PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Quality Detective: A Case Study [and Discussion]

S. Bisgaard, L. P. Fatti, David Cox and J. Bibby

Phil. Trans. R. Soc. Lond. A 1989 **327**, 499-511
doi: 10.1098/rsta.1989.0006

Phil. Trans. R. Soc. Lond. A **327**, 499–511 (1989)
Printed in Great Britain

The quality detective: a case study

BY S. BISGAARD

Center for Quality and Productivity Improvement and Department of Industrial Engineering, University of Wisconsin–Madison, 610 *Walnut Street, Madison, Wisconsin* 53705, *U.S.A.*

The idea of modern quality improvement is to improve processes and remove causes for problems rather than passively inspect. People engaged in quality improvement projects work as detectives. Scientific method plays a major role, and statistical methods and design of experiments are the tools for fact finding. Data and graphs are sparks that light up the imagination of the investigator. The ideas coming from the data are catalysts that bring out the best of engineering.

An industrial case study is presented that illustrates the quality detective idea. I show how engineering and statistics played dual roles in an iterative learning process. Design of experiments, statistical methods of analysis and engineering were used to remove causes for excessive variation. Practical constraints, time pressure, psychology and relations with management and other team members had an important impact on the experimental strategy. This example illustrates the tremendous economic gains industry can make by using designed experiments.

1. INTRODUCTION

The driving force behind the eighteenth-century Industrial Revolution was scientific method. For example, the members of the Lunar Society and, in particular, Josiah Wedgwood learned by careful experimentation how to produce high-quality products at low cost. The same ideas that set the Industrial Revolution in motion are behind the modern quality revolution. By application of scientific methods similar in spirit to those used, for example, by Charles Darwin, modern industrialists can improve their products, processes and services for the benefit of mankind. However, they are in a better position than their eighteenth-century predecessors because scientific methods themselves have been improved. In particular, statistical methods of experimental design and analysis as developed by R. A. Fisher in the 1930s constitute a major improvement. But the potential of modern methods of statistics in the industrial context is only beginning to be realized. Part of the reason for this delay is perhaps the traditional teaching of statistical methods in engineering and business schools.

A single frame of a film leaves no impression of motion. Similarly, most examples in statistics textbooks are snapshots that do not show the dynamics of investigations and the role statistics play. Too much emphasis is placed on formal procedures of hypothesis testing and probability models, and too little on the role of statistics in the context of scientific investigations and the interplay between deduction and induction in an iterative learning process (see Box 1976). Thus potential users, that is, engineers and business students, often leave universities without a proper understanding of the power of statistics and are likely in their later careers in industry to regard statistics as useless.

Real-life problem solving with statistical methods is a dynamic process that is often like detective work. Successful statistical practitioners involved in quality improvement projects in industry use data the same way Sherlock Holmes solves a murder mystery, by using fingerprints on the murder weapon and footprints on the front lawn. 'Quality detectives' follow hunches and get inspired by evidence collected in conjunction with their subject-matter expertise. Data inspire them to follow new leads in pursuit of never-ending quality and productivity improvement. It is with a sequential use of experimental design and analysis that the real power of statistics applied to industrial problems is realized. For a refreshing discussion of scientific method relevant to this paper, see Beveridge (1951).

In this paper a case study is presented which shows how a quality problem was solved through a dynamic problem-solving process using statistics and experimental design as well as engineering. The emphasis is on the problem-solving process and non-statistical issues such as teamwork, relations with management, strategy, time pressure and psychology.

2. Background for the experiments

A few years ago a company which manufactures consumer goods was confronted with unexpected difficulties in getting a prototype machine to produce a saleable product because of excess variability in a cutting process. The machine was designed for a new product line, and the company was planning to build seven additional machines based on the prototype as demand for the product increased. Trial runs preceding full-scale manufacturing were in progress when the quality problem was discovered and a team of two engineers, three technicians and two operators was assigned to solve it. However, after eight months of hard work, and a budget over-run of several hundred thousand dollars, the cause of the problem was still a mystery.

3. Technical description of the problem

A major component of the machine was a conveyor belt, about 40 ft (*ca.* 12 m) long, 4 in (*ca.* 10 cm) wide and $\frac{1}{2}$ in (*ca.* 1 cm) thick. It transported, at high speed, a very light product that was about 6 inches long. The conveyor belt was composed of a set of interlocking hard plastic links much like a chain. A vacuum created by a box mounted under the perforated conveyor belt held the products in place on the belt. See figure 1 for a simplified layout of the machine.

