P

APPENDIX

DISCUSSIONS AND INFORMATION

APPENDIX:

DISCUSSIONS AND INFORMATION

WELCOME

RECOMMENDED "OUT-OF-HOURS" PREPARATION

As mentioned near the end of the "Welcome" section, this is the reference list of preparatory tasks recommended for you to carry out between the different days of the course. I am not providing the full details here but rather referring you to the pages where you will find the relevant information. This will enable you to see the context in which the recommendations are made. The only necessary tasks that will take up any significant amount of time are those before Days 10 and 11 (these will get you much more ready and able to proceed with the Second Project) and, to a lesser extent, one to get you prepared for carrying out the Funnel Experiment on Day 3. All that remains are various small amounts of extra printing or (if convenient) enlarging.

Before Day 2. Just a reminder here that an ordinary basic calculator is likely to be useful to you on both Days 2 and 3.

Before Day 3. The more important task here is to prepare for the Funnel Experiment: see Day 2 pages 43–44 and Day 3 page 36. However, you may also find it useful to print a separate copy of Day 3 page 19 since the charts on that page are discussed in detail on the subsequent pages.

Before Day 6. For some extra pages of printing for your convenience during Day 6's Major Activity, etc: see Day 5 page 28. On this page there is also a reminder to prepare for "your organisation" if you don't belong to one (!).

Before Day 8. For some extra copies to aid you in Day 8's Major Activity: see Day 7 page 36 and Day 8 page 3.

Before Day 9. For just a little extra printing, mostly to aid you in Day 9's Major Activity: see Day 8 page 28.

Before Day 10. For details of your preparation for the Second Project: see Day 9 pages 26–30.

Before Day 11. For details of preparation for the second half of the Second Project: see Day 10 page 33.

Before Day 12. For a little equipment and one or more copies or enlargements: see Day 11 page 36.

Finally, if and when you decide to venture into the **Optional Extras** but without printing that file or the Workbook, there are just two pages in both Parts A and B which you will need to print out before studying those first two parts. They are pages 6 and 7 in Part A and pages 24 and 25 in Part B.

DAY 1

DISCUSSION ON THE FIRST PARADOX

In the "Springboard" article just mentioned on Day 1 page 8, as indeed in much other writing, I have largely followed the guidance of Dr Donald J Wheeler by using some language which is different from traditional terminology. Amongst living statisticians, I regard Don as the "world master" in understanding, using and teaching control charts in a way that is consistent with what Dr Shewhart gave us. Several years ago, Don became concerned that the word "control" was causing misunderstandings. Indeed, we shall show (and, more particularly, *you* will show) on Day 3 that attempts to "control" processes can often do more harm than good. So it was he that popularised the term "process behaviour chart" as an alternative to "control chart"—more of a mouthful but a far more appropriate description. Similarly, he preferred the term "process limits" rather than "control limits". I used Don's terms in the Springboard article, but I have already explained why I've mostly retained the traditional language during this course (see the small-print paragraph beginning at the bottom of Day 1 page 6).

Another great value of Don's work is that, whereas it used to be the case that control charts were generally considered to require small *samples* of data to be available at each time-point represented on the chart, he popularised an excellent and simple method of how to construct control charts when only *one* value is obtainable at any particular time. Samples are often easily obtainable in manufacturing processes, but single values are all you can get with the majority of other types of process. Since most people only have "one-at-a-time" data available in their processes, that type of data is all I consider in the main material of this course, in the Springboard article, and in the case studies covered in both *ST* and *EST*. However, in the latter little books, I do also briefly describe charts that are suitable for "a-few-at-a-time" data. If you are interested in reading much more on control charts than is needed for this course then there is plenty in the "Optional Extras" section (file S) which was introduced on page 8 of the introductory "Welcome".

In the extract from *EST* on Day 1 page 6, I referred to the control chart (process behaviour chart) as being "a remarkable statistical technique which, to all intents and purposes, was unknown at the time of the original edition". It was actually the above point to which I was then referring. In contrast, versions of control charts suitable for "a-few-at-a-time" data had already been well-known for decades, although mainly in manufacturing areas. Versions for some very specific kinds of "one-at-a-time" data were equally well-known. We shall be using one of those when we get to the Red Beads Experiment on Day 2 since the data there turn out to be suitable for that version. But what was still largely unknown around 1980 was the version suitable for virtually *any* type of "one-at-a-time" data—i.e. the only kind of data that the large majority of people ever see or use! This is why Don's contribution was, and is, so incredibly important.

If, either now or subsequently, you wish to obtain more substantial help and guidance on constructing and interpreting control charts than is included in my Springboard article or during this course (including the Optional Extras), I can do no better than point you to two of Dr Wheeler's many fine books. These are firstly the superb and relatively slim introduction I have already mentioned on Day 1 page 5: *Understanding Variation—the Key to Managing Chaos*, and then the more substantial *Making Sense of Data: SPC for the Service Sector.*

But now let us return to the first paradox summarised on Day 1 page 7: the apparently strange matter of neither the all-important tool of the control chart nor its creator, Dr Shewhart, even being mentioned in many Statistics courses nor, I suspect, even being known to most teachers of Statistics. Why?

I am tackling the first paradox here rather than in the main text because I do need to refer to matters covered in the usual kind of Statistics courses. If you have "no previous knowledge" of Statistics then there will be several matters that I'll mention here which will be unfamiliar to you. But don't worry about that—I think you'll still get the gist of what I'm saying!

The subject of Statistics has largely been developed by mathematicians: indeed it is sometimes regarded simply (and wrongly) as merely a branch of Mathematics. As you know, with regard to Statistics we are concentrating on "understanding variation", and I believe that most Mathematical Statisticians would not object to that little phrase as an apt description of their subject. So what's the difference?

Let's consider how a mathematician tackles his problems. It might appear ambitious to try to describe that in a single sentence, but here's my attempt! Mathematics essentially consists of virtually legalistic arguments: initially some definitions and terms and conditions are specified, and then everything that follows is deduced by pure logic in accordance with those terms and conditions. And that's fine for arithmetic, algebra, trigonometry, calculus, etc.

Mathematics has, of course, contributed greatly to the development of Statistics. And, reflecting our interpretation of "Statistics" as "understanding variation", Mathematics has indeed contributed greatly to understanding variation of many kinds, rather than just with our focus on *process* data, i.e. data from processes that are recorded over time. But, with the latter being our focus here, Mathematics turns out to be overly restrictive. Since a mathematical argument depends on its terms and conditions, what follows by logical argument will naturally be true under those terms and conditions. *But not necessarily otherwise.*

And there's the difficulty. What kind of "terms and conditions" could possibly be specified that would apply in practice to data from processes of all the kinds that we might want to study (manufacturing processes, administrative processes, financial processes, service processes, management processes, sales processes, medical processes, etc, etc)? In particular, Mathematical Statisticians love to assume that their data are normally distributed: beautiful mathematics can be carried out if that's included in the terms and conditions. And it *is* included in the terms and conditions for several statistical methods that are included in the conventional Statistics courses, e.g. *t*-tests, *F*-tests (including the tests used in Analysis of Variance and Experimental Design), the derivation of tests and confidence intervals in regression and correlation problems, etc. But the truth is that most (some would say *all*) real data from real processes are *not* normally distributed, maybe are not even *approximately* normally distributed—or maybe don't even have a distribution at all! What can be done then? That's a question which mathematicians tend to shy away from, because their beautiful mathematics just cannot be carried out in the absence of some such assumptions.

