CHEMTECH, **1991**, *21*, 274-280.

ETHICS IN SCIENCE

George M. Bodner
Department of Chemistry, Purdue University

Introduction

Sir Peter Brian Medawar once described scientists as follows [1].

> Scientists are people of very dissimilar temperament doing very different things in very different ways. Among scientists are collectors, classifiers, and compulsive tidiers-up, many are detectives and many are explorers. There are poet-scientists and philosopher-scientists, and even a few mystics.

He later said that he was sorry he had not added: "... and a few crooks" [2].

Questions about the frequency of unethical activities by scientists have been raised in publications directed toward the general population [3], the scientific community [4,5], and specialists in the history and philosophy of science [6]. Are unethical activities so common that they threaten the scientific enterprise? Or are they so rare they are routinely discovered, and attract comment primarily because of their rarity? Has fraud and other unethical activities been a component of science throughout history? Or is there something about the nature of modern science that fosters such activities?

Answers to most of these questions are beyond the scope of this article. One, however, can be answered explicitly. There is nothing new about unethical activities in science. More than 150 years ago, Charles Babbage proposed a classification system for scientific fraud [7]. His first category was **hoaxing**, "... deceit [which] is intended to last for a time, and then be discovered, to the ridicule of those who have credited it." Babbage classified **trimming** as "... clipping off little bits here and there from those observations which differ most in excess from the mean, and in sticking them on to those which are too small." **Cooking**, on the other hand, was defined as follows [7].

> An art of various forms the object of which is to give ordinary observations the appearance and character of those of the highest degree of accuracy. One of its numerous processes is to make multitudes of observations, and out of those select those only which agree, or very nearly agree. If a hundred observations are made, the cook must be very unlucky if he cannot pick out fifteen or twenty which will do for serving up.

The most serious category in Babbage's hierarchy was **forging**, which was done by "... one who, wishing to acquire a reputation for science, records observations which he has never made."

For our purposes, it will be useful to classify unethical activities in three categories: (1) fraud, (2) embellishment, and (3) self-deception. Examples of all three categories will be presented. Particular attention will be paid to last category, however, because we can consciously avoid the sins of fraud or embellishment and still be trapped by self-deception.

Fraud

Fraud can be defined as an act of malicious, intentional dishonesty. Its most common forms are the forging of data and plagiarism. The case of William T. Summerlin [2, 8-10] provides an example of the first. Shortly before 7:00 AM on 26 March 1974, in an elevator on the way to his mentor's office, Summerlin used a felt-tip pen to paint the skin of two mice which were to be used as evidence of a successful skin graft between different strains of mice.

Summerlin later explained his action as a spur-of-the moment impulse, which occurred when an overworked scientist was pressed by his supervisor for confirmatory results. There is reason to believe, however, that it was not an isolated instance [11]. There is less doubt about the case of John Darsee [12-13], who apparently forged data throughout his career. When several of his co-workers observed Daress faking data while he was an NIH fellow at Harvard Medical School, the work he had done previously at Emory University was subjected to internal review. Eight of the ten papers he had published before coming to Harvard had to be withdrawn or corrected, one was judged to be fictitious, and only two out of 45 abstracts of papers presented at conferences stood up to scrutiny.
Mark Spector [14-17] was described by colleagues as one of the most brilliant young people they had ever met; a prodigy with golden hands. Within months of his arrival in Ephraim Racker's laboratory at Cornell, Spector isolated a series of four kinase enzymes involved in a cascade mechanism that eventually phosphorylated the β-subunit of the ATPase enzyme which pumps sodium through the cell membrane. Similarities between one of these kinase enzymes and the product of the transforming gene of the Rous sarcoma virus led Spector to propose a hypothesis in which this cascade mechanism enabled a virus to transform animal cells into tumor cells.

Attempts to repeat parts of Spector's work failed, as did experiments that extended his hypothesis. The most telling indictment of his data was obtained by one of his co-authors, who discovered ^{125}I on a radioautographic gel that should have been labeled with ^{32}P. Radioactive iodine can be incorporated into a protein by a simple chemical reaction; radioactive phosphorus requires an enzyme-catalyzed reaction. The only possible explanation was that Spector had manipulated his sample to get the desired results. Post-hoc analysis of work done before he came to Cornell suggested that Spector's tendency to fabricate data did not begin when he became a graduate student.