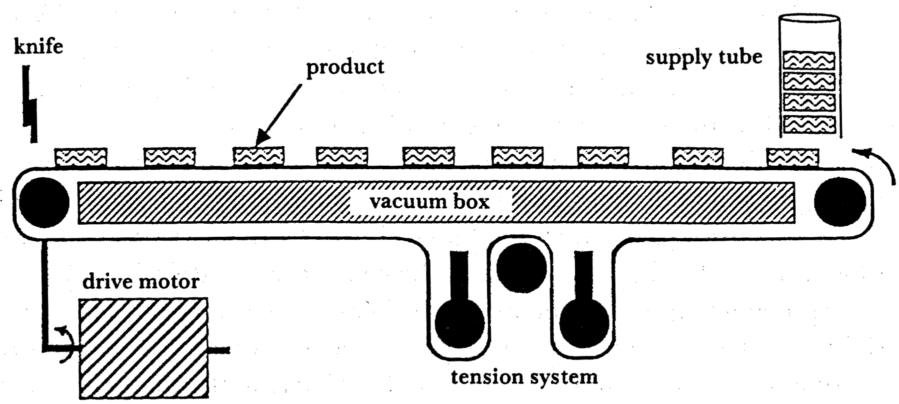

Figure 1. Simplified layout of the machine showing conveyor belt, products, original tension system, knife, vacuum box, supply tube and drive motor.

The cause of the quality problem was, as mentioned above, excess variability in the final cutting process at the end of the conveyor belt. If the products did not arrive for cutting at precisely the right moment, and in the right position, the product would be cut off outside specifications. It was easy to adjust a *constant* difference in the timing of the knife relative to the speed of the conveyor belt. Excessive *random* variation, however, in either the positioning of the individual products on the conveyor belt or in the speed of the conveyor belt, could not be compensated for.

In their efforts to solve the problem the engineers had rigged up a process computer connected to photo-eye and magnetic proximity switch sensors. The photo-eyes were installed along the conveyor belt and monitored the distance between the products. The magnetic proximity switches, which were activated by the passage of steel screws put into the side of the conveyor belt at equal intervals, were used to monitor the speed of the conveyor belt.

At times the engineers were reading off numbers from as many as 25–30 monitors, thus creating very large quantities of data. As the frustration about the problem was spreading, they seemed to think that clues would emerge if they could just read off more data from more monitoring points. As a result they flooded themselves with data.

As the inundation increased one of the engineers who had previously taken an introductory statistics course suggested that they should seek help from a statistician, and I was contacted. From our initial telephone conversation I got the impression that the problem was quite complicated and that my task would be to make sense out of the large sets of time-series data already obtained. I found it difficult to get a clear idea of what the problem really was and, with some reluctance, I agreed to come and visit the company.

4. The first meeting

During the first part of my one-day visit I listened carefully to the engineers' explanation of the problem. They showed me many time-series plots and I visited the shop, saw the machine in operation, and talked to the operator. By lunchtime my understanding of the problem was still superficial but it seemed clear to me that analysing their old time-series data would be a waste of time. It was not clear under which circumstances their data were collected, and it was easy and not too expensive to run new trials.

Later that day I explained that an experimental approach, whereby the settings of the machine variables were systematically changed, would be more useful. Neither the managers nor the engineers were familiar with experimental design, so I used the late Professor William G. Hunter's explanation that looking at old data is like passively listening to a process. By contrast, experimentation is like actively engaging in a conversation with a process. By asking directed questions one is more likely to get answers to what one wants to know. The managers and engineers seemed to like this explanation, so I proceeded to explain the basics of factorial design and response surface methods. Nevertheless, the engineers had worked hard and gone through much trouble to get to where they were, so when they realized that I was serious about turning the knobs on their machine, the excitement cooled off and the prevailing attitude was 'Don't touch anything!'.

When I left the company late that day I did not know whether they wanted to follow my advice and run experiments. However, about a month later I was invited back for two days and the management agreed to let us experiment. In the following sections I shall recapitulate what happened. Before going on it is necessary to explain briefly the personalities involved and the psychological environment, because these factors influence experimental strategy. I shall provide a short description of the three team members who conducted the experiments, and the manager.

5. The key players

The engineer, who was also the person who had taken the initiative to contact me, was very practical, willing to take risks, try new things, and even break company rules that prevented us from proceeding swiftly. He was a person of action and good with computers. The operator came to play a crucial role. He had been with the company for 15 years and his extensive experience was invaluable to us. He was very observant, mechanically inclined, and quick to learn. The manager of engineering projects mostly played an indirect, yet decisive role. His belief (in my opinion) in 'management by fear' had a profound influence on the way we conducted the experiments. I benefited from my own background as toolmaker and manufacturing engineer before I got my education in statistics.