In our context, and with our special interest in processes being in, or at least heading toward, statistical control, we might try to narrow the breadth of the difficulty by proposing that the terms and conditions need relate *only* to the state of statistical control. And that is precisely what Shewhart initially did. I'll let him explain in his own clearly autobiographical words what happened (this is from page 12 of his 1939 book: *Statistical Method from the Viewpoint of Quality Control*):

> "Some of the earliest attempts to characterise a state of statistical control were inspired by the belief that there existed a special form of frequency function *[nowadays usually called a probability density function]* *f* and it was early argued that the normal law characterised such a state. When the normal law was found to be inadequate, then generalised functional forms were tried. Today, however, all hopes of finding a unique functional form *f* are blasted."

He could hardly have expressed it more pointedly than that! So what's the alternative?

I'll prepare the ground by quoting *verbatim* from a magnificent presentation that Deming gave to an audience at the Palace of Versailles, France in 1989, transcribed in BDA Booklet A6: *Profound Knowledge*; here I am quoting from pages 3–4 of this booklet. (You will see further extracts from this presentation this afternoon and on Day 3.)

He was talking about the two obvious kinds of mistakes that can be made when analysing process data: i.e. concluding that a process is in statistical control when it isn't, and *vice-versa*:

> "So what shall we do? Anyone can set for himself a clean record from this hour henceforth on *one* of the two mistakes. But, if he does, he will achieve the maximum loss from the other kind of mistake. This is a very easy way of making decisions: attribute anything that happens to a special cause—or attribute anything that happens to common causes. (A lot of things are easy to do!) So you minimise the loss from one kind of error at the expense of maximising the loss from the other kind. Yes, you can always minimise one, but not both—not both.
>
> Either kind of mistake causes loss. There is no way to avoid all of it—you can forget that! So we must resign ourselves to making both kinds of errors now and again—we hope not too often. We must aim to make them *with minimum overall economic loss.* It is all a matter of give and take. What shall we do? How shall we do it? Dr Shewhart helped us with these important questions, and this was a great contribution to man's process of thought and ability to manage.
>
> How can we aim for minimum economic loss? It is *nothing* to do with probabilities of the two kinds of mistakes. No, no, no, no; not at all. What we need is an operational definition of when to look for a special cause, and when not to. That is, a rule which guides us when to search in order to try to identify and remove a specific cause, and when not to. It is not a matter of probability. It is nothing at all to do with how many errors we make on average in 500 trials or 1,000 trials. No, no, no—it *can't* be done that way. We need an operational definition of when to act, and which way to act. Shewhart provided us with a communicable operational definition: the control chart using 3σ-limits. Shewhart contrived and published the rules in 1924: 65 years ago. Nobody has done a better job since."

After that final paragraph, those of you "in the know" will, I think, now begin to understand why "traditional" Mathematical Statisticians have tended to shy away from the control chart! But statisticians who are genuinely concerned with analysing real data from real processes owe Shewhart a considerable debt of gratitude in that he created and developed the control chart in a way which, truth to tell, does *not* depend on any such assumptions nor on "beautiful mathematics"!

Yet some people still try to claim that you must have normally distributed data in order to be able to use control charts—and many of them still insist that Shewhart said so!

> Three quick notes here. Firstly, Deming's use of the terms "common cause" and "special cause" will be described this afternoon. However, if you'd like to check on that now, see "bare bones" 3 and 4 on Day 1 page 32. Secondly, operational definitions will be studied on Day 11. And finally, "σ" (a Greek letter, pronounced "sigma") is a symbol commonly used by statisticians to represent a measure of variation.

There will be much more about these matters, including the "3σ-limits" mentioned near the end of the above extract, during the Optional Extras. Here I'll focus directly on the main point of contention as far as the orthodox Mathematical Statistician is concerned.

Shewhart's general ideas leading to the construction of the control chart are not contentious. But the stage comes when he needs to propose *in detail* the criterion for judging the process to be in or out of statistical control. Bearing in mind some of what Deming said above, the nature of what follows may now be of little surprise (although it might have been before you read that extract from the Versailles presentation).

In effect, Shewhart was considering the tightness or the looseness of the criterion which should be used as the dividing-line between judging the process to be in statistical control or out of statistical control, and then acting on the basis of that judgment. On page 277 of his 1931 book he delivered his proposed criterion in a single short sentence of the form:

"Experience indicates that *[his proposal]* seems to be an acceptable economic value."

No probabilities, no fancy theory, no assumptions about any particular probability distribution, etc—*experience*! Shewhart carried out a lot of experimentation with real data. Just like Deming in the Versailles presentation, Shewhart's reference to "economic" implied reducing as far as practical the cost of the control chart giving the wrong advice: i.e. regarding the process as in statistical control when it isn't and *vice-versa*. (Recall Deming's discussion of the "two mistakes" on the previous page.) Shewhart's extensive experimentation convinced him that his proposal (which we shall use on Days 2 and 3) seemed to do the trick. And as far as Deming was concerned more than half a century later, "Nobody has done a better job since."

But, yet again, it was *experience* that was Shewhart's deciding factor in deciding upon his proposal—*not* the "pure logic" of the mathematician's usual way of solving problems. And I would suggest that most practical people would agree that, regarding practical problems, experience surely has a rather important part to play!

And so I think we now have several reasons to explain the paradox being considered. Why isn't the control chart in most introductory (or even later) Statistics courses? First, we have the mistaken idea that the control chart is solely or primarily suitable just for manufacturing processes. But there are *hosts* of statistical methods used for studying manufacturing processes. So of course, if that were the prevailing idea, why should the control chart find a place in the standard courses on Statistics rather than any of the others? Awareness that the control chart is *fundamentally* important in studying processes of *all* kinds, especially the most important *management* processes, was effectively absent on this side of the world before Deming arrived on the scene in the 1980s. And, unfortunately, relatively few Mathematical Statisticians ever attended his four-day seminars. But, even if that awareness emerges, most mathematicians will still baulk at the idea of the crucial decision about the criterion to be used depending on *experience* rather than *pure logic*. That is not how most mathematicians think, and it is not how most mathematicians teach. So, even if its importance comes to be realised, it's still not "real" Mathematics, so it has no place in their courses.

I know. I was educated as a mathematician. So I cringed at the very idea. I avoided teaching it. I was in print a long while ago describing Shewhart's version of control charts as being "rough and ready". At the time, of course, that was intended as a criticism. I now re-interpret the same description as praise.

You know that in 1985 I initially thought of Dr Deming as a statistician and a fine mathematician. Of course, I soon found out that he was much more than either of those. The same could surely be said of Dr Shewhart, and of Dr Wheeler, and of many others. But these three (and, I suggest, just a very few others) should be described as unusually *wise* members of that group. Let's return to what we read from Shewhart near the bottom of page 3. His first thought about the nature of statistical control was that of a mathematician: i.e. maybe the relevant terms and conditions *could* include the Mathematical Statistician's favourite of normality. But he found that, in practice, many real processes sadly do not fit that condition. So he continued as a mathematician would and tried more complex mathematical models. But they also turned out to be unsuitable. So eventually he concluded that Mathematics just could not provide a solution that would be appropriate in practice (rather than just in mathematical theory). *And so he looked outside Mathematics.*

And there is the special wisdom, shared by a few but (I believe) not by many Mathematical Statisticians. Their first port of call when tackling a new practical problem is indeed likely to be Mathematics. Their wisdom is to be able to accept, after due investigation, that *Mathematics might not be able to provide the answer*, and then to proceed accordingly.

That is wisdom which I most certainly did not have in 1985. So I shall be ever grateful that both Drs Deming and Wheeler proved to be such excellent and patient teachers and friends.

(Continue on Day 1 page 11.)