Reported examples of plagiarism in science are rare, with the exception of the well-publicized case of Elias A. Alsabti [18-19]. Alsabti's vita of more than 60 papers has been shown to include two reviews that were direct copies of material from a grant application written by one of his employers; a paper copied from a manuscript submitted for publication in another journal; and several papers that were verbatim copies of papers published in other journals, with changes in only the title of the paper and the list of authors.

EMBELLISHMENT

The term "embellishment" comes from the Old French stem meaning "to make beautiful." It can be defined as the process of improving a story by adding touches or details, often of a fictitious kind. Actions that Babbage would classify as hoaxing, trimming, or cooking might all fall with the category of embellishment.

A hoax, when properly carried out, lacks the malicious character of fraud [7]. The most famous scientific hoax was apparently carried out by Charles Dawson and Pierre Teilhard de Chardin [20-23]. In 1912, Dawson and Arthur Smith Woodward announced the discovery of fragments of a skeleton in a gravel pit in Piltdown that provided evidence of the long sought after "missing link" in the evolution of man. Piltdown Man, or *Eoanthropus dawsoni*, possessed a mandible like that of an ape and a cranium like that of a human skull.

In his laboratory at the British Museum, Smith Woodward carefully pieced together these two fragments, filling in the missing space with modeling clay. Evidence for the legitimacy of Piltdown Man was obtained in 1913 when Teilhard found one of the missing teeth, which had a remarkable similarity to the canine tooth Smith Woodward had reconstructed with modelling clay the year before. Further evidence was provided by the discovery of another specimen of Piltdown Man at a second site two miles distant.

It was not until 1953 that analysis of the original Piltdown Man specimens showed that they were a hoax. The cranium was indeed a human skull which had been stained with chromate so that it looked older. But the mandible was the jaw of an orangutan, and the canine tooth showed obviously evidence of having been crudely filed and then stained with Vandyke brown.

Piltdown Man was a hoax, however, that went too far. Teilhard returned to his native France in 1914 to join the French army at the outbreak of World War I, and Dawson suddenly became ill and died in 1916. By the time the war was over, major figures in British anthropology and paleontology had staked their careers on the legitimacy of Piltdown Man. When the hoax was propagated, Teilhard was an amateur, who could see the humor in playing a joke on gullible professionals. After the war, he was a professional paleontologist who was in no position to confess his role in the hoax.

Mendel's study of the inheritance of genetic traits in peas might provide an example of what is undoubtedly a common phenomenon in the scientific literature: the trimming of data. Fisher has shown that Mendel's data were systematically "sophisticated," most likely by an overenthusiastic assistant [24]. This trimming is so famous it was described as follows, under the heading "Peas on Earth" [25].

> In the beginning there was Mendel, thinking his lonely thoughts alone. And he said: "Let there be peas," and there were peas and it was good. And he put the peas in the garden saying unto them "Increase and multiply, segregate and assort yourselves independently," and they did and it was good. And now it came to pass that when Mendel gathered up his peas, he divided them into round and wrinkled, and called the round dominant and the wrinkled recessive, and it was good. But now Mendel saw that there were 450 round peas and 102 wrinkled ones; this was not good. For the law stateth that there should be only 3 round for every wrinkled. And Mendel said unto himself "Gott in Himmel, an enemy has done this, he has sown bad peas in my garden under the cover of night." And Mendel smote the table in righteous wrath, saying "Depart from me, you cursed and evil peas, into the outer darkness where thou shalt be devoured by the rats and mice," and low it was done and there remained 300 round peas and 100 wrinkled peas, and it was good. It was very, very good. And Mendel published.

