Eight Centuries of Sampling Inspection: The Trial of the Pyx

STEPHEN M. STIGLER*

A sampling inspection scheme bearing striking similarities to modern procedures has been in more or less continuous operation at the Royal Mint in London for about eight centuries. The history of this scheme is reviewed, and a defect in the manner in which the critical levels were set is noted. The possibility that Master of the Mint, Isaac Newton might have grasped the nature of the defect is discussed. Finally, we review the situation at the United States Mint, where similar trials have long been held, but where steps to correct the defect were taken early in the nineteenth century.

KEY WORDS: History of statistics; Newton; Coinage; Quality control; Hypothesis testing.

1. THE TRIAL OF THE PYX

The construction of optimal tests of significance is a twentieth century innovation in statistical theory, but some of the concepts underlying this theory have a longer history. One early appearance of statistical tests in sampling inspection bears the mysterious name, "the trial of the Pyx," and has been in continuous operation since three-quarters of a millennium before Neyman met Pearson. The trial of the Pyx is an ancient ceremony of the Royal Mint of Great Britain. The avowed purpose of this ceremony is to ascertain that coinage issued by the Royal Mint meets the Crown's specifications, although the actual motives of the participants have varied over the centuries. The ceremony and its strange name are perhaps best explained by describing the procedure followed at a typical trial as they were held before 1870.

The trial of the Pyx was the final stage of a sampling inspection scheme for the control of the quality of the Mint's product—primarily gold and silver coinage. Over a period of time, one coin would be taken out of every journey (from journée, for day's work) of the Mint's production; i.e., one coin taken out of every 15 pounds produced in the case of gold and out of every 60 pounds in the case of silver. The coin would be placed in a box called the Pyx. Prior to the minting, a section of a trial plate of the king's gold would be safely stored in a box in a thrice-locked treasury room called the Chapel of the Pyx in Westminster Abbey for later use as a control.

The word Pyx, often spelled pix or pixe before the middle of the nineteenth century, derives from the Greek πυξις, a box. In early (before 1400) ecclesiastical literature, a pyx was the vessel in which the bread of the sacrament was reserved. It is possible that the present (at least since 1598) usage as the Mint's strongbox derives from the ecclesiastical usage and the early location of the box in Westminster Abbey.

At irregular times, perhaps as frequently as each year but usually separated by three or four years, a trial of the Pyx would be declared and a jury of members of the Worshipful Company of Goldsmiths (or Mystery of Goldsmiths, an independent tradesman's guild) would assemble. At the trial, the Pyx would be opened and the contents counted, weighed, and assayed. The results would then be compared with the standard set in the Mint's indenture. Prior to 1851, the Royal Mint was not a governmental body but operated under a contract or indenture to the Crown, and the indenture specified that a tolerance, called the remedy, above and below the standard would be allowed, the amount depending upon the type of metal and the denomination of the coin being tested. After a successful trial a banquet would be held in celebration.

The antiquity of the ceremony of the trial of the Pyx is well documented. Chisholm (Great Britain 1866) suggests that the trial may date back to the reign of Henry II, 1154–89, although Craig (1953) thinks that so early an origin is unlikely and argues that coinage inspection by the Barons of the Exchequer in 1179 was intended more to determine the values of tax contributions than to verify the standard of the minters' work. Also, while the principal source for information about pre-1179 trials, the *Dialogus de Scaccario* by Richard Fitz Neal (or Fitz Nigel, son of Nigel of Ely), was written in 1176–79, it is only known through sometimes erroneous or altered later transcripts. In any case, a public trial was held in 1248, and all authors are agreed that the trial of the Pyx was well established by 1279 when Edward I issued a proclamation describing the procedure to be followed. Many particulars of this procedure have changed over the centuries but the basic premise has not: Since the time of Edward I, even to the present

* Stephen M. Stigler is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. This research was supported by the National Science Foundation under Grant soc 75-02922. The author is grateful to F.H. MacDonald, Deputy Director of the U.S. Mint, for information on the current procedures followed by the Annual Assay Commission and to G.P. Dyer, Librarian and Curator of the Royal Mint, for information on both modern and ancient trials of the Pyx and for permitting him to inspect the ancient plates.

© Journal of the American Statistical Association
September 1977, Volume 72, Number 359
Applications Section

day, the ceremony has consisted of an independent jury inspecting the selection of coins kept in the Pyx in order to determine whether or not the Mint has met the standard.

The motivation for the trial has, however, undergone some change. The trial was originally designed as a means by which the king could maintain a check on the use of ingots furnished to the independently operated Mint. If, for example, a minted sovereign contained too much gold, the king was literally not getting his money's worth, and a collector of overweight sovereigns could have them melted down and returned to the Mint at a profit. If the sovereigns tended to weigh too little, the currency would be debased and worse—the profit from this debasement would go to the Mint!