PAUSE FOR THOUGHT 1–c

My acquaintance was, in fact, also now a member of staff in another part of the same university. This was around the time that British universities were first being subjected to "quality assurance" initiatives, quality audits, etc. Having suffered from these, my acquaintance jumped to the conclusion that we in the Quality Unit were responsible for such things! If you, the reader, similarly have thoughts that the Deming philosophy might have any such connections, you will soon find the truth to be very different.

Why should a university teacher dislike "quality assurance"? Doesn't he want to do a good job?

I believe that the large majority do indeed want to do a good job. But mostly they find that "quality assurance" hinders rather than helps. Yes, it can introduce some useful disciplines. But, to my friend and many others, the main effects are more and more paperwork, more and more "statistics", judgment, fear, blame, "massaging the figures", rules and regulations—which, the truth is, most teachers find *obstruct* the good job that they would like to do rather than *aiding* it. The comments about "performance indicators" on Day 7 will also be relevant here. What is needed is quality *improvement*, which turns out to be very different from "quality *assurance*".

(Return to Workbook page 2 / Day 1 page 21 or continue on Day 1 page 22.)

MAJOR ACTIVITY 1–f

Of course your behaviour would change—*massively*—depending on those different circumstances.

But *you* are still the same person. It's differences in *the system* around you, and the *effects* of those differences on you, that change your behaviour and performance.

I venture to suggest that if just this one message coming from Dr Deming's teaching were generally understood and appreciated then this world would be a better place. The emphasis would move away from attempting to improve performance by merely "doing things" to people toward concentrating instead on improving the system in which those people work and live. Please reread Day 1 page 41 which introduced this Major Activity—and see what you think.

DAY 2

PAUSE FOR THOUGHT 2–b

I can give some personal answers to these questions which have considerable relevance to what follows. (The "bullets" here relate to the four parts of this Pause for Thought.)

- A *very* long while ago, I used to teach control-charting the "conventional" way, i.e. trying to interpret control limits in terms of probabilities, saying that it is necessary to have normally distributed data, quoting the Central Limit Theorem to "justify" $\bar{X}$-charts, etc. Yet again, as previously, don't worry if you've never even heard of any such things—*you don't need to know!* But read on.

- I did this because of the way I had been taught—the conventional way. At the time I thought that the subject of Statistics was dependent on probability theory and all that that implies. So control charts, like everything else, had to be based upon such mathematical requirements.

- I changed my mind for two reasons. The first is that eventually I had two excellent and patient teachers: Dr Deming and Dr Don Wheeler. Secondly, I found in practice that control charts were enormously useful *without* being dressed up in the traditional mathematical finery, including situations where the traditional assumptions were clearly contradicted.

- Yes!

(Return to Workbook page 10 / Day 2 page 7 or continue on Day 2 page 8.)

PAUSE FOR THOUGHT 2–c

- "Illusion of knowledge" comes from how we are taught, both formally and informally. And, sometimes, the more impressive the teacher, the more dangerous he is! It also comes from jumping to conclusions: presuming something is "obvious" without thinking it through. Referring far ahead to Day 11, it comes from "experience and examples without theory".

- Yes!

(Return to Workbook page 11 / Day 2 page 8 or continue on Day 2 page 9.)

PAUSE FOR THOUGHT 2–e

What do we see? All the data are contained between the Lower and Upper Control Limits (LCL and UCL). Even Ernie's 3, just after the forthcoming performance appraisal was announced, is above the Lower Control Limit. There is *no* indication here that the process is out of statistical control. The control chart's assessment of the data is that essentially all the variation in the results is due to common causes, to the system, *not* in any way to the Willing Workers.

That is not what the Foreman thought—he thought it was *all* due to the Willing Workers!

(Return to Workbook page 13 / Day 2 page 16.)

ACTIVITY 2–f

Day 1

Audrey's 16	*"What a terrible start. But you've only just been trained—weren't you watching?"*
John's 9	*"That's better."*
Al's 4	*"Excellent: continual improvement."*
Carol's 7	*"What a disappointment."*
Ben's 9	*"No, no—it's WHITE beads that we want. Have you forgotten?*
Ed's 9	*"Don't just copy Ben: he's no example to follow."*

Day 2

Audrey's 10	*"At least it's not as bad as yesterday."*
John's 11	*"Hey—that's worse than yesterday. Concentrate!"*
Al's 9	*"You were yesterday's top performer—must have let it go to your head."*
Carol's 11	*"This is ridiculous."*
Ben's 17	*"Hold it—stop the line!"*
Ed's 7	[to them all:] *"**65?** That's a lot worse than the first week! Quite dreadful."*

Day 3

Audrey's 7	*"Audrey, you're in danger of impressing me."*
John's 12	*"Must I remind you all again?—the customer will not accept red beads."*
Al's 13	*"From bad to worse."*
Carol's 14	*"Even more ridiculous."*
Ben's 9	*"I'm glad you learned your lesson."*
Ed's 12	*"No, no, no. Remember how you did it yesterday."*
To them all:	*"How could you have done it again? That's even worse than yesterday."*

Day 4

Audrey's 6	*"You're a slow learner. But I'm proud of you."*
John's 10	*"Very consistent. Consistently bad."*
Al's 11	*"Look—if Audrey can get down to 6, anybody can get 6."*
Carol's 11	*"Has the rot stopped?"*
Ben's 13	*"I thought you'd learned your lesson."*
Ed's 7	*"You did that on Day 2. So why didn't you do it yesterday?"*

(Continue on Day 2 page 27.)

ACTIVITY 2–g

The upper and lower control limits for the statisticians' data are UCL = 16.4 and LCL = 0.5.

(Return to Workbook page 22 / Day 2 page 34.)

The upper and lower control limits for the Spaniards' data are UCL = 20.1 and LCL = 2.4.

(Return to Workbook page 25 / Day 2 page 37.)

Technical Aid 4

Statisticians' data

The statisticians' unusually low total number of red beads was 203.

So the average number of red beads obtained by the statisticians was $\bar{X} = 203 \div 24 = 8.458$.

The average *proportion* of red beads was $\bar{p} = \bar{X} \div n = 8.458 \div 50 = 0.1692$ which gives $1 - \bar{p} = 0.8308$.

So $\bar{X}(1-\bar{p}) = 8.458 \times 0.8308 = 7.0269064$, and $\sqrt{\bar{X}(1-\bar{p})} = \sqrt{7.0269064} = 2.6508$.

Finally, the distance from $\bar{X}$ out to the two control limits is $3\sqrt{\bar{X}(1-\bar{p})} = 3 \times 2.6508 = 7.952$. This gives the upper and lower control limits as

$$\text{UCL} = \bar{X} + 3\sqrt{\bar{X}(1-\bar{p})} = 8.458 + 7.952 = 16.41, \text{ and}$$
$$\text{LCL} = \bar{X} - 3\sqrt{\bar{X}(1-\bar{p})} = 8.458 - 7.952 = 0.51.$$

(Return to Workbook page 22 / Day 2 page 34.)

Spaniards' data

At the other extreme, the Spaniards produced no less than 270 red beads.

Then the average number of red beads obtained by the Spaniards was $\bar{X} = 270 \div 24 = 11.250$.

The average *proportion* of red beads was $\bar{p} = \bar{X} \div n = 11.250 \div 50 = 0.2250$ which gives $1 - \bar{p} = 0.7750$.

So $\bar{X}(1-\bar{p}) = 11.250 \times 0.7750 = 8.71875$ and $\sqrt{\bar{X}(1-\bar{p})} = \sqrt{8.71875} = 2.9528$.