6. Initial series of experiments

I will now give a chronological description of how we solved the problem of excessive variation in the production process. I returned to the company on Tuesday 19 November 1985. From the outset I was aware that I did not completely understand all the details of the problem. Furthermore, I knew that the engineer and the operator did not know about factorial experiments. Because I had to rely on them as team players it was important that they were carefully instructed. Therefore I spent considerable time explaining my ideas to them. I wanted them, based on their past experiences, to identify the most important factors for reducing variance and then run a simple two-level factorial experiment with these. Through a sequence of experiments I hoped that we would be able to identify better ways of running the machine.

At the initial stage an overriding consideration was simplicity. In the past the engineers had been reading off measurements from about 20 to 30 sensors. I reduced this number to 2. The most important measurements had to stem from the observation of the distance between products at the end of the conveyor belt right next to the knife. As a second response we used the speed of the conveyor belt next to the knife because that was not necessarily the same as the speed of the products. It was measured by a proximity switch at the end of the conveyor belt. Other measurements had to be subordinate to these and could always be considered later if necessary.

We decided that three minutes of production would be adequate for each individual test. The operator told us that in his experience the most important factors for reducing variation were the tension of the conveyor belt and the amount of vacuum. Next the operator told us what would be reasonably high and low levels of these two factors.

We were now ready to run the first two-factor, four-run experiment which we completed in an hour. The results showing the distance between the individual products, as measured by the photo-eye for each of the four runs, are displayed in figure 2. From inspection of these plots we learned that increasing either or both vacuum and tension increased the variation. The best combination of the two variables was low vacuum and low tension. But there also seemed to

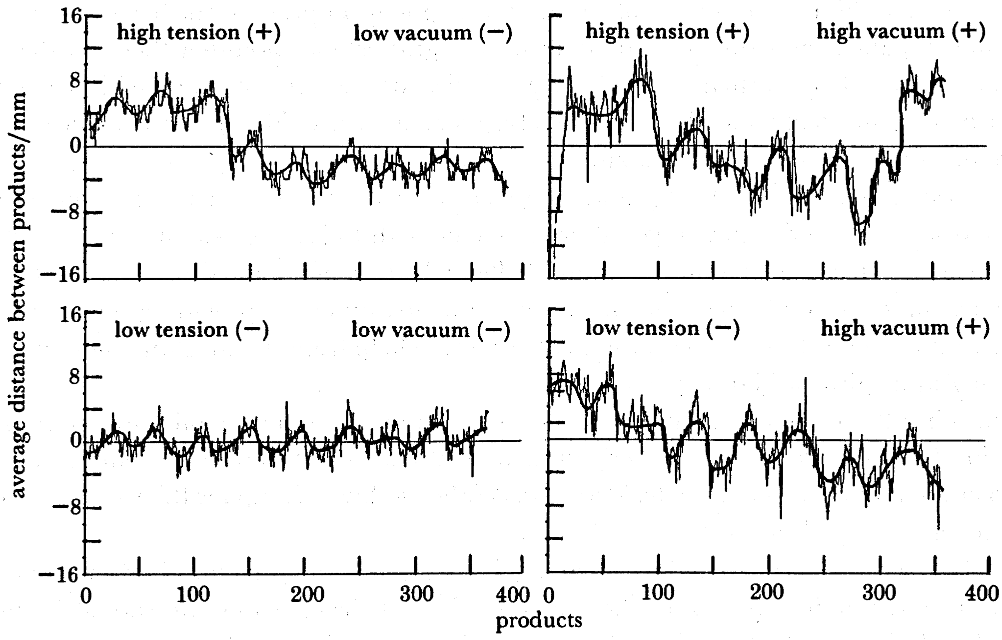

FIGURE 2. Results of the first experiment, showing the four authentic time-series plots for the four combinations of low and high vacuum and tension from a two-level factorial experiment in two factors. The smooth curves indicating a systematic cycle were drawn by hand by the experimenters at the time of the experiment.

be a systematic cycle in the data. The cycle, traced by hand by me at the time of the experiment, is shown in figure 2. It was no surprise to the engineer or the operator, but they had no idea what caused it. There was no obvious reason for this except, perhaps, that it was caused by the pneumatic tension system or friction.

The purpose of the tension system was to take up slack in the conveyor belt. It consisted of two pneumatic pistons exerting pressure on two rollers over which the belt passed. By increasing the air pressure the two rollers would be pushed outward and take up slack. It seemed plausible that the air in the pistons' cylinders could be compressed and cause the rollers to bounce back and forth. This could, in turn, cause the cycle in the data.