The trial of the Pyx was intended to maintain the standard and insure that any debasement of the currency was as intended by the monarch. To this end, if the trial revealed that the amount of gold in the coinage was too little, the early masters of the Mint were required to reimburse the Crown for this deficit, extrapolated to cover the entire quantity minted since the last trial. This is the reason for the word remedy, which probably began by meaning the amount that must be remedied and later came to mean the amount that would be tolerated without remedy.

In later years, particularly since 1851 when the status of the Mint was redefined as a branch of the Exchequer, a secondary purpose of the trial became primary; a principal motivation for the modern trial of the Pyx is to inspire public confidence in the coinage. Thus verdicts of the Pyx jury, which in earlier years were not publicized, have been printed with the House of Commons reports since 1870. The trial of the Pyx is not the only inspection undergone by newly minted coinage; constant checks are carried out by the Mint, and even in early times checks by the king's assayer were routine. Indeed, the selection of coins for the Pyx was only made after the coinage had passed the Mint's own inspection by the king's assayer. The trial is, however, the only such inspection performed by an independent jury.

2. THE PYX FROM A STATISTICIAN'S VIEWPOINT

From the viewpoint of a statistician the trial of the Pyx is a marvelous example of a sampling inspection scheme for the maintenance of quality. Indeed, having been in more or less continuous operation for perhaps eight centuries, it is certainly one of the oldest such schemes. One would expect that Roman coinage may have been subjected to some sort of inspection, but I'm aware of no documentation of this. The problem can be formulated in terms of the standard textbook model for a simple hypothesis testing problem. The Mint Indenture sets up a standard as a simple null hypothesis. The Crown is worried about deviations either above or below the standard thus calling for a two-sided test. Because it is uneconomical for the jury to weigh the entire coinage (and impossible to assay it all, as assaying was then a destructive process), it becomes necessary to resort to sampling. Since minting is an ongoing process, some sort of sampling and testing procedure must be implemented. A sampling scheme must be devised, an interval between tests must be decided, and a test statistical and critical region must be determined.

All of these considerations of course raise questions. How, in light of modern statistical concepts and techniques, does the ancient trial of the Pyx fare? How was the sampling performed? What test was actually used, and how was the remedy (or tolerance) determined? Was the test an appropriate one? The passage of time, and the fact that what documents do exist were usually not designed to answer these questions, makes a detailed investigation impossible. Yet some information on these questions is available, and it is to this I shall now turn.

3. THE SAMPLING SCHEMES

How was a coin selected for the Pyx; was the sampling done randomly? Since not all coins in the Pyx were assayed (or, in the early years, even weighed), how were the coins selected from the Pyx for testing? These questions are very difficult to answer, as the exact procedures have not been set down in all cases. Indeed, we should not expect a mathematically exact description of the selection procedure, as the concept of a random sample is a modern one, and an exact description of even a nonrandom sampling scheme almost presupposes the notion of a random sample.

Official documents relating to the Pyx state the number of coins to be selected for the Pyx, but say nothing about how this selection should be carried out. The Mint Indenture of the 18th year of the reign of Edward II, 1345, states that the warden of the Mint shall

> take out of every *C* pound w^eit i j^a starlinge, and of every five pound w^eit of gold one peece, which peece shall be kepte in one chest with two keys, and sealed with two seales, th' one to remayne with the king's deputie, and the other with the master. (Ruding 1840, p. 70).

Similarly uninformative statements survive from other reigns, down to more modern times. For example, the Mint Indenture of 1817 specifies that

> before any deliverance be of the whole sum a portion of the said monies shall be taken and put into a box . . . , (that is to say) for every journey weight of gold not exceeding fifteen pounds weight two pieces, whereof one be for the pyx and the other to be for the assay . . . (Great Britain 1866, p. 30).

The assay referred to was an in-house assay called the Mint Pyx Assay, a procedure that was accomplished much more frequently and less formally than the trial of the Pyx (Great Britain 1866).

A naive view would be to suppose that as coins which were eligible for the Pyx had already passed various Mint inspections and been judged fit for circulation, they must then have been indistinguishable under a casual inspection, and the selection, therefore, must have effectively been random. That this may have been the case in recent times is plausible, but it has not always been true. For example, Craig (1953, p. 104) notes that at the trial of the Pyx in 1534 the jurors reported that the coins were sufficiently uneven in weight that it would be profitable to cull the heavier pieces, and in the mid-1600's overweight coins were nicknamed "come again guineas" by the Mint because they would be returned to be remelted and reminted at profit to the receiving merchant (Craig 1953, p. 212). While this suggests the feasibility of a biased selection, it still leaves us in the dark as to whether or not the selection was practiced in such a manner as to make it a reality.

Whatever may in fact have been the case, in the nineteenth century the public perceived the selection as more or less at random. An 1815 account of the 1814 trial of the Pyx in the *Gentleman's Magazine* describes the coins in the Pyx as having "been previously taken at hazard for this Trial" (R. 1815). An 1892 account in *Chambers's Journal* describes the coins as having been taken out from each journey weight "impartially" (unsigned 1892).