Then the distance from $\bar{X}$ out to the two control limits is $3\sqrt{\bar{X}(1-\bar{p})} = 3 \times 2.9528 = 8.858$. This gives the control limits as

$$\text{UCL} = \bar{X} + 3\sqrt{\bar{X}(1-\bar{p})} = 11.250 + 8.858 = 20.11, \text{ and}$$
$$\text{LCL} = \bar{X} - 3\sqrt{\bar{X}(1-\bar{p})} = 11.250 - 8.858 = 2.39.$$

(Return to Workbook page 25 / Day 2 page 37.)

MAJOR ACTIVITY 2–h

There is, of course, so much to learn from the Experiment on Red Beads that I could not possibly attempt to cover everything here. Regarding messages from the experiments studied during today, along with Chapter 6 of *DemDim*, hopefully you now have a pretty comprehensive list of notes which can be used during this Major Activity. But in these few pages I shall include some further important thoughts and emphases that have not been sufficiently explored earlier.

Common-cause variation is much larger than expected

One important fact to learn and appreciate is how *large* common-cause variation (variation produced by the system) can be. With 20% of the beads being red, it might seem reasonable to assume that the long-term average number of red beads per worker per week would be 10. Actually, even that is untrue, as is evidenced on the next page. But suppose for the moment that it *is* true. If the overall average was 10 out of 50, naturally we would not expect everyone to get exactly 10 every time. But what would be a reasonable departure from 10? Now that you have seen various sets of results and a typical control chart, you know that the answer is something like 7 or 8 either way. It *is* reasonable to occasionally be as bad as 17 or 18. And you do *not* have to be specially gifted to occasionally get down to 2 or 3!

But take a look back at Activity 1–e (Day 1 page 25 *[WB 4]*) concerning the manager of the call-centre. What were your answers? The question there was in effect the same: how much variation away from 10 would you expect "by chance"? I presented that question to many of my seminar audiences before they knew the nature of the "work" in the Red Beads Experiment (during which they would, of course, then *see*—to their considerable surprise—the amount of variation to be expected by chance). The answer which I was given more often than all other answers put together was ... 10 plus or minus *2*! Anything between 8 and 12 is fair enough. But anything better than 8 or worse than 12 apparently invites comment!

This is a very common phenomenon (no pun intended!). Even when people get to understand the concept of common-cause variation, they almost always hopelessly underestimate how large it is, often by a factor of two or three times or even more. This is yet another reason why use of the control chart is so crucial. For without it, even *with* some understanding of common-cause variation, you'd still be likely to "see" more special-cause variation than actually exists. The result is that your good intentions about trying to improve the system will be far more likely to result in *tampering* with the system rather than improving it. ("Tampering" was Dr Deming's favourite word for describing this effect.) As we shall see tomorrow, the almost inevitable consequence is not only to fail to *improve* performance: it is actually to make performance *worse*. Hopefully, some of those words will sound familiar to you—refer back to the bottom of Day 1 page 17 and the top of Day 1 page 18. As you already know, in tomorrow's Major Activity you will be carrying out a version of the Funnel Experiment, and the problem of tampering is the main focus of that experiment.

The Tribus experiment

Many years ago the late Dr Myron Tribus, a respected consultant and speaker, took around with him some randomly-generated data, displayed in a format somewhat similar to that used in the Red Beads Experiment. When visiting any company or management team with whom he hadn't previously been involved, he presented them with these data and asked them to give their opinions on what should be done—as a useful indication about their current style and method of management. As Myron reported (see page 13 of BDA Booklet W2: *The Germ Theory of Management*[a]), during a lengthy period of time "only three out of thousands of people have suggested that perhaps the problem was in the system itself"—i.e. everybody else interpreted the data in a way appropriate to special-cause variation rather than to common-cause variation. Yet Myron's data *were* merely random—just like the red beads.

Financial information

One of the many needs for understanding the Red Beads process as a stable system is that, were this a real production process, such understanding would enable sensible decisions to be made on costing and pricing. With an unstable system, such decisions are always something of a gamble, but with a stable system the necessary information for prediction is available—prediction, not with certainty (which is never possible) but with a high, though never exactly quantifiable, degree of belief.

The big message

Practically, surely the most important conclusion of all is that, if improvement is desired (as it should surely be), it is the *system* within which the workers work that must be improved; the workers cannot do it by themselves. Improvement of the system, as has already been argued (and is similarly argued in *DemDim* Chapter 4), is *management's* responsibility. The red beads are already there in the system before our Willing Workers have had anything to do with them. Of course, having realised this, the job of the Willing Workers could be changed to include some sifting operation so that not so many red beads reach the customer (or even the Inspection Department). But this is a costly way to improve the figures. The red beads have already been made, and therefore paid for. If they came direct from the supplier, it is no good arguing that there may be a penalty clause so that he refunds us for the bad product. Like us, he has to balance his books, so *someone* has to pay, directly or indirectly. What about all the extra administrative costs, both his and ours, involved in our returning the faulty goods and getting the money back? Who pays? Or maybe, instead, the red beads have been produced early in our own operation. Again, they've been paid for. And, once they've been made, any subsequent inspection and sifting process is additional cost. As always, the moral must surely be that prevention is better than cure. As Bill Scherkenbach put it (on *Out of the Crisis* page 303[355]): "Search upstream provides powerful leverage toward improvement ... ."

Mechanical *vs* random sampling

Several further "profound messages" are to be found in the references provided earlier. But I shall content myself here with just one more. And this is (or should be) a particularly scary one for most statisticians. I implied earlier that the presumption of a long-term average of 10 out of 50 red beads per worker per week (based on there being 20% red beads in the container) is invalid. Before arguing why, let's restate some unarguable evidence which Deming himself provides (*Out of the Crisis* page 300[pages 351–352]). Over the years he tried four different paddles, two of them for large numbers of experiments. "Paddle No. 1, used for 30 years, shows an average of 11.3" whereas "the cumulated average for paddle No. 2 over many experiments in the past has settled down to 9.4 red beads per lot of 50". We do not know the numbers of experiments on which those figures were based; however, bearing in mind the length of time involved, statisticians would have little option but to conclude that not only is the difference between these averages "very highly significant" but that both of those averages are rather significantly different from 10!

> Here's a little exercise for the statisticians—all others should skip this small-print section.
>
> If you are keen on hypothesis tests, you might like to confirm that, if we try sample sizes (i.e. the number of experiments) of just 100 for both paddle No. 1 and paddle No. 2, the standard two-sided test for difference in means produces significance at about the α = 0.0002% level! The actual numbers of experiments were surely much larger than 100 and so the real significance would have been even stronger than that.

The almost automatic presumption that the long-term proportion of red beads obtained in the experiment must equal the proportion of red beads in the container is based on a false premise: that the sampling method used is—or is equivalent to—*random* sampling. Random sampling implies that every possible selection of 50 beads from the container is precisely as likely to occur as each and every other selection of 50 beads. This implies, in particular, that each and every bead (red or white) has exactly the same chance of being selected as any other at any time—the "mathematically ideal" model. Does the paddle do that? The numerical evidence plainly shows that it does not. Should we expect it to? A little thought would indicate not. The red and white beads are obviously different as regards colour—but is that their *only* difference? Unlikely. Very probably there are differences in weight, size, and surface "tackiness" or roughness. They might not just *look* different—they might *feel* different; and, if they feel different to our hand, why not to the paddle? Clearly, paddle No. 1 had a greater attraction to its red beads than did paddle No. 2 to its red beads. Just as clearly, 20% was a wholly irrelevant figure in both cases.

Random sampling can only properly be done by allocating a number to each bead and then selecting the beads according to genuine random numbers (or, more practically, using a well-researched and thoroughly tested system of computer-generated pseudo-random numbers). The paddle is clearly not random sampling. It is instead *mechanical* sampling.