Another possible cause for the cycle was friction. There seemed to be an excessive amount of it in the 40 ft track in which the conveyor belt ran, making the movement of the belt erratic. If friction was the problem, it would be aggravated by increasing tension on the conveyor belt or by using more vacuum. It seemed to me that we could remove some of the conveyor belt edge guides and this would reduce friction. We therefore removed as many as we thought we could and looked for other ways to smooth the path for the conveyor belt.

From the first experiment it was clear that we had made the differences in the factor levels too large. Thus we decided to run a second experiment with a smaller difference in both vacuum and tension levels. For reference we kept the old low levels the same, so that we could compare and see the impact of removing friction.

We were ready to start experiment number two on the second morning, Wednesday 20 November, and finished it in a few hours. Simple inspection of the plots showed no major difference between the results of the first and the second experiment, and the systematic cycle was still there. I am not embarrassed to admit that our 'statistical analysis' up to this point only amounted to simple inspection of the plots. The results were quite obvious and we were under great time pressure. Moreover, it was only later that we got appropriate statistical software. For this and all subsequent experiments I plotted out the graphs and mounted them on the wall for everyone to see. That turned out to be a good idea because it not only facilitated much discussion among the team members, but it also seemed to provide visible evidence which assured management that we were making progress.

After the second experiment we became more suspicious of the pneumatic tension system. After some discussion we decided it would be simple enough to jury-rig an alternative system. We drafted a design for a pivoting arm and roller which could push down on the belt to increase tension. This simple mechanical arm had the advantage of providing an even load, and any vibration could be observed directly. With some help from the machine shop we built it in one day in a spirit of high expectation.

After finishing the tension system we looked over the machine once again and eliminated more friction from the track. The last experiments had led us to believe we should lower the levels of vacuum and tension. So we decided to make the old low settings for the factors vacuum and tension the new high levels.

By now the operator and the engineer had a good grasp of what a factorial experiment was and I had a better understanding of the machine. Thus I decided that we should be more daring and include more factors. One obvious candidate was the speed of the conveyor belt. Perhaps the cycle we had observed was a resonance effect. Speed would then be important. A fourth factor, a pressure belt, was also included, but later turned out to be of no importance.

We had spent all of Thursday 21 November rebuilding the machine, but on Friday morning, 22 November, we were ready to run one more experiment. I suggested that we conduct a four-factor, two-level fractional factorial experiment in eight runs, and I wrote down the design. The main idea behind this experiment, as indicated above, was to see if the new tension system was better. But we also wanted to explore new regions of the design space. A third consideration, in case the cycle was not due to the tension system, was to see if we could minimize the amplitude of the apparent cycle in the data.

We quickly conducted the eight-run experiment and plotted the data. To our disappointment the cycle was still there, so we eliminated the hypothesis that the cycle was due to the tension system. However, we had clearly made some improvements: we found that an increase of speed and vacuum resulted in a decrease of the variance, and that tension and vacuum seemed to interact.

It was already late Friday afternoon when we completed the analysis. We had no more time to run additional experiments before the weekend. I felt we had come a long way already. Our team worked well together and I was confident that we could reach a solution following a response surface strategy. However, management did not indicate whether they wanted me to continue, so I left the company without knowing if I would return.

About a week later they asked me to conduct a new series of experiments. When I returned we decided to run a full two-level factorial experiment in four factors. We were suspicious that some of the variation was due to products being put on the conveyor belt in an irregular fashion. Thus in addition to the factors vacuum, speed, and tension we included a vacuum cleaner rigged up in reverse to blow air into an overhead supply tube. This would force the products to drop more regularly from the tube on to the conveyor belt. The two levels of the factor – the vacuum cleaner – were 'on' and 'off'. As a last preparation we further reduced the friction of the track.

Based on the experience from the third experiment we decided to make the new low level of speed the same as the old high speed, and the new high level somewhat higher. Similarly, the old high vacuum was used as the new low level, and the new high vacuum level was somewhat higher. Tension was unaltered from the previous experiment.

We were ready to run our next experiment at 10 a.m. on Thursday 5 December. The experiment lasted a few hours but the results of this experiment were disappointing. All the trials looked almost alike and none of them showed a real reduction in the variance. But the systematic cycle was now even more clear. It was as if our effort to decrease friction had removed the most irregular components of the variation and made the cycle more apparent. The question which increasingly bothered me was 'Where did the cycle come from?' By now we had eliminated the obvious possibilities. Disheartened and wanting to do something, without knowing exactly what, I suggested we run one more factorial experiment exploring the effect of vacuum. But this yielded no useful clue.

7. Despair and frustration

The week had come to an end. The operator and the engineer attempted one more experiment over the weekend. But they encountered numerous problems which led to the experiment's early termination. When I returned on Monday morning we ran still another experiment which also was ended because of technical trouble.