The most carefully studied example of cooking is the work of Cyril Burt [26-30]. Burt is best known for his work on the relationship between heredity and intelligence which was purported to be the largest collection of IQ data of its kind. Leon Kamin was the first to notice ambiguities and oddities in Burt's papers. The centerpiece of Burt's data was a study of the intelligence of identical twins who were separated at birth. Since identical twins have the same genes, any difference in their intelligence must reflect differences in the environment in which they were raised. In theory, the correlation of their IQ scores therefore gives a direct measure of the influence of heredity on intelligence.

Burt published his first report on identical twins in 1955, when he had located 21 pairs. A second report was published in 1958, when the collection included "over 30" pairs. A final report, in 1966, contained 53 pairs. In all three studies, the correlation between the IQ scores of the separated twins was the same, 0.771. Furthermore, the correlation between the IQ of identical twins who were reared together was 0.944, regardless of sample sizes. No fewer than 20 instances have been found in Burt's papers in which correlation coefficients remain constant in spite of changes in sample size.

Several explanations have been offered for these remarkable coincidences. His supporters have argued: (1) most of the data were published after Burt retired, when he was ill, and therefore careless with his data; (2) to avoid the chore of recalculating the correlations with the new data, Burt simply carried over the correlations from earlier papers; (3) as he got old he remembered correlation coefficients that stuck in his mind from earlier papers; or (4) he was so convinced of the correctness of his views that he became careless with the data. A less generous explanation has also been offered: Burt concocted data from the beginning of his career and became less careful at covering his tracks as he grew older.

It should be noted that the papers which contain the questionable data list coauthors whose existence has been debated. These coauthors also appeared frequently as book reviewers for the British Journal of Statistical Psychology while Burt edited this journal. Their reviews were written in a style very much like Burt's, tended to praise Burt's work, and attacked his opponents. It should also be noted that Burt rarely reported his data, but referred to unpublished reports which are as elusive as his co-authors.

It is legitimate to ask whether Burt's activities differ in either magnitude or intent from those described in the previous section. His supporters note that deleting Burt's data from the literature has no appreciable effect on conclusions about the inheritance of intelligence [29]. In my mind, however, the following quotation irrevocably indicates the severity of his actions [3, p. 204].

> [Burt] was fond of accusing his opponents of basing their criticisms "not on any fresh evidence or new researches of their own, but chiefly on armchair articles from general principles. My co-workers and I, on the other hand, were engaged in on-going research."

SELF-DECEPTION

Fraud and embellishment involve conscious acts of deception that are clearly unethical, regardless of how common they might be. Self-deception is a very different phenomenon. It can be defined as behavior that results from unintentional violations of the normal practices of science. It is tempting to view fraud or embellishment from a self-righteous perspective, as traps into which we would never fall. Analysis of incidents of self-deception, however, shows that this is a naive view of ethical practice in science.

There are abundant examples of scientists who observed what they expected (or hoped) to find. Gould provided one in his analysis of the work of Samuel G. Morton [31]. Between 1830 and 1851, Morton amassed the world's largest collection of skulls in order to replace speculation about differences among the races with hard evidence. Over a period of years, Morton used mustard seed and then lead shot to measure the cranial capacity of 623 skulls, because he believed it was the best index of overall intelligence.

Morton's three major works on the cranial capacity of the human races each contained a summary table that showed the average internal skull volume of the different races. The first version of this table showed that Caucasians (87 in^3) had a larger average skull volume than Mongolians (83 in^3), American Indians (82 in^3), Malay (81 in^3), or Blacks (78 in^3). This table and subsequent versions were used repeatedly during the 19th century as irrefutable evidence of the relative mental worth of the human races.

When Gould recalculated Morton's data, he found no significant difference between the volume of Native American (86 in^3), Mongolian (85 in^3), Caucasian (85 in^3), Malay (85 in^3), or African (83 in^3) skulls in Morton's sample. He concluded that Morton unconsciously finagled his data to obtain the results he expected.

> Yet, through all this juggling, I find no indication of fraud or conscious manipulation. Morton made no attempt to cover his tracks, and I must assume that he remained unaware of their existence. He explained everything he did, and published all his raw data. All I discern is an a priori conviction of racial ranking so powerful that it directed his tabulations along preestablished lines. Yet Morton was widely hailed as the objectivist of his age, the man who would rescue American science from the mire of unsupported speculation [31].