But what kind of sampling does one carry out in industrial experiments and other "real-life" data-gathering: random or mechanical?

Right! So where does that leave those who depend in practical applications only on standard statistical theory based on ideal mathematical models and random sampling? Worrying, isn't it?

I hope so.

(If, at some time, you would like to study such matters in greater depth, you will find much to read in Parts C, D and E of the "Optional Extras".)

(Return to Workbook page 26 / Day 2 page 38.)

Dave Kerr's "Wrap-Up Brief" following the Red Beads Experiment

So indeed, at one level, this is a "stupidly simple" experiment. At another level, however, it is enormously rich and profound—and is it not scarily close to many real-life organisational situations and experiences?

Within the context of a desire to improve our organisations, I'd like to suggest to you that there are some key principles we can take and learn from the experiment that will help us to ensure our organisations, existing and new, do not become (either intentionally or accidentally) like the Red Beads company.

We need ...

- to "lift our eyes up" and *see the wider system*—customers, suppliers, employees, managers, stakeholders …, and we may also need to "look back in time".
- to understand something about numbers from processes that is still not widely taught, i.e. that they are *always* "uppy and downy". *[Dave thanks Mark Sheasby of the West Midlands Police for what he describes as this "wonderfully evocative phrase". My friend Peter Worthington, who contributes to Day 3, would refer to it as "wibble-wobble": clearly, the two descriptions are not unrelated!]* More technically, this is what we call *variation*—it is *not* the same as variety. Some key questions follow. As examples, when is the "uppy and downy" meaningful of anything, and how do we know; and what are the implications for decisions and consequent actions? You were all quick to note that the numbers in the experiment were "random"—yet the Foreman treated them as a factual basis for firing people!

- a practical and effective method for helping us to *really learn* about what is necessary for future improvement and success. Further, we need a method that helps to expose many of our false and limiting beliefs, such as the Foreman's apparent belief in the Red Beads Experiment that "what we need is better people".

- to understand something about how we *behave in organisations*—in particular, to know and understand the *interactions* between the *system* in which we are working and our behaviours. *[Think back to Day 1's Major Activity.]* In his work, Deming put very great emphasis on the distinction between intrinsic and extrinsic motivation, and he was strident in challenging the prevailing view that we are only motivated extrinsically: this, he believed, was both wrong and highly destructive. In his later work, Deming put great emphasis on *everyone's* (thus including managers) right to "Joy in Work". It is when we have "Joy in Work" that we are most likely to join in and contribute to the development and improvement of our organisation.

We also, very significantly, need to understand that *those four needs are* ***not independent*** *of each other*: indeed, the *interactions* between them are probably more significant and potent than any of them individually. In effect, we can think of these needs as *four interdependent components of a "system of thinking and seeing"*: a system that helps us to understand how organisations really work and thus helps us to understand how to be able to improve them—both continually and sustainably.

As an example, using this "system" to "view" the Red Beads company, we can e.g. see and understand how:

- they are not "seeing" the whole production system.
- the (poor) system of work and management is imposed on the employees (including the Foreman) and causes them to, for example,
  - ascribe all "uppy and downy" to people, and thence focusing action on "getting the best people;
  - treat the "Willing" Workers as *unintelligent units of production*. Just think how much creativity goes begging because of this kind of attitude!

Finally, but very significantly, we need to clearly understand that an absolutely vital function of leading is to give an organisational system *vision, meaning, direction and focus*—continually. Unclear and/or inconsistent purpose and direction and lack of focus create organisational ambiguity and individual confusion. In consequence, the organisation suffers from chaos and dysfunction.

(Return to Day 2 page 43.)

DAY 3

ACTIVITY 3–b

- We treat the words "variation" and "variety" in this context rather more precisely than is the case in general English usage. "Variety" indicates differences that have been deliberately introduced in order to provide a greater range of products and/or services compared with what would otherwise be available. Thus customers are able to choose what more closely meets their needs and desires. In this sense, variety is indeed the "spice of life".

  Differences may also be deliberately introduced for *experimental* purposes: one or more details involved with a process are altered in order to see what happens as a result—are those alterations beneficial or not? There is more on this in the final paragraph below.

  In contrast to both of those above, "variation" implies differences that most certainly have *not* been deliberately introduced: they are unwanted, they are a nuisance, they cause inconvenience or worse. They may be differences within the service or production process, making that process more inefficient than necessary, thus raising costs; they may be differences in the resulting service or product, whereas customers want to be able to rely on getting what they thought they were buying.

- Variation produces undesirable differences both internally and externally. Both are bad for the organisation concerned. As just observed, variation within processes causes inefficiency and thus extra cost. It may result in the need for increased inspection and troubleshooting—again costly. Variation experienced by the customer results in expensive repairs or replacements under warranty. Even worse for the customer, the problems may arise soon after the warranty has expired. In either case the company's reputation for poor reliability and service is thereby increased, leading to poorer future sales, etc. All such waste reduces the company's resources which could otherwise be available for innovation, experimentation, research—every one of which is necessary in order to successfully provide increased variety. So indeed, all this implies that reducing variation increases the feasibility of greater variety becoming available.

  There is a close analogy in the branch of traditional Statistics known as the Design of Experiments or Analysis of Variance (so if you're Stats-level 0 then you can skip this!). This topic had its origins in Agriculture: experimenting with various factors which might result in increased yield of crops—and some of the terminology from those origins still remains although the method is used extensively in many other areas. The "levels" (e.g. amounts, concentrations, etc) of one or more factors that are thought to probably affect results are deliberately varied in order to study the consequences. The resulting data are then analysed by comparing the variation resulting from different levels of the factors with what is usually referred to as the Residual Error or Unexplained Variation or something similar. The latter effectively plays the role of common-cause variation (though it is still quite rare for control charts to be used in this context, despite the fact that the mathematical assumptions made concerning the Residual Error are in fact much more stringent than simply being in statistical control!). Common-cause variation can be regarded as "fog": the denser the fog (i.e. the larger the common-cause variation), the harder it is to see anything; the thinner the fog, the easier it becomes to see things—including whether e.g. different levels of a fertilizer do indeed affect the yield. Thus the thicker the fog, the less possible it becomes to see the effects of experimentation, again harming the possibility of developing greater variety in what the company can offer to its customers.

(Return to Workbook page 32 / Day 3 page 4 or continue on Day 3 page 5.)

ACTIVITY 3–g

Using the first 12 values we have

$$\overline{X} = (18 + 19 + 17 + 17 + 16 + 17 + 16 + 15 + 14 + 15 + 15 + 13) \div 12 = 192 \div 12 = 16.00.$$

As before, the moving ranges are printed in italics:

18	19	17	17	16	17	16	15	14	15	15	13
	1	*2*	*0*	*1*	*1*	*1*	*1*	*1*	*1*	*0*	*2*

So $\overline{MR} = (1 + 2 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 0 + 2) \div 11 = 11 \div 11 = 1.00$ and $2.66 \times \overline{MR} = 2.66 \times 1.00 = 2.66$. This puts the control limits at 16.00 – 2.66 = 13.34 and 16.00 + 2.66 = 18.66, confirming the obvious message from the run chart that the process is trending downward.

(Return to Workbook page 37 / Day 3 page 18.)