Some further light is cast on the matter, at least on the way the selection was viewed by Mint officials, in nineteenth century reports to the House of Commons. In 1866, H.W. Chisholm, then Chief Clerk of the Exchequer and later father of the mathematician Grace Chisholm Young, suggested a decrease in the number of coins 'pyxed,' and wrote:

> It appears, therefore, wholly unnecessary to place so many coins in the Pyx box, and that it would be quite sufficient to take indiscriminately one coin out of the whole number coined during the day at the Mint, one denomination only being coined in a single day, and to place this single coin in the Pyx box (Great Britain 1866, p. 7).

Chisholm's recommendation was based on the fact that only about one in 200 Pyx coins were actually examined closely, rather than on a mathematical analysis of the appropriate sampling fraction. At another point Chisholm describes the Pyx coin as "a specimen" of the melting (Great Britain 1866, p. 20).

It then appears that all parties viewed the selection as impartial, even if this may not have been guaranteed by a mathematically sound randomization in the modern sense. Further information on the manner in which sampling may have been accomplished for the Pyx is furnished by 1837 testimony to a Select Committee of Commons on the manner in which sampling was done for assay within the trial of the Pyx. J.W. Morrison was, in 1837, the Deputy Master of the Mint, in the 34th year of the 47 he filled that position. The members of the Select Committee questioned him on the conduct of the trial, particularly as regards the portion of the trial devoted to assaying the fineness of the trial coins, and the following exchange occurred:

> Question: Is a trial made of every piece?
> Answer (Morrison): The whole is weighed and tried, and from the heap they promiscuously take a certain quantity to melt; they melt it into a bar or ingot, a piece is then cut off for the assayer to try; so that it is an average of the whole.
> Question: Under that trial some of the pieces might be defective?
> Answer: They are all mixed up in a large heap, and taken out by chance.
> Question: They might be on average, but some of the pieces might be defective?
> Answer: I cannot tell how that should happen. They might happen to take those very defective pieces. The assayer runs the risk of that.
> Question: Each individual piece is not assayed, but a certain number from the heap taken out by chance and melted, and the average trial made?
> Answer: Yes.
> Question: Supposing the process of melting was improperly performed, would not that affect the result of the assay subsequently made?
> Answer: Of course it would if it was improperly made, but we are to suppose the assayer is a skilful person; we are all looking on at the same time, to see it is melted in the usual way.

Earlier in his testimony, Morrison had described the selection of coins for the Pyx in the following terms: "...the comptroller takes from those bags, given to him promiscuously, two pieces, one for the King's assayer, and one for the public trial of the pix" (Great Britain 1837). In light of Morrison's later testimony it does not seem unreasonable to suppose that, at least in the early eighteenth century, the selection was done more or less blindly, certainly without prolonged study of the pieces.

Further evidence on this point, and on the remarkable persistence of British tradition, can be found in a surviving manuscript dating from about 1280. The manuscript describes the selection of coins from the Pyx for assay:

> When the Master of the Mint has brought the pence, coined, blanched and made ready, to the place of trial, e.g. the Mint, he must put them all at once on the counter which is covered with canvas. Then, when the pence have been well turned over and thoroughly mixed by the hands of the Master of the Mint and the Changer, let the Changer take a handful in the middle of the heap, moving round nine or ten times in one direction or the other, until he has taken six pounds. He must then distribute these two or three times into four heaps, so that they are well mixed (Oresme 1956, p. 91).

The timing of the trial of the Pyx has been changed frequently in its long history. The indentures of Edward III and of Elizabeth I specified that the trial should be held every three months, although these schedules were not taken seriously for long. The Coinage Act of 1870, which contained the first statuatory recognition of the trial of the Pyx, specified that the trial should be held annually. However, throughout the long history of the trial, the timing was more haphazard. Sometimes a variant on a fixed sample-size procedure was employed—the trial was held when the Pyx could hold no more coins. Sometimes a political motive seems to have lay behind the king's call for a trial. But often, the trial was held to coincide with a change in the directorship of the Mint. In the years around 1800, a new Master of the Mint was required to post a £20,000 bond, and this bond was not released upon his departure from the Mint until a trial of the Pyx was successfully completed.

4. THE TRIAL AND THE REMEDY

As we have seen, it appears that some semblance of randomness seems to have occurred in the selection of coins for the Pyx. What can be said of the use to which this sample was put? A complete evaluation of the procedures used would require facts about the ancient coinage of Great Britain that were not recorded in Mint reports of the time, but sufficient information does exist to advance some tentative judgments. In brief, it appears that before 1870 the trial was conducted in such a manner as to be rather more favorable to the Mint than may have been intended. Had the king's councillors possessed a greater knowledge of probability theory, a more exacting trial might have been required by the Mint Indenture.