An increasing feeling of despair was emerging. I felt we were stuck. I was running out of ideas and didn't know how to proceed. Time pressure was increasing. I felt that management was losing patience and I, myself, could allocate no more than the rest of the week to this project.

That evening I stayed up very late in my hotel room, organized my notes, and studied our experimental diary in hope of finding a clue. We had run so many experiments, and at such a fast pace, that I felt a need for reviewing what we had learned so far. I could hardly remember the logical steps any more. Hence I thought it would be useful to write a summary. Unfortunately this exercise did not provide the miracle for which I was beginning to hope so desperately.

I began to fear the whole project would fail. I envisioned that management would blame me for experimenting rather than analysing the large piles of old data as they had initially asked me to do. I was particularly upset because I had been so sure of myself at the beginning of the project and had claimed I could solve this problem with statistical methods. Now it looked as if I would have to give up. But I was also getting stubborn. I wanted badly to solve this problem. My professional pride was at stake.

As I was going over my notes I looked at some plots I had made of the autocorrelation for some of the earlier experiments. Figure 3*a* shows an example of the autocorrelation for the photo-eye data and figure 3*b* shows the autocorrelation plot for the corresponding proximity switch data. There was a clear peak at 80 in the photo-eye data and a very distinct peak at 40 for the proximity switch data. But what caused this cycle as indicated by the peak? I had no clue. That evening it was clear to me that we could make no progress unless we tracked down the source of this systematic cycle.

However, I had to think about strategy, too. It was safer for me to write a report justifying every step we had taken and state that the problem was unsolvable. At least with a report we had something to show and could cover our backsides. I was inclined to gamble and run more

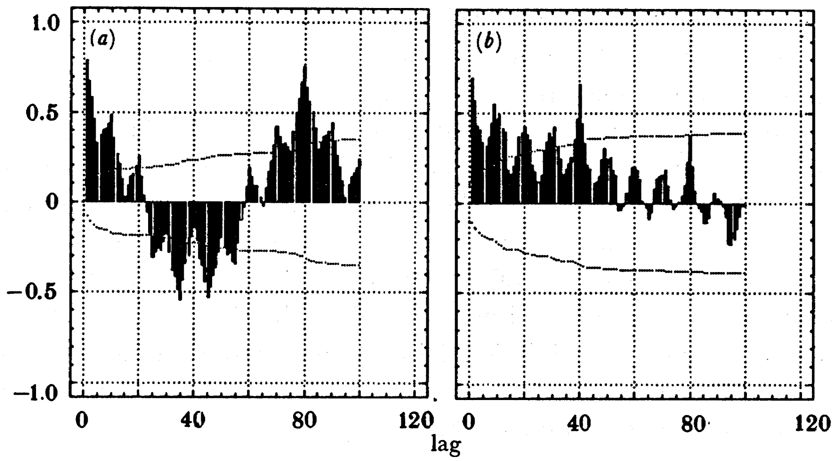

FIGURE 3. Plots of the estimated autocorrelation function for (*a*) the photo-eye data and (*b*) the proximity switch data. The autocorrelation function for the photo-eye data shows a peak at lag 80. The autocorrelation function for the proximity switch data shows a peak at lag 40.

experiments, but the risk was that we could end up empty-handed without a solution and nothing to show. Given the 'management by fear' environment at the company it seemed most sensible that evening to play it safe and write a report. However, as we will see later, despite all common sense we ended up running more experiments.

8. Breakthrough

When I went into the shop early the next morning I was determined to follow the safe strategy. But first I wanted to spend a few moments talking to the operator and ask him one more time if he had any idea why the cycle's length corresponded to 80 data points. After some discussion he suddenly said 'Maybe the number of products on one loop of the conveyor belt is 80.' We immediately went over to the machine and turned it over by hand so we could count the number of products on one loop. He was right. There were exactly 80 products on one loop. Next we counted the number of screws and found exactly 40. We were now very excited because the number 40 corresponded to the well-defined peak on the autocorrelation plot of the proximity switch data. This had to be pursued. I forgot everything about my 'play-it-safe' strategy. But how should we pursue this lead? I had no clear idea. I recalled that Box (1966) had written 'To find out what happens to a system when you interfere with it, you have to interfere with it (not just passively observe it).'

While considering this, it suddenly occurred to me that perhaps we could add one more link to the conveyor belt's length. This would increase the total number of products on one loop from 80 to 82 and the total number of screws to 41. If the cycle was length-dependent this would be the kind of interference we should try. It turned out it would be possible to take up the additional slack in the conveyor belt caused by adding a link. Within minutes we added the link and were ready for a trial run. The excitement was high. We quickly ran approximately three minutes of production and got the data loaded into the computer and calculated the autocorrelation for the photo-eye data. Although one really had to know what to look for, there definitely was a shift in the peak. Next we looked at the autocorrelation for the proximity switch data and here it was even clearer. We had changed the cycle and thereby established that it was somehow associated with the length of the conveyor belt.