MAJOR ACTIVITY 3–h

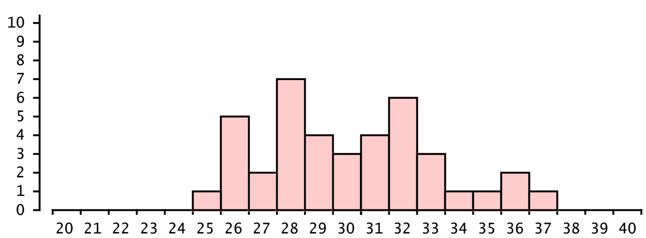
In order to aid my own learning from the Funnel Experiment, I wrote a computer simulation of the whole of this Major Activity, including fixing the first five dice-scores at the values you have been using. I then ran the simulation numerous times in order to study the kinds of similarities and differences that may occur. As a result, I can tell you about the kind of things that *usually* happen. But of course, by their very nature, data *vary*—and just now and again they vary in annoyingly *unusual* ways. And be sure that that happens with data other than those produced by throwing dice!

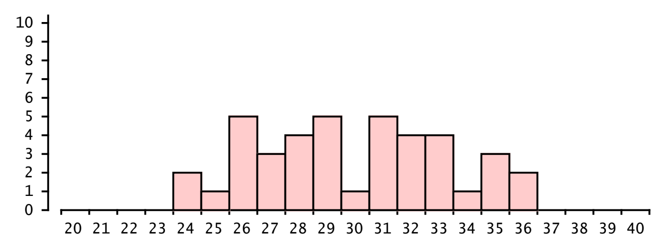
I have chosen to show you a couple of those simulations that demonstrate rather well both the similarities which *typically* occur and the kind of differences which *may* occur. (The sequences of dice-throws used in these simulations are those which I provided on Day 3 page 37 in case you couldn't find any dice!)

First, let's compare the histograms for Rules 1 and 2. If you look back to the similar histograms in the Ford example, remember that the Ford people were initially using Rule 2 but then eventually (on their statistician's advice!) tried Rule 1. I'll show the histograms here in that same order.

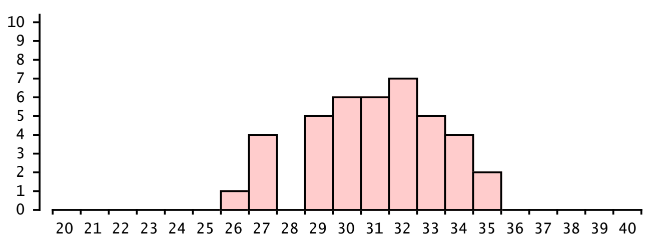
Simulation A *Simulation B*

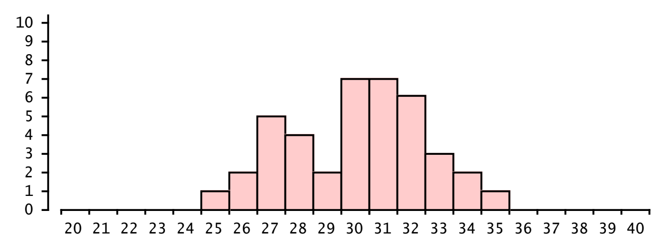
RULE 2





RULE 1



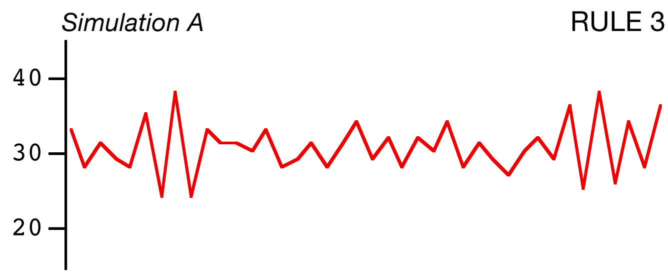


Because of our familiarity with the Ford example, there are no surprises here. Rule 2's histograms are rather wider than Rule 1's histograms, matching what was seen to occur in the Ford example. Rule 2 produces more variation than Rule 1. Using the kinds of measures of variation familiar to statisticians, Rule 2's variation is generally calculated to be around 40% greater on average than that of Rule 1. The automatic compensation device was indeed "tampering" with the system, making things worse rather than improving the system in any way whatsoever.

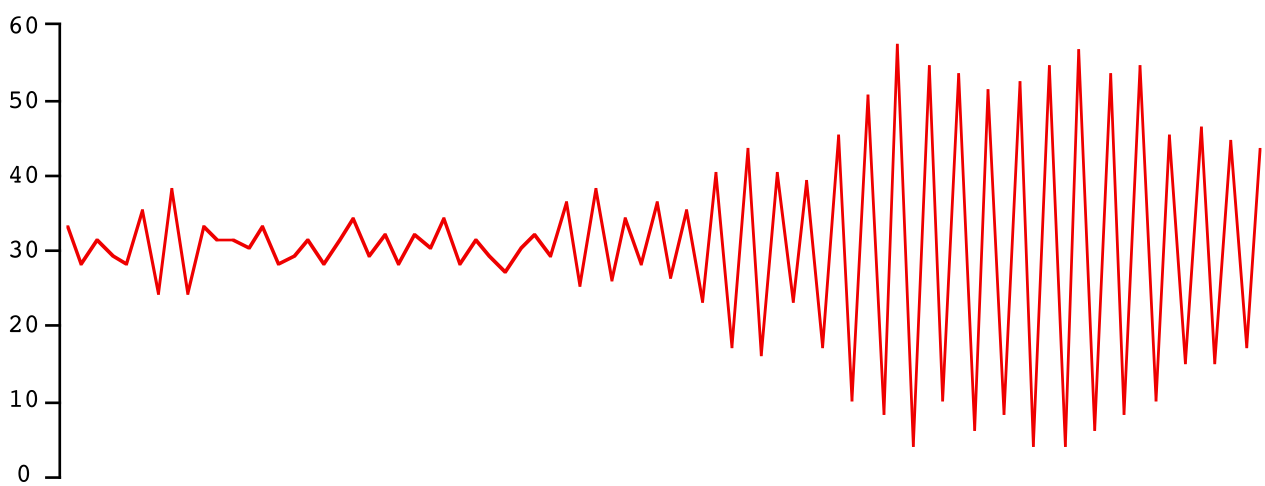
But beware of "over-analysing" histograms. The eye might be drawn to unusually short bars such as at 30 in Simulation B's Rule 2 and, even more so, the non-existent bar at 28 in Simulation A's Rule 1. They mean nothing—just "the luck of the draw"! As I said at the beginning, data by their very nature *vary*, and just now and again they vary in annoyingly untypical ways.

I imagine that one can still purchase automatic compensation devices. I suppose they may do some good with processes that are horribly unstable. But I think it would be rather less expensive and more sensible to use control charts, thus enabling focus on bringing the processes into statistical control and then using Rule 1 or, better still, get working on improving the process since more will then be known about it.

Let's move on to the run charts for Rules 3 and 4.



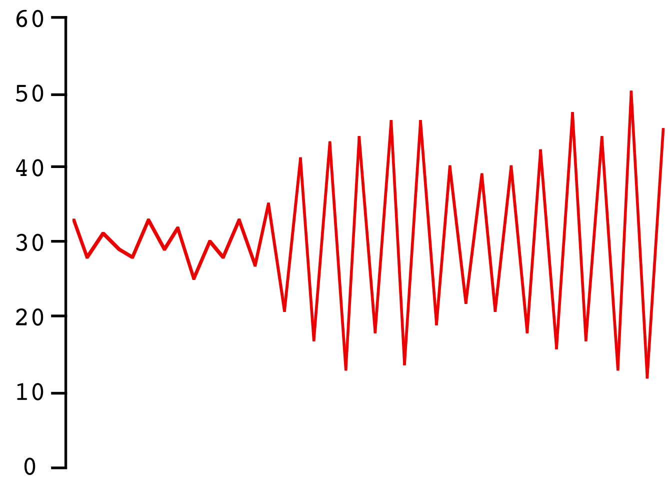
The run chart for Simulation A's Rule 3 surprised me. I imagine it may also surprise you, assuming you had the kind of fun and games with Rule 3 that often occur! This actually doesn't look too bad, although (again recalling Dr Worthington's teaching near the start of his seminars), the zig-zag tendency in the chart looks suspicious. However, this chart is very different from my own general experience and also was a strange exception to pretty much all the other simulations of Rule 3 that I looked at. Consequently I reran this particular simulation but this time let it develop over 80 stages rather than 40. And then I obtained:



That was more like it! Sooner or later, Rule 3 *always* produces horrific zig-zags—but they began considerably later in Simulation A than is usually the case. The Rule 3 run chart in Simulation B at the top of the next page was more typical.