From the earliest times the trial was charged with testing three measurements of the coins in the Pyx: the count (called the tale), the weight, and the fineness. The checking of the tale seems to have been a formality which was done to provide a basis for judging the weight. From the earliest times, the master of the Mint was allowed a tolerance or remedy with respect to both weight and fineness; Ruding (1840, p. 69) lists the remedies as far back as the reign of Edward I, 1272–1307. The size of the remedies varied slightly over the centuries, only decreasing substantially in 1815. The details of procedure in these earliest trials are not all reported, but apparently the coins of the Pyx would be tested to see if they were within the remedy. If they were, the master of the Mint was expected to make restitution to the Crown for any implied deficit; if they were not, the master's future was somewhat in doubt. As a c. 1280 treatise put it, "the Master will be at the prince's mercy or will in life and members" (Oresme 1956, p. 80).

The Indenture of 1345 stipulates,

> The said box to bee opened once every three monthes before the councell of the kinge, the warden, and the master, and the said moneys to bee assaid before them, and being found good and covenable, the said master to have letters pattents for his discharge; and beinge found otherwyse, the master to pay the kinge or his deputy that which shall apperteyne ... (Ruding 1840, p. 70).

Apparently, after about 1550, restitution for deficits within the remedy was often not required but allowed to the master in lieu of an adequate salary (Great Britain 1848, p. 41).

Curiously, in the long history of the early trials of the Pyx there are only two recorded instances (both before 1550) where the remedy was not met. We shall see why this occurred if we look to a later time, where more detailed descriptions of procedure exist. By 1799, the trial procedure had become set much as described by Morrison in his 1837 testimony. We shall describe the testing of the gold coins as it was done in 1799 (Ruding 1840, p. 75–76).

The Pyx was opened and the gold coins counted. Over a four-year period, 10,748 gold coins in three denominations had been accumulated, with a face value of £8,914 13 s. 6d. Their total weight was found to be 190 lb 9 oz 8 dwt. (In troy weights, 24 gr = 1 dwt, 20 dwt = 1 oz, 12 oz = 1 lb; thus 5,760 gr = 1 lb.) According to the standard in force at the time they ought to have weighed 190 lb 9 oz 9 dwt 15 gr, thus they were in deficit 1 dwt 15 gr. However, the remedy for this quantity of coin being, in 1799, 1 lb 3 oz 18 dwt; they were well within the remedy. The assay trial was less perfectly described, but it was based on an assay by fire of a sample of 59 coins of face value £46 14 s 6d, compared to an assay of trial pieces produced by an officer of the Mint.

The key to the outcome of the trial, and the key to the long run of the Mint's successes, was the manner in which the remedy for the given weight of gold was calculated. For the indenture specified that the remedy would be $\frac{1}{6}$ of a carat, or 40 gr per lb. The fact that the Crown specified the remedy on a per pound basis and the fact that all the coin weights were combined, together with the central limit theorem, guaranteed that no reasonably careful master of the Mint could exceed the remedy!

To substantiate this claim that the master's position was a remarkably safe one, we need to know more about the distribution of weights of gold coins in this period. The only such evidence which seems to be avaiable is a Mint study performed in 1848 at the request of a House of Commons Select Committee. By 1848, Mint technology had improved, and the remedy had been reduced to 1/20 of a carat, or 12 gr per lb of gold, but the results are informative nonetheless. At the Committee's request, a clerk at the Bank of England weighed 10,000 gold sovereigns. His description of his experiment contains the earliest use of the word random I have encountered in the literature of the Mint trials, although it should not be read here in the modern, technical sense.

> I have taken for this trial, at random, four bags of sovereigns containing 1000 each from deliveries as near to standard as I could find; three bags from a delivery upon which there was a gain by tale; and three bags from a delivery upon which there was a loss; altogether 10,000 (Great Britain, 1848, p. 218).

His sampling method was defective in that it seems unlikely to be even a stratified truly random sample of bags, and we are not informed of the relative frequency of occurrence of the three types of bags. However, it is apparent from his data (which he presents by bag, in grouped form, see Table 1) that the variation in distribution from bag to bag was not so great as to invalidate his results. He found that 454 of the 10,000 sovereigns were above or below the legal remedy as calculated from the per pound basis for a single sovereign. Allowing for round off, he took the figure to be 5 percent of the sovereigns weighed. Incidentally, probability plots of the grouped data exhibit no pronounced departure from normality; indeed, the plots are surprisingly straight, and anticipated heavy-tails are not apparent (although they may be masked by the coarseness of the grouping).

Thus it would appear, in this case at least, that the remedy was specified so that about 5 percent of a repre-

1. Weights of 10,000 Sovereigns, Classified According to Deviation from Standard, in Grains

Bags	Below −R	(−R to −.2)	(−.2 to −.1)	(−.1 to 0)	(0 to .1)	(.1 to .2)	(.2 to R)	Above R	Totals
1 & 2	34	57	172	630	597	366	116	28	2,000
3	11	17	100	412	172	218	57	13	1,000
4	20	22	135	350	184	222	50	17	1,000
5	30	102	107	289	209	184	50	29	1,000
6	32	27	267	210	236	144	56	28	1,000
7	47	65	141	380	157	135	50	25	1,000
8	11	21	110	215	361	156	71	55	1,000
9	10	38	103	228	425	140	36	20	1,000
10	14	13	126	309	290	168	50	30	1,000
Totals	209	362	1,261	3,023	2,631	1,733	536	245	10,000