One should not be carried away by excitement and jump to conclusions. Thus it occurred to me that if we looked at some of the old data produced before changing the tension system we should be able to see a cycle longer than 82. The old tension system had the conveyor belt looped over several rollers and was 84 products long. We found one of the old data-sets from before the change and computed the autocorrelation. As I had conjectured, the peak was at 84. I was now convinced that the cycle was due to the conveyor belt and associated with its length. However, to be absolutely certain, I proposed we reduce the length back to 80 products. Once again the cycle moved.

We had now established the source of the cycle. But we still did not understand the cause. Was there some structural pattern in the belt or was it because of wear? We took the belt off the machine and found a place where we could stretch it out in its full length to study it more carefully and measure it. From a visible inspection we could see no cause for a repetitive pattern.

What next? Suddenly I got the idea that we might learn something if we broke down the belt into smaller pieces and reassembled them in a random order. It was a long shot. But why not try it? Perhaps this randomization would break the cyclic pattern. We took the belt off again, broke it into seven pieces, and randomized them before putting the belt back together. Once again the excitement was high. Perhaps this experiment would provide a clue. A plot of the autocorrelation showed we had broken the pattern somewhat. That provided additional evidence that the cycle was a property of the conveyor belt. Based on this result I suggested the cycle was caused by a wear pattern in the belt and that by randomization we had broken some of this pattern. If that was the case, then the solution would be to get a new belt. I mentioned this and the engineer told me they did, in fact, have new belts on hand.

We all agreed it would be interesting to try one. We could hardly wait to get the new belt out of storage and on to the machine. We were now convinced we had found the solution to the problem and would soon be big heroes. However, instead of having a smaller amplitude the new belt made it much worse. The disappointment was immense, but we quickly recovered. This new fact was added evidence that the problem was a physical property of the conveyor belt. Contrary to what we thought at first, wear and use of the belt reduced this problem. Thinking about this, it seemed plausible that the cycle was caused by the elastic properties of the conveyor belt. The conveyor belt was made out of some kind of stiff plastic which could flex substantially over its entire length. Perhaps a new belt flexed more than an old and stretched-out one. If that was so it would be important to reduce friction and smooth any rough spots in the track for the conveyor belt because the surface in the track for the belt was still quite poor.

9. Final experiments and report

Time was running out for us, but we were anxious to solve this problem. It was now Wednesday afternoon, 11 December. There was little time for doing much more and we could hardly manage to write a report before Friday. Desperately we ran a few more experiments including a total rebuilding of the track for the conveyor belt, a task which lasted the whole night till early Thursday morning. But we achieved only a marginal improvement. As a substitute for writing a report I decided to invite management down to the plant for a presentation. For the presentation we put additional charts on the wall to document the last few days' hectic activities.

When management and the engineering staff arrived on Friday afternoon I explained what we had done. I showed them we had obtained convincing evidence that the cycle which caused the excessive variation was associated with the conveyor belt. The most likely explanation was that the cycle was caused by the conveyor belt flexing much like a closed-loop spring. I showed how we had redesigned the track to reduce friction, but that this reduction was insufficient to achieve the necessary low level of variation. Furthermore, we had demonstrated that it was unlikely that any adjustment of other process parameters, such as tension, vacuum and speed, could achieve the desired reduction in variance. Against this background I recommended that management find a conveyor belt of a less flexible material.

Management seemed to like the work we had done and were, in particular, impressed by our use of experimental design and statistics. But they said our proposed redesign would be costly and they wanted to be absolutely sure our recommendations would help. After some discussion and one more experiment which convincingly defeated the engineering manager's hypothesis that the cycle was caused by the gear-boxes, they decided to rebuild the machine as we proposed.

Of course I could not be absolutely certain my recommendation would help. So when the engineer contacted me in May 1986, I was relieved to hear that the redesign had been a great success. They had purchased a much less flexible conveyor belt and redesigned the track, resulting in a dramatic reduction in product-length variation. Now not only could they produce a saleable product but they were able to remain well within specifications. The cycle was gone and, as an unexpected benefit, they could now run the machine 50% faster because of the smaller variation. The economic ramification of this was enormous. The company, by improving quality, improved productivity and could meet its production demands with fewer machines than expected. This meant a substantial reduction in capital expenditure.