Now, although Rule 2 resulted in wider variation than Rule 1, the difference might not be regarded as particularly dramatic. But the behaviour of Rule 3 usually *is* dramatic! Recall that I asked you to predict what you thought Rule 3 would produce before you started working with it—I wonder what you suggested! For remember that, when it was introduced, Rule 3 may have seemed to simply be a relatively innocent variant of Rule 2. Well, not so. It *was* rather easier to carry out: I expect you may have developed your Rule 3 data in maybe half the time that you took for Rule 2. (That's unless you became worried when huge zig-zags began to occur and you spent time checking for your mistake!) What was the difference between Rules 2 and 3? One way of expressing it is that, in Rule 3, attention to the vital factor of the funnel's position was instead replaced by undue and inappropriate attention to the target. That may raise some analogies in your mind!

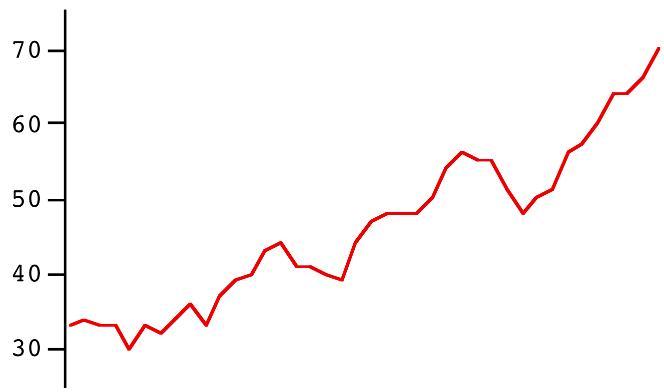
Simulation B RULE 3



Finally, onto Rule 4.

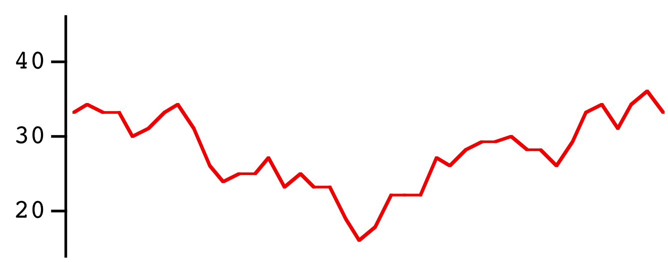
There are no wild zig-zags with Rule 4. Instead, Rule 4 usually just gently "wanders around" (mathematicians often refer to this as a "random walk"). It may mostly wander in one direction, as in Simulation A. But although the wanderings are "gentle", the consequences may well be rather unpleasant! Toward the end of this chart the school bus is arriving at around 9.10 rather than 8.30, and the socket is now around 2.7 cm wide rather than 2.3 cm. Not exactly improvement! Yet it *is* true that, as the motivation for Rule 4 was expressed, it does indeed reduce the point-to-point (short-term) variation compared even with Rule 1.

Simulation A RULE 4



In Simulation B, Rule 4 initially slowly drifts downward but then eventually moves back up to roughly where it started—i.e. close to the desired value of 30. At that stage, maybe this would have inspired some relief in those involved with the process: "It took a while to settle down, but now it looks OK!". But not for long … !

Simulation B RULE 4



Since there is considerable discussion on the Funnel Experiment in *DemDim* Chapter 5 and also in both *Out of the Crisis* (pages 280–284[327–332]) and *The New Economics* (Chapter 9), there is no need for much more here. Being wise after the event (if not before), Rules 2, 3 and 4 show differing levels of stupidity in trying to improve results. If a process is out of statistical control, the *only* way to improve it is to identify and deal with special causes. If a process is in statistical control, the *only* way to improve the results is to improve the process—not just "tamper" with it. So, as previously mentioned, isn't it frightening that, after learning from the Funnel Experiment, the delegates in both Dr Deming's seminars and mine would then so readily recognise having seen these Rules active in their own work and elsewhere in life? Several examples will be presented on pages 20–21 in the discussion on Activity 3–j.

So, returning to our version of the Funnel Experiment, if tampering doesn't do any good (to put it mildly), how might we really "improve the process", i.e. reduce the variation it produces? How about using some redesigned dice on which three of the faces show a 3 and the other three show a 4? That would do it! Or how might we improve the process in Lloyd Nelson's original version? Using a *very* thick tablecloth would help. And so would lowering the funnel closer to the table. Can you see the difference? These actions would *improve* the process: they would not be *tampering* with it.

I shall content myself here with just one further observation on each of Rules 2, 3 and 4.

Compared with what happens with Rules 3 and 4, Rule 2 doesn't look too bad. But that *is* only relative. Think back to your initial reaction to the Ford compensation example near the beginning of Day 3—as you now know, that was a direct application of Rule 2.

Rule 3's behaviour is clearly catastrophic—though did you suspect that such would be the case when (as recalled on the previous page) we may have initially regarded it as an "apparently rather innocent variant of Rule 2" on Day 3 page 48? A natural reaction when people first see its behaviour illustrated on a run chart is that, of course, this wouldn't be allowed to continue very long in practice: the strategy would be abandoned even if there was no real understanding of *why* its results were so horrible. Well, yes—that's if the time between successive data-points is short enough for the behaviour to be so apparent. But sometimes my delegates mused over longer-term changes in cultural or economic matters … . Even in Simulation B's run chart for Rule 3 (at the top of the previous page), things didn't start to get any worse than Rule 2 until around a dozen points had been plotted. If the time-interval between points were, say, a year, what is the likelihood that anyone would realise the subsequent catastrophic behaviour was caused by a policy instigated some 12 years earlier?

And Rule 4 is *really* dangerous. Remember its motivation is to reduce *short-term* variation—which it succeeds in doing. Thus often it really does appear to be even better than Rule 1—*in the short term*. And, yet again, it might be quite a while before it wanders off very far from the target—so that, when it eventually does so, as with Rule 3, it might not be easy to recognise the real source of the trouble.

You know that Dr Deming sometimes described the Red Beads Experiment as "stupidly simple". As I've indicated, the Funnel Experiment could surely be similarly described. But, as with the Red Beads, the messages to be learned from it are most certainly *not* stupid. They are profoundly important.

Forewarned is forearmed.

(Return to the bottom of Day 3 page 56.)

ACTIVITY 3–i

As you know, **Rule 1** simply produces the unadulterated data coming from the underlying process of throwing the dice or, in Dr Nelson's original version of the experiment, dropping the marble through the funnel. In particular, in our dice version the process is surely in statistical control with the control limits giving a fair indication of the likely range of the bulk of the data from that stable process. Even the extreme values of the dice-throws, i.e. 2 and 12, are not especially rare occurrences and so we would be rather unlucky if our computed control limits fail to enclose *all* the 11 possible outcomes.