NOTE: R = (12/5760)·123 = .25625 is the remedy, in grains, for a single sovereign. Bags 1–4 selected as "near to standard," bags 5–7 as below, bags 8–10 as above.
Source: Adapted from Great Britain (1848, p. 220).

sentative collection of freshly minted coins would exceed it; i.e., if all coins were of the same denomination and the remedy prorated to the standard weight of a single coin, 5 percent of the population distribution would exceed it. Small wonder then that the aggregated weight of a sample of n coins did not exceed n times the remedy for a single coin! If the sampling were indeed random, $\sqrt{n}$ times the remedy for a single coin would be a more appropriate yardstick, and typically n would be more than 5,000. (In 1866 the Pyx contained over 45,000 sovereigns!) The critical value represented by the remedy might thus correspond to 200 to 300 times the standard deviation of the total weight, assuming random sampling, if the coins were in fact being minted to standard.

This circumstance would permit quite a bit of room for maneuver by a clever master of the Mint. That is, he could aim at a lower standard than that of the indenture without risk of exceeding the remedy, and pocket the difference. In early years this would have been prevented by the requirement that the master repay any deficit, but evidence (Great Britain 1848, p. 41; 1837) suggests that it may have been practiced in England when salaries were low in Elizabethan times, or even encouraged, as an indirect form of remuneration for the master, and was always a standard source of revenue for the minters in France. ("Q: Have not the French contractors an interest in coining money as near to the worse side of the remedy as they can do, provided they keep within it? A: Decidedly, and they do take advantage of that." (Great Britain 1837).) But in later years the English minters seem to have not availed themselves of this opportunity. Perhaps the best one can say about the trial of the Pyx is that by safeguarding against only the grossest abuses by the Mint, it avoided the chance of a scandal that would have shaken public confidence in the coinage.

5. NEWTON

The most illustrious master of the Mint was Isaac Newton, who served in that position from 1699 until his death in 1727, having been Warden since 1696. It is natural to ask if this great scientist might have benefited from the remedy during his tenure with the Mint, either with or without the approval of the Crown. Such evidence as exists seems to indicate that he did not. Newton's manuscripts show him to have had a conscientious concern with the integrity of the Mint's product and to have put particular emphasis on the reduction of the variation in individual pieces. He was acutely aware of the possible benefit of the remedy to importers (through their returning of overweight coins for remelting) and sought to maintain a high standard for accuracy. (Craig 1953, p. 212; 1946, p. 37).

Was Newton's grasp of statistical theory sufficient to allow him to take advantage of the remedy, had he wished? His published works have little to say about probability (Sheynin 1971), and his interpretations of data relevant to his physical theories were sometimes more influenced by his preconceptions than by statistical analysis (Westfall 1973). Nonetheless, in order to obtain some insight into Newton's understanding of the behavior of sample means, we shall examine a work he wrote while master of the Mint, where an interval estimate of a mean is presented in a situation conceptually similar to the trial of the Pyx.

The work in question was Newton's last, *The Chronology of Ancient Kingdoms Amended,* published posthumously in 1728. It had been written primarily as a defense of a short chronology of ancient history he had prepared for the Princess Caroline which had been printed in an unauthorized version in France in 1725. (Manuel 1963, ch. 1; More 1962, ch. 16; Pearson 1928). As part of his investigation of the dates of ancient events, Newton sought to show that earlier chronologists who had reckoned the average reign of ancient kings as being 35 to 40 years had been in error. By appealing to data from the more accurately recorded periods of history he meant to determine an estimate of the mean length of a reign, a quantity that could in turn be used to estimate lengths of eras when numbers of kings were more accurately recorded than were years. He wrote:

> For by the ordinary course of nature Kings Reign, one with another, about eighteen or twenty years a-piece: and if in some

> instances they Reign, one with another, five or six years longer, in others they Reign as much shorter: eighteen or twenty years is a medium. So the eighteen Kings of *Judah* who succeeded *Solomon*, Reigned 390 years, which is one with another 22 years a-piece ... (Newton 1728, p. 52).

Newton went on to list the data for eleven more periods of time; these data are presented in tabular form in Table 2. After presenting the data, he repeated his

2. Newton's Data on Length of Kings' Reigns

Kingdom	Number of kings	Years	Mean reign (Newton)	Mean reign
Judah	18	390	22	21.67
Israel	15	259	17¼	17.27
Babylon	18	209	11⅔	11.61
Persia	10	208	21	20.80
Syria	16	244	15¼	15.25
Egypt	11	277	25	25.18
Macedonia	8	138	17¼	17.25
England (1066–1714)	30	648	21½	21.60
France (first 24)	24	458	19	19.08
France (second 24)	24	451	18¾	18.79
France (last 15)	15	315	21	21.00
France (all)	63	1224	19½	19.43

Source: Adapted from Newton (1728, pp. 52–3).

