deal. Independent technical review of notebooks would require auditors to have significant amounts of scientific skill. But, as Judson notes, the research system is already heavily burdened by the costs of prospective proposal reviewing and publication refereeing. Adding auditing costs to this burden as well implies a large reduction in the amount of resources available for new research. Estimating the cost/benefit trade-offs of such a policy would be very difficult, but it's hard to imagine a system in which large amounts of auditing would be justified.

Moreover, it seems clear that the major cost of fraud in science is not that of wasted money. Much more important is the consequences of fraud for the course of research. Bad data enters the system and confuses work down the line; good models may be rejected and bad models accepted. Careers of promising and productive by-standers (and whistle-blowers) are severely injured or ruined. Potentially productive lines of research may become tainted, so that talent enters other fields.

These sorts of failures cannot be dealt with by simple adoption of institutional arrangements developed for forensic issues. The technical problems of detection are solved by encouraging robustness-checking, and this work should be located in the work of the disciplines.

Once a potential problem of fraud is revealed, then the appropriate forensic apparatus can reasonably be called into play. The organizational and administrative problems should be solved by suitable procedures that give incentives first to practices that prevent and detect fraud, and that work to correct and minimize the damage when it occurs.

REFERENCES

Babbage, Charles. 1830. *Reflections on the Decline of Science in England, and Some of Its Causes*. London: B. Fellowes. (http://www.gutenberg.org/etext/1216)

Benkler, Y. 2002. "Coase's Penguin, or, Linux and The Nature of the Firm". *Yale Law Journal* 112: 1 - 72.

Chang, H. 2004. *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press.

Collins, H. M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Beverly Hills, CA: Sage Publications.

Golan, T. 2004. *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America*. Cambridge, Mass: Harvard University Press.

Hollinger, D. 1983. "The defense of democracy and Robert K. Merton's formulation of the scientific ethos". *Knowledge and Society* 4: 1 -1 5. Reprinted in *Science, Jews, and Secular Culture: Studies in Mid-Twentieth Century American Intellectual History*. Princeton: Princeton University Press, 1996.

Hull, D. L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Merton, R. K. 1942. "Science and technology in a democratic order". *Journal of Legal and Political Sociology* 1: 115-126. Reprinted as "The normative structure of science", pp. 267-278 in *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Mirowski, P. 2004. "Why econometrician don't replicate (although they do reproduce)". Pp. 213-228 in (Ed.), *The Effortless Economy of Science?* Durham, N.C: Duke University Press.

Neuliep, J. 1991. (Ed.) *Replication Research In The Social Sciences*. Newbury Park, CA: Sage Publications.

Shapin, S. 1994. *A Social History of Truth: Civility and Science in Seventeenth Century England*. Chicago: University of Chicago Press.

Shapin, S., and S. Schaffer. 1985. *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press.

Zammito, J. H. 2004. *A Nice Derangement of Epistemes: Post-Positivism in the Study of Science From Quine to Latour*. Chicago: University of Chicago Press.