I'll move straight on to **Rule 4** since that's the one you have dealt with most recently. The clue as to what happens to the control limits here follows directly from Rule 4's objective: "Since it is good to reduce variation, *[the management]* strive to minimise variation *at least in the short term*." The obvious result of that strategy is to *reduce* the moving ranges from what they would normally be under Rule 1. In other words, the gap between the control limits will be *narrower* than under Rule 1; on average, it is actually around 30% narrower. Because of Rule 4's "wandering" nature, it is likely to soon be producing plenty of points outside *either* type of control limits. However, that wandering nature soon becomes obvious with or without control limits; and so, unlike most kinds of special causes, there is nothing to learn from when exactly you start to get points outside the limits.

Apart from the feature of possibly wandering off to relatively enormous distances from wherever it starts, there is another aspect of Rule 4 which is often seen in real-world processes. This is that adjacent values are fairly closely "tied together" compared with the overall range of variation. For example, let's recall Process D of the Six Processes (described on Day 3 page 21). Suppose that, instead of my pulse-rate being measured just once a day, I was fitted with one of those monitors which records it, say, every minute. Or it could be measuring the systolic and/or diastolic blood pressure, or my body temperature, etc. All such measurements and many more may vary quite a lot over a reasonable amount of time but not usually over just a minute! So, in that sense, their variation could never be "random". If you want to impress your colleagues, this effect of adjacent measurements being closely "tied together" is called "positive autocorrelation"! The important point of which to be aware is that, as with Rule 4, control limits computed from such data will be *unnaturally close together*, and so you will be pretty much bound to get lots of points outside those limits before very long. However, now that you are aware of the problem, the remedy is obvious enough: make the time between readings long enough for the autocorrelation effect to become negligible. If you were to do this with pulse-rates etc then, as Chart D1 clearly showed (Day 3 page 19), it is entirely possibly to get a typical-looking stable control chart. However, with Rule 4 itself (because of its wandering nature) the time will come when virtually all points will be outside the limits.

Rule 3's zig-zag effect is the complete opposite. (This is *negative* autocorrelation.) "Zig-zag" *means* up-down-up-down-up … : i.e. here the moving ranges almost immediately start becoming *larger* than would be expected from the common-cause variation in Rule 1 and sooner or later become huge! So control limits from such data are often *much* wider than in the other Rules—but, even so, as the zigs and zags increase in size then points start falling outside even those very wide limits!

Now of course, in a sense, part of this discussion has become academic: in practice, when either Rule 3 or 4 behaviour starts to become at all extreme then it will surely be noticed and some kind of emergency action taken even if the cause of the trouble is not understood. But serious damage might already have been caused before then. However, there are still some very practical lessons to be learned regarding the computation of control limits. Even if there is no actual Rule 3 or 4 affecting the process, it is possible that, just by bad luck, your data may be behaving rather untypically during the baseline (the period which is used for computing the limits)—either unusually smoothly or with some big "zig"s and "zag"s. Regarding this matter, and whether or not you are on Stats-level 0, refer back to Technical Aid 9 on Day 3 page 27.

Rule 2 is yet another matter. Unusually, this overcompensation problem is often quite difficult to recognise on a control chart. There is some tendency for a rather low value to be followed by a rather high one and *vice-versa*, but this effect is not usually very pronounced nor long-lasting. For most of the time there is little difference to see compared with a properly stable process. The experienced eye might possibly notice a tendency for more "jaggedness" than usual. However, the main difference from Rule 1 is that which has already been seen from the histograms: Rule 2's higher level of variation compared with Rule 1. As previously mentioned, this *can* also be seen when comparing their control charts, but only relatively indirectly. I'd say the histograms have a big advantage here.

Regarding the control limits, there are two influences with Rule 2 that will widen the gap between them. Firstly, there is the approximately 40% increase in variation compared with Rule 1. Secondly, there is likely to be *something* of a zig-zag tendency although it is, of course, nowhere near as apparent as in Rule 3. The combination of these two influences may sometimes produce at least a hint of a "hugging the Central Line" effect, though it will certainly be nothing like as pronounced as in the "favourite example" (Day 3 page 24).

Some control-charting computer packages include a picture of a histogram turned through 90° (so that its scale coincides with the control chart's vertical scale) at the end of the control chart. This can occasionally be quite useful, especially when processes appear to be in statistical control.

As mentioned in the main text, if you'd like to examine the actual control charts discussed here, take some "time out" to read through Part A of the Optional Extras.

(Return to Workbook page 46 / Day 3 page 57 or continue on Day 3 page 58.)

ACTIVITY 3–j

Here is a selection of illustrations *suggested by delegates* in the earliest four-day seminars at which Dr Deming included the Funnel Experiment. As suggested, I have arranged it in two lists: one for Rules 2 and/or 3 and one for Rule 4.

Rules 2 and/or 3 ("zig-zag")

- Procurement of materials, volume planning, MRP systems
- Rifle adjustment, golf swing
- Adjustments made to inventories based on surge demand
- If I want my kids to be in bed by 8.00 pm and last night they could not get ready until 8.30 pm, to-night I tell them bedtime is 7.30 pm
- If you miss this month's shipment by $25,000, increase next month's goal by $25,000
- Reacting to a single customer complaint
- Calibration of an instrument
- Using reduced or tightened inspection based on results of prior lot(s)
- Adjusting work standards to reflect current performance
- Balancing an unbalanced oscillating ceiling fan

Some time ago my friend Fran Wheeler sent me a Calvin and Hobbes cartoon which she had recognised as a rather sweet illustration of the "zig-zag" effect. Particularly instructive is the way it showed not only what was happening to the factor suffering from the effect but also the sad consequence that had as a side-effect on another factor. Copyright restrictions prevent me from reproducing that cartoon, and so I must be content to just describe it to you:

> A little boy is in the bath (or, in view of the American source, I guess I should say "tub"!). He shouts for help:
>
> *Little boy:* The water's too cold!
>
> His mother rushes up to remedy the situation by running hot water into the bath. A little while later,

Little boy: Now it's too hot.

Mother adjusts the process to remedy the situation. A little while later,

Little boy: Now it's too cold.

Mother further adjusts the process. A little while later,

Little boy: Now it's too deep!

Rule 4 ("wandering")

- Matching colour by saving the last swatch
- Adjustment of time of a meeting based on the last actual starting time
- Copy examples. Learn by example with no theory.
- Hanging wallpaper
- If part does not fit gauge, fit gauge to part
- Reacting to rumour
- The man who blew the mill whistle sets his watch by the jeweller's clock on the way to work. The jeweller goes by the mill whistle.
- Interpretation of a law based on precedence
- Getting together to share ideas
- Changing company policy based on the last attitude survey

(Finally, see the bottom of Day 3 page 58.)

DAY 4

ACTIVITY 4–a

(1) In the event that you are having difficulties here, I believe that plenty of thoughts will come to you as we work through the 14 Points and the Deadly Diseases. But here are a few guidelines which may trigger some thoughts right now: anything that encourages quantity rather than quality, unreasonable deadlines, judgmental and unhelpful management, lack of cooperation at any level, quotas, commissions, fear, targets.

(Return to Workbook page 49 / Day 4 page 4.)

(2) Control limits inform us about the extent of variation being produced by the system. It follows that *where* a result lies between those limits—relatively high, medium or low—is in effect just a matter of luck. Thus the size of the *difference* between two values which are both between the limits is similarly just a matter of luck or, as Deming put it, will literally "mean nothing". And, almost always, the large majority of what or whom are being ranked *do* lie between the limits.

(Return to Workbook page 50 / Day 4 page 5.)

(3) Some that we shall consider later are pride of workmanship, indeed joy in work (see Day 1 pages 38–39), self-esteem, respect, reputation, willingness to