contention that in "the course of nature," the reign of kings should be reckoned "at about eighteen or twenty years a-piece" (Newton 1728, p. 54). Where did Newton get his interval estimate of the mean? That he repeated "eighteen or twenty years" three times rather than even once quoting nineteen as a "medium" would indicate that he did indeed think of his estimate as an interval. Newton presented no calculations, but it is instructive to use the twelve average reigns of the third column of Table 2 to compute an overall mean plus or minus an estimated standard error of the mean, or $\bar{X} \pm s/\sqrt{12}$, in the manner of present day physicists. We find 19.10 ± 1.01. An alternative analysis would (a) note that the kings of France should not be included twice, eliminating the twelfth row of the table, (b) use the actual means of column 4 rather than Newton's approximations, (c) use a weighted (by number of kings) mean of the eleven remaining numbers in column 4, and the corresponding maximum likelihood estimate of the standard deviation of this weighted mean. This would give 19.03 ± 1.01. Now, the point is not that Newton did either of these calculations. We can be certain he did not, although he may have calculated the mean of the third column. The point is that the ad hoc or intuitively derived interval he did present was in rough accord with something reasonable, i.e., roughly a 65 percent confidence interval. Had Newton specified an interval on the same basis as that apparently used to determine the remedy in the trial of the Pyx, he would have made a very different statement. As we have seen, the limits set for the trial of the Pyx correspond to values such that a sizable proportion (about 95 percent) of single coins fall between them; these limits were then, in effect, applied to average weights of large numbers of coins. An analysis of the length of reigns of the kings of England, data easily available to Newton, suggests that an interval such as 19 ± 11 includes about 65 percent of king's reigns. Yet Newton, who knew he would be applying his statement to aggregates of reigns, based his analysis upon aggregates and gave an interval appropriate to aggregates. Newton must have realized that 19 ± 1 would only contain a small fraction of individual reigns, and it seems safe to say on the basis of this that he had at least an approximate intuitive understanding of the manner in which the variability of means decreased as the number of measurements averaged increased. The application of this understanding to the trial of the Pyx would have been well within Newton's capabilities.

Newton's term as master was not free of scandal. Following a 1710 trial of the Pyx, the charge was made that his gold coinage was less fine than the trial plate, but he survived with his reputation intact after successfully arguing that the 1707 trial plate being used was too fine a standard for fair comparison (Craig 1953, p. 216; 1946, p. 77; Newman 1975). Newton did become wealthy as master of the Mint. Prior to 1695, "his pecuniary circumstances are said to have been rather straightened," (De Morgan 1840, p. 200) yet when he died he left a princely estate of £32,000. However, Craig (1946, p. 124) feels that this can be accounted for by savings from salary and fees of about £1,500 per year and investment income; indeed, de Villamil (1931, pp. 19–29), in a partisan but apparently factual account of Newton's finances, provides a detailed scenario of how Newton, despite investment losses of £4,000 and considerable generosity to friends and relatives, could have amassed such a fortune. Thus while the inventor of the calculus and discoverer of the law of gravity may or may not have fathomed the intricacies of the significance test, there seem to be no grounds for believing that he took advantage of this knowledge for personal gain.

6. THE SITUATION IN THE UNITED STATES

Trials similar to that of the Pyx have long been carried out in the United States but with an interesting difference. The Constitution reserved the right to coin money to the federal government, and the U.S. Mint was established by congressional action in 1792. The Act of 1792 provided that of each "separate mass" of metal made into coin, no fewer than three coins be reserved for a yearly assay in the presence of the Chief Justice, the Secretaries of State and Treasury, and the Attorney General. The Act (Section 18) further specified that results must be within one part in 144 parts of the given standards, else the Mint officers "shall be deemed disqualified to hold their respective positions." (United States 1894, p. 6). Thus, at the outset, the American system was much like the British.

However, in 1837 another Act of Congress introduced a subtle change into this procedure. In Section 25 of the Act of January 18, 1837, a double system of tolerances was introduced. It was specified that while a deviation of $\frac{1}{4}$ grain would be permitted in a single gold coin, if a thousand eagles (=\$10 each) were weighed together, the allowed deviation would be only 2 dwt = 48 gr, or only 192 times the allowance for a single eagle! No reason for the change is given in the Act, but it had the effect of greatly diminishing the tolerance from the corresponding level under the British system if large numbers of coins were weighed together. The yearly trial became more exacting and the levels closer to those of a modern statistical test.

It seems plausible that this provision of the Act of 1837 was included as a response to an invitation dated December 2, 1833 by the Director of the Mint, Samuel Moore, in a report to Congress. Moore had suggested as desirable "An amendment of existing provisions in regard to the allowance for casual error in the weight and fineness of the coins issued from the mint." (United States 1833–34, p. 7). More specifically, he noted that technological improvements had led to increased precision in minting, and thus a more restricted tolerance would be feasible, particularly as in the past the annual assay had "... very rarely discovered a deviation ..., either in excess or defect, amounting to one-fourth part of the legal allowance ..." (United States 1833–34, p. 8). Moore did not spell out a schedule of tolerances but seems to have had a double system in mind when he wrote:

> That there should be an allowance in regard to weight in small quantities of coin taken alone, is indispensible, since perfect accuracy is not to be effected but by a degree of care and expense obviously disproportioned to any benefit which would result from the attainment of this extreme precision. The average, however, of large amounts should be conformable to the weight contemplated by law. (United States 1833–34, p. 8).