Morton was not alone in his belief that there was a direct relationship between the size of the skull and the intelligence of the individual [32]. Paul Broca was quoted as follows [33, p.304].

> In general, the brain is larger in mature adults than in the elderly, in men than in women, in eminent men than in men of mediocre talent, in superior races than in inferior races.

Nor was Morton alone in adjusting experimental data to fit this theory. When told that analysis of the brain sizes of five eminent professors at the University of Gottingen gave results close to average, Broca suggested they must not have been so eminent after all [33, p.165]

The most telling example of self-deception was the N-ray affair. A popular reference to this controversy starts with the following statement by R. W. Wood [34].

> In the late autumn of 1903, Professor R Blondlot, head of the Department of Physics at the University of Nancy, member of the French Academy, and widely known as an investigator, announced the discovery of a new ray, which he called the N ray, with properties far transcending those of the x-rays. Reading of his remarkable experiments, I attempted to repeat his observations, but failed to confirm them after wasting a whole morning.

The remainder of this quotation is a delight to read, but a better understanding of the nature of the controversy can be obtained from the following truncated version of the paper in which Wood reported his experience in Blondlot's laboratory [35].

> The inability of a large number of skilful experimental physicists to obtain any evidence whatever of the existence of the n-rays, and the continued publication of papers announcing new and still more remarkable properties of the rays, prompted me to pay a visit to one of the laboratories in which the apparently peculiar conditions necessary for the manifestation of this most elusive form of radiation appear to exist. I went, I must confess, in a doubting frame of mind, but with the hope that I might be convinced of the reality of the phenomena, the accounts of which have been read with so much scepticism.
>
> After spending three hours or more in witnessing various experiments, I am not only unable to report a single observation which appeared to indicate the existence

of the rays, but left with a very firm conviction that the few experimenters who have obtained positive results have been in some way deluded.

The first experiment which it was my privilege to witness was the supposed brightening of a small electric spark when the n-rays were concentrated on it by means of an aluminum lens. The spark was placed behind a small screen of ground glass to diffuse the light, the luminosity of which was supposed to change when the hand was interposed between the spark and the source of the n-rays.

It was claimed that this was most distinctly noticeable, yet I was unable to detect the slightest change. This was explained as due to a lack of sensitiveness of my eyes, and to test the matter I suggested that the attempt be made to announce the exact moments *at which I introduced my hand* into the path of the rays, by observing the screen. In no case was a correct answer given, the screen being announced as bright and dark in alternation when my hand was held motionless in the path of the rays, while the fluctuations observed when I moved my hand bore no relation whatever to its movements.

I was next shown the experiment of the deviation of the rays by an aluminum prism. The aluminum lens was removed, and a screen of wet cardboard furnished with a vertical slit about 3 mm. wide put in its place. In front of the slit stood the prism, which was supposed not only to bend the sheet of rays, but to spread it out into a spectrum. The positions of the deviated rays were located by a narrow vertical line of phosphorescent paint, perhaps 0.5 mm. wide, on a piece of dry cardboard, which was moved along by means of a small dividing engine. It was claimed that a movement of the screw corresponding to a motion of less than 0.1 of a millimetre was sufficient to cause the phosphorescent line to change in luminosity when it was moved across the n-ray spectrum, and this with a slit 2 or 3 mm. wide. I expressed surprise that a ray bundle 3 mm. in width could be split up into a spectrum with maxima and minima less than 0.1 of a millimetre apart, and was told that this was one of the inexplicable and astounding properties of the rays. I was unable to see any change whatever in the brilliancy of the phosphorescent line as I moved it along, and I subsequently found that the removal of the prism (we were in a dark room) did not seem to interfere in any way with the location of the maximum and minima in the deviated (!) ray bundle.

I then suggested that an attempt be made to determine by means of the phosphorescent screen whether I had placed the prism with its refracting edge to the right or the left, but neither the experimenter nor his assistant determined the position correctly in a single case (three trials were made). This failure was attributed to fatigue.