10. Conclusions

This case study illustrates a number of important points. First of all, modern quality improvement is not based on inspection but is much more like detective work. A scientific attitude whereby workers and engineers, like detectives, discover and remove causes for quality problems using statistical methods of experimental design and analysis can bring about quality and productivity improvement as well as lower costs. The use of statistics only as a means for inspection is expensive and will not lead to improved quality. It is important that we provide present and future genertions of engineers and business students with an impression of this investigative role for statistics.

The case study also illustrates the iterative nature of experimentation. The experimenters started from a limited knowledge about the problem's causes and proceeded through several steps in an iterative, adaptive, learning process to a solution. Each step in this learning process was guided by an idea or hypothesis which led to experments which, in turn, resulted in additional data. When the hypothesis was tested with the new data, it was either updated or rejected and replaced by a new one. This iterative adaptive model for problem solving is discussed further by Box *et al.* (1978).

The investigators' approach was clearly personal. Other investigators would probably have proceeded on a different path or developed a different solution. We saw how the investigators pursued an unproductive path or made mistakes. It is not usual to report the mistakes investigators invariably make and, of course, that is legitimate. However, one purpose of this paper is to discuss the process of experimentation rather than the specific results, and in that context it is important to be aware mistakes are made. For, as Box *et al.* (1978) write, 'the best time to design an experiment is after it is finished'.

In hindsight everything could have been done more efficiently. For example, why did we not pursue the cyclic pattern in the data right away? But again, as Box *et al.* (1978) remind us, 'At the end of an investigation, on looking back at the first experiments, one is frequently impressed (and even embarrassed about) how odd and pathetic they appear.' This remark applies to the young investigator who has just learned about experimental design as well as to any seasoned investigator beginning a new project.

A key to our success was creating an environment where each team member felt himself an equal partner. Contributing factors to a cooperative spirit were the use of graphics, careful explanations and the use of simple methods.

Engineering, product development and manufacturing have always involved a lot of experimentation, as any study of the history of technology will show. Only recently has industry recognized the importance of modern statistical methods of experimental design and analysis as tools for making the experimental effort more systematic and efficient. But experimentation does require management's support. Any experimental effort involves risks, which few are willing to take unless they have proper management support. In our example we nearly stopped short of a solution because we perceived it as safer to write a report than to run a few more experiments. If management wants engineers to conduct experiments they must nurture such efforts and help create an environment in which experiments can take place.

A valid criticism of my effort is that it was only firefighting. To prevent this quality problem from arising the company should have been using statistical and experimental methods from the very beginning and in all stages of design and manufacturing. In hindsight, if the management had appreciated the importance of designed experiments, the company might never have encountered this problem or, at least, they would not have been delayed as much as they were.

Finally, this case study touches on an important issue concerning the current debate on competitiveness of Western nations. Statistics properly used can bring about dramatic quality improvement of product designs and manufacturing processes, and reduce costs. Unfortunately, public debate often becomes an 'either/or' discussion. Should we focus on robots, computers and sophisticated manufacturing equipment or should the focus be on improving quality using statistical methods? In my view we should do both. It is wise to take advantage of new technology, but robots and computers can also lead to faster production of more scrap if the processes are not adequately understood. Statistical methods of design and analysis, as exemplified in this case study, can be powerful tools for improving the performance of high-technology manufacturing equipment.

I thank Professor George Box for encouraging me to write this paper and for many helpful discussions. Rocky Van Asten and Mark Jung deserve many thanks for their efforts in conducting the experiments. I also thank Dr Charles Schoengrund for discussion of the final manuscript. The work on this manuscript was sponsored by the U.S. National Science Foundation Grant No. DMS-8420968 and the Vilas Trust of the University of Wisconsin.

References

Beveridge, W. I. B. 1951 *The art of scientific investigation*. New York: Norton.
Box, G. E. P. 1966 Use and abuse of regression. *Technometrics* **8**, 625–629.
Box, G. E. P. 1976 Science and statistics. *J. Am. statis. Ass.* **71**, 791–799.
Box, G. E. P., Hunter, W. G. & Hunter, J. S. 1978 *Statistics for experimenters*. New York: Wiley.

Discussion

L. P. Fatti (*Department of Statistics, University of Witwatersrand, South Africa*). In this case study, as often happens when statisticians act as consultants in industry, it turned out that data which were painstakingly collected beforehand were not very informative or useful for solving the problem. It is a difficult task for the consultant to discard these data and then request that more data be collected according to his experimental design, and still retain a good relationship with his client. This seems to have been achieved in this case.