Over succeeding years the tolerances were often changed, although the double system was retained. The double eagle was introduced in 1849, and the Act specified that for this new coin a deviation of $\frac{1}{2}$ gr would be allowed for a single piece, 3 dwt = 72 gr for a thousand weighed together. It is in connection with this coin that we find the first intrusion of which I am aware, of the calculus of probabilities into the testing of coins.

In an 1871 paper read to the American Association for the Advancement of Science, J.E. Hilgard, then with the U.S. Coast and Geodetic Survey, noted that the law specified the U.S. Mint trial must be held once a year and that at that time only a single coin was reserved from each day's work. In an actual case, this left 328 double eagles to be tested, but the law only specified tolerances for one and 1,000 pieces. After arguing that the severity of the possible penalty made the application of the 1,000 piece tolerance unjust, Hilgard wrote:

> The law of probabilities, according to which accidental errors accumulate in the ratio of the square root of the number of cases involved, would give, upon the basis of one-half grain for a single piece, a tolerance of sixteen grains for 1,000 pieces. The statute allowing seventy-two grains, it is clear that the law of probabilities has not been held to apply; but an ampler margin has been allowed, which is doubtless proper and necessary, since that law disregards *constant* sources of error, which are likely to obtain in the operations here in question (Hilgard 1872).

He went on to suggest that a curve of the form $y = n^{\epsilon}$ be fitted to the statuatory tolerances, found them consistent with a relationship $y = n^{.72}$, and deduced that the appropriate tolerance for 328 double eagles was 32 gr. Estimates of the serial correlation of the weights of selected double eagles are not available, but it seems plausible that they would be positive and would approximately validate Hilgard's conclusions.

7. CONCLUSIONS

Clearly, routine sampling inspection for the control of quality has existed for considerably longer than is generally realized. Mathematical statistics appears to have played no role in the formulation of these early procedures, but early versions of modern concepts were present nonetheless. The trial of the Pyx even in the Middle Ages consisted of a sample being drawn, a null hypothesis (the standard) to be tested, a two-sided alternative, and a test statistic and a critical region (the total weight of the coins and the remedy). The problem even carried with itself a loss function which was easily interpretable in economic terms. One could argue that the problem was in reality a three-action problem, that the two sides of the alternative corresponded to two traits (greed and incompetence) for which different actions might be appropriate; but if this were so, the Crown valued the traits equally little.

From our present vantage point we can find much in the trial of the Pyx to criticize. The sampling scheme was sufficiently ill-defined that many sources of bias could have crept in. The linear dependence of remedy on sample size effectively reduced the probability of a type I error to zero and all but destroyed the power of the test to detect nearby alternatives, thus providing a steady and sure income to masters of the Mint at different times in different countries. But we must not be too harsh on the trial, despite these shortcomings. Insofar as public confidence in the coinage was a goal, a small type I error probability was a necessity, in order that the Mint not be frequently shaken by groundless scandal, and there is little doubt that even an easy test kept the Mint alert. Even though the Mint must have developed over the years some sense of the safety of its position in the trial of the Pyx, its officials (with the intriguingly possible exception of Newton) could not have understood the mathematics of this statistical problem any more than did the Crown.

With many ideas of mathematical statistics in the air after 1800, it is curious that changes in the relationship between remedy and sample size came so late. This may be evidence that the trial served its political role rather well and did not require close reexamination of its procedures. It is interesting that when a change did occur it came first in America, in the 1830's when mathematics in America was suffering from retarded development and mathematical statistics was nonexistent. Freedom from the bonds of centuries of tradition but not from the technological changes of the nineteenth century, apparently provided a climate sufficiently congenial to procedural innovation to compensate for whatever lag may have existed in mathematical research.

Another point bears mention in connection with the trials of the Pyx: the conspicuous absence of a routine examination of the variability of the weights of the freshly minted coins. To a certain degree the existing test provided a check on variability because the critical values were not dependent on the sample variability, and a large increase in variance would have increased the chance that the null hypothesis and the master would be rejected. The fact that a check on variability was not introduced into these early tests is evidence that the economic effects of increased variance (e.g., the "come again guinea") were not so great as the effects of changes in the average and that the very idea of measuring dispersion was not yet a well-formed concept. The only means of checking variability which seems to have been considered was the application of the remedy to single coins. For example, the royal warrant of 1663 provided for a limitation on the variation of individual coins, but there is no evidence that this was incorporated into the trial of the Pyx. Such a procedure would have been extremely cumbersome and was apparently not used routinely in trials before 1870. As late as 1848, the Deputy Comptroller of the Mint testified that

> Provided the manufacture of the money is conducted in good faith by the Moneyers, ... I am of the opinion that as great a degree of *uniformity of weight upon the piece* is to be obtained by the remedy being allowed upon the pound weight as upon the piece (his emphasis) (Great Britain 1848, p. 166).