I was next shown an experiment of a different nature. A small screen on which a number of circles had been painted with luminous paint was placed on the table in the dark room. The approach of a large steel file was supposed to alter the appearance of the spots, causing them to appear more distinct and less nebulous. I could see no change myself, though the phenomenon was described as open to no question, the change being very marked. Hold ing the file behind my back, I moved my arm slightly towards and away from the screen. The same changes were described by my colleague. A clock face in a dimly lighted room was believed to become much more distinct and brighter when the file was held before the eyes, owing to some peculiar effect which the rays emitted by the file exerted on the retina. I was unable to see the slightest change, though my colleague said that he could see the hands distinctly when he held the file near his eyes, while they were quite invisible when the file was removed. The room was dimly lighted by a gas jet turned down low, which made blank experiments impossible. My colleague could see the change just as well when I held the file before his face, and the substitution of a piece of wood of the same size and shape as the file in no way interfered with the experiment. The substitution was of course unknown to the observer.

> I am obliged to confess that I left the laboratory with a distinct feeling of depression, not only having failed to see a single experiment of a convincing nature, but with the almost certain conviction that all the changes in the luminosity or distinctness of sparks and phosphorescent screens (which furnish the only evidence of n-rays) are purely imaginary. It seems strange that after a year's work on the subject not a single experiment has been devised which can in any way convince a critical observer that the rays exist at all.

SOURCES OF UNETHICAL CONDUCT

An important source of unethical conduct was captured many years ago in a note entitled "Reinvestigation of the so-called bis (triphenyllead)mercury" [36].

> We greatly regret the inaccuracies in our earlier publication. They arose from the imaginative contributions of a junior assistant, who is no longer with us.

A recurring theme in any analysis of unethical actions by scientists is fraud which results from the work of junior colleagues. (Particularly those who have not been carefully supervised.)

This paper provides the basis for understanding another important factor in scientific fraud. With the exception of the case of Alsabti (and possibly Darsee) there is a common thread that binds together all of the incidents described in this paper. In virtually every case, one or more scientists fell into the trap of seeing what they expected to see.

The photograph on the back cover of Kohn's book [4] shows one of the mice ("Old Man") on which Summerlin carried out a successful skin transplant before he arrived at Sloan-Kettering. Success with this animal may have led Summerlin to believe in the efficacy of his tissue culturation technique in spite of limited or no success in later experiments. The results with "Old Man" were deceptive, however, because this mouse was later shown to be a hybrid (F_1) of the C3H donor strain and the white A acceptor strain, not a pure bred C3H mouse. As a result, the skin graft was much less likely to be rejected.

After he was confronted with evidence that he had forged the results of some of his experiments, Spector successfully repeated some but not all of his experiments under the careful supervision of Ephraim Racker. There is reason to believe that success with early experiments led Spector to believe in the validity of his cascade mechanism to the point that he was willing to forge experimental data to prove what he believed was correct.

The key to any successful hoax — a category into which the Piltdown Man undoubtedly belongs — is self-deception on the part of those who fall for the hoax. In this case, the evidence provided by the specimens was exactly what gullible scientists of the time hoped to see. Evidence of this is provided by the following quotation from the work of a contemporary of Dawson and Teilhard [22].

> If there is a Providence hanging over the affairs of prehistoric men, it certainly manifested itself in this case, because the three fragments of the second Piltdown man found by Dawson are exactly those which we would have selected to confirm the comparison with the original type. ... Placed side by side with the corresponding fossils of the first Piltdown man they agree precisely; there is no shadow of difference.

Neither Mendel, Broca, Morton nor Blondlot were involved in malicious, intentional dishonesty. Each was guilty, however, in seeing in his experimental data what he expected (hoped?) to see. Some would argue that the same cannot be said of Cyril Burt; that he took a fraudulent approach to data throughout his career. I am tempted to disagree. The fact that elimination of his data makes no effect on the overall conclusions about the heredity of intelligence suggests that at least some of his initial data were honestly collected and interpreted. From that point, however, he was so convinced of the truth of his expectations that he could no longer either adequately judge the validity of data he collected or resist the temptation to improve these data.