S. Bisgaard. I think the answer ought to be that one should never disregard data that have already been collected. Often, however, as in my case study, it becomes clear that the data do not supply the information needed to solve the problem. Perhaps I did not make it clear in my presentation that I did in fact look at the data, but not in great detail and it was my judgement, in the heat of the situation, that we would only be able to solve the problem by running new experiments. A factor that put me in a rather strong bargaining situation with my client was that I do not normally consult for a living, and I was at the time quite busy with other things. I was therefore not to eager to become involved in this project.

I agree that the question is a difficult one and that it may be hard to retain a good relationship with the client. Most good consulting in my opinion, however, comes from long-term relations with the client. A long-term relationship will to some degree eliminate the problem of poor experimental designs because the client will learn the benefits of working closely with a statistician at the design stage. May I add that I have always been a little embarrassed about this case study because of its 'Rambo style' which, although successful in this instance, should not be taken as the norm for statistical consulting. However, owing to our success the company now is putting much greater emphasis on the use of statistics in a more orderly fashion.

Sir David Cox, F.R.S. (*Imperial College, London, U.K.*). One general point which is implicit in the paper is the undesirability of drawing strong methodological distinctions between science and technology. In both, 'understanding' is a key.

S. Bisgaard. Sir David raises an important issue. Lately a zealous group of Taguchi disciples have been claiming that 'engineers don't want to know why' and because of this they are different from scientists and should therefore be served by different statistical procedures. I often wonder who has nominated these people to be spokespersons for such a large and diverse group of professionals. Incidentally, I also consider myself an engineer and I wholeheartedly agree with Sir David that 'understanding', or what George Box has called 'scientific feedback', is a key to progress in technology as well as science. If anyone would like 'data' that support this statement I recommend reading the biography *Edison's electric light* by Friedel *et al.* (1987).

J. Bibby (Edinburgh, U.K.). Sir David Cox notes the analogy between the iterative approach to quality improvement and the development of science as a whole. The parallel to which he refers is presumably Sir Karl Popper's image of scientific knowledge growing piecemeal by means of 'conjectures and refutations'. However, Popper's model is today seen as a rather partial view, to which has been added (or, erroneously, often counterposed) Thomas Kuhn's notion of paradigm shift of 'scientific revolution'. This contains a non-iterative, serendipitous leap, and a general question is what can we do about iterative methods that might embed us deeper and deeper in a wrong or suboptimal model? The speaker showed some awareness of this possibility when he indicated at several stages in his talk that his detective work might have been following completely the wrong lines.

Professor George Barnard has commented to me informally at this meeting that 'conjecture and confirmation' is the appropriate engine for quality improvement work, rather than Popper's maxim. However, popperian 'refutation' involves seeking out the weakest point of the scientist's current hypothesis, which is similar to Myron Tribus's proposal yesterday to replace the 'if it isn't broken, don't mend it' maxim by 'if it isn't broken, stress it to find the weak features – then improve it'. Thus, as George Box has pointed out, statisticians may have something to learn from philosophers of science.

S. Bisgaard. Mr Bibby's comment is a very welcome one. Preparing for this talk has given me the opportunity to reflect over philosophical issues involved in statistical work. I agree with the general spirit of the question; statistics should be considered in a broader scientific context and not as a sub-branch of mathematics (see Cox 1981). However, Mr Bibby quotes George Box as having said that 'statisticians may have something to learn from philosophers of science'. I don't know of George Box having said that, but I think it would be more to the point to say 'statisticians may have something to learn from the philosophy of science', but that 'philosophers of science may have something to learn from George Box!'

In reply to Mr Bibby's specific question about 'scientific revolutions' against iterative evolution of scientific understanding, may I quote Albert Einstein (1934), the scientist whose work most often is used in defence of the 'scientific revolution' paradigm, (my italics).

> Turning to the theory of relativity itself, I am anxious to draw attention to the fact that this theory is not speculative in origin; it owes its invention entirely to the desire to make physical theory *fit observed fact* as well as possible. We have here *no revolutionary act* but the natural continuation of a line that can be traced through centuries.

Personally, I think that scientific progress stems from an iterative process where some iterations (although not very often) may be serendipitous leaps. The model that in my view best 'fits observed facts' on how scientists and engineers actually work is provided by Box (1976). The feedback loop he discusses will most often ensure that we do not 'embed ourselves deeper and deeper in a wrong or suboptimal model'. We saw this corrective feedback loop in action several times in my case study.

Additional references

Cox, D. R. 1981 Theory and general principle in statistics. *Jl R. statist. Soc.* A **144**, 289–297.
Einstein, A. 1934 On the theory of relativity. In *Essays in science*, p. 48. New York: Philosophical Library Inc.
Friedel, R., Israel, P. & Finn, B. 1987 *Edison's electric light.* New Brunswick, New Jersey: Rutgers University Press.