But if some aspects of statistical tests were slow in evolving, the appearance of a test did occur at an early date. The longevity of the trial of the Pyx testifies to its perceived importance.

The trial of the Pyx, in Great Britain, and the Annual Assay Commission, in the United States, have continued on an annual basis to the present day. The procedures followed at these modern trials differ but little from those of past centuries. Modern scales and assay techniques are used, but in the main the same tests are performed. In the United States, the test of weight is now performed separately on individual coins selected at random from previously selected mint bags. This change improves over early trials in that the use of inappropriately large tolerance levels for large batches of coins is avoided, and it helps guarantee a greater uniformity in the coinage than did the early procedures. However, it provides low power and little protection against slight changes in average weight, unless such a change is accompanied by an increase in variance. No doubt the great decrease in the use of precious metals in coins and the increase in the use of automatic coin machines have contributed to the present emphasis upon uniformity in individual pieces over accuracy in aggregate weight.

In the present British trial of the Pyx, gold coins are weighed in aggregate and a sample weighed individually, while other standard coins are only weighed in bulk. However, there has been one major innovation in these trials: The "average diameter" of the cupronickel coins is now measured and compared with a permitted variation from a standard. Neither the barons of the twelfth century nor the goldsmiths of the thirteenth foresaw the demands of twentieth-century telephones and vending machines!

[*Received January 1977. Revised February 1977.*]

REFERENCES

The Rev. Rogers Ruding's three volume treatise (Ruding 1840) is the definitive work on the coinage of Great Britain before the nineteenth century. Sir John Craig's book (Craig 1953) gives an excellent, modern, scholarly treatment of the London Mint, and is likely to be more accessible than Ruding.

Craig, John (1953), *The Mint*, Cambridge: Cambridge University Press.
——— (1946), *Newton at the Mint*, Cambridge: Cambridge University Press.
De Morgan, Augustus (1840), "Newton, Isaac," in *The Penny Cyclopedia*, Vol. 16, London: Charles Knight, 197–203.
de Villamil, R. (1931), *Newton: The Man*, London: Knox.
Forbes, J.S., and Dalladay, D.B. (1958–9), "Metallic Impurities in the Silver Coinage Trial Plates (1279 to 1900)," *Journal of the Institute of Metals*, 87, 55–8.
Great Britain, House of Commons (1837), "Report from the Select Committee of the House of Commons on the Royal Mint," in Vol. 16 of *House Documents for 1837*.
——— (1848), "Report of the Commissioners Appointed to Inquire into the Constitution, Management, and Expense of the Royal Mint," in Vol. 28 of *House Documents for 1849*.
——— (1866), "Copy of a Report to the Comptroller-General of the Exchequer upon the Trial of the Pyx, and Coinage Subjected to this Public Trial, Dated the 10th Day of February 1866, by Mr. H.W. Chisholm, Chief of the Exchequer," in Vol. 40 of *House Documents for 1866*.
Hilgard, Julius E. (1872), "An Application of an Exponential Function," *Proceedings of the American Association for the Advancement of Science*, 20th Meeting, August 1871, 61–3.
Manuel, Frank E. (1963), *Isaac Newton: Historian*, Cambridge, Mass.: Harvard University Press.
Newman, E.G.V. (1975), "The Gold Metallurgy of Isaac Newton," *Gold Bulletin*, 8, No. 3, 90–5.
Newton, Isaac (1728), *The Chronology of Ancient Kingdoms Amended*, London: Tonson, Osborn, and Longman.
Oresme, Nicholas (1956), *The De Moneta of Nicholas Oresme and English Mint Documents*, trans. by Charles Johnson, London: Thomas Nelson.
Pearson, Karl (1928), "Biometry and Chronology," with Appendix by C.F. Trustam, *Biometrika*, 20A, 241–62, 424.
R. (1815), "Trial of the Pix," *The Gentleman's Magazine and Historical Chronicle*, 85 (8 of New Series), Part 2, 207–8.
Ruding, Rogers (1840), *Annals of the Coinage of Great Britain and its Dependencies*, 3rd ed. London: John Hearne (especially pp. 69–77 of Vol. I).
Sheynin, Oscar B. (1971), "Newton and the Classical Theory of Probability," *Archive for History of Exact Sciences*, 7, 217–43.
United States, Laws, Statutes (1894), *Coinage Laws of the United States, 1792 to 1894*, 4th ed.—Revised and Corrected to August 1, 1894, Washington, D.C.; U.S. Government Printing Office.
United States, Public Documents (1833–34), "Report of the Director of the Mint," Document No. 8 in *Executive Documents*, Vol. I, 23rd Congress, 1st Session, 1833–34, U.S. Serials Vol. 254.
Unsigned (1892), "Trial of the Pyx," *Chambers's Journal*, 69, 454–5.
Westfall, Richard S. (1973), "Newton and the Fudge Factor," *Science*, 179, 751–8.