I am indebted to Alexander Kohn [4] for bringing my attention to the work of Irving Langmuir on pathological science — scientific studies based on non-existent phenomena [37]. The symptoms of pathological science proposed by Langmuir can be paraphrased as follows.

> The magnitude of the effect is close to the limits of what can be detected. In spite of this, claims of great accuracy are made which serve as the basis for fantastic theories that are contrary to experience. When the work is criticized, its proponents make ad hoc excuses on the spur of the moment. The ratio of supporters to critics rises to near 50% and then gradually falls to oblivion.

Most of the instances of self-deception described in this paper suffer from one or more of these symptoms.

REFERENCES

[1] Medawar, P. B. *The Art of the Soluble*, London: Methuen, **1967**.
[2] Medawar, P. B. *New York Review of Books*, April 15, 1976.

[3] Broad, W., and Wade, N. *Betrayers of Truth*, New York: Simon and Shuster, **1982**.
[4] Kohn, A. *False Prophets: Fraud and Error in Science and Medicine*, Oxford: Basil Blackwell, Ltd, **1986**.
[5] Zurer, P. S. *Chemical and Engineering News*, **1987**, April 13, 10-17.
[6] Holton, G. J. *The Scientific Imagination: Case Studies*, Cambridge: Cambridge University Press, **1978**.
[7] Babbage, C. *Reflections on the Decline of Science in England and on Some of its Causes*, London: B. Fellowes, Ludgate Street, **1830**.
[8] Culliton, B. J. *Science*, **1974**, 184, 644-650.
[9] Culliton, B. J. *Science*, **1974**, 184, 1154-1157.
[10] Hixson, J. *The Patchwork Mouse*, Garden City: Anchor Press, **1976**.
[11] McBride, G. *Journal of the American Medical Association*, **1974**, 229, 1391-1410.
[12] Culliton, B. J. *Science*, **1983**, 220, 31-35.
[13] Culliton, B. J. *Science*, **1983**, 220, 936.
[14] Newmark, P. *Nature*, **1981**, 293, 93.
[15] Newmark, P. *Nature*, **1981**, 293, 329.
[16] Stein, M. D. *Nature*, **1981**, 293, 93.
[17] Wade, N. *New Scientist*, **1981,** *91*, 781-2.
[18] Broad, W. J. *Science*, **1980**, 208, 1438-1440.
[19] *Nature*, **1980**, 285, 429-430.
[20] Wade, N. *Science*, **1978**, 202, 1062.
[21] Gould, S. J. *Natural History*, **1980**, 89(8), 8-28.
[22] Gould, S. J. *The Panda's Thumb*, New York: W. W. Norton and Company, **1980**.
[23] Mathews, L. H. *New Scientist*, **1981,** 90, 280-282, 376, 450, 515-516, 578-579, 647-648, 710-711, 785, 861-862, and 91, 26-28.
[24] Fisher, R. A. *Annals of Science*, **1936**, 1, 115-137.
[25] *Hort. Sci*. **1972**, 7, 438.
[26] Kamin, L. J. *The Science and Politics of IQ* Potomac, Md: Lawrence Erlbaum Associates, **1974**.
[27] Wade, N. *Science*, **1976**, 194, 916-919.
[28] Gillie, O. *Phi Delta Kappan*, **1977**, 469-471.
[29] Jensen, A. R. *Phi Delta Kappan*, **1977**, 471, 492.
[30] Dorfman, D. D. *Science*, **1978**, 201, 1177-1186.
[31] Gould, S. J. *Science*, **1978**, 200, 503-509.
[32] Gould, S. J. *The Mismeasure of Man*, New York: W. W. Norton, **1981**.
[33] Broca, P. *Bulletin Societe d'Anthropolgie Paris*, **1861**, 2, 304.
[34] Wood, R. W. In *A Random Walk in Science*, R. L. Weber, Ed. London: The Institute of Physics, **1973**.
[35] Wood, R. W. *Nature*, **1904**, 70, 530-531.
[36] Deganello, G., Carturan, C., and Uguagliati, P. *Journal of Organometallic Chemistry*, **1969**, 18, 216.
[37] Langmuir, I. *Speculations in Science and Technology*, **1985**, 8, 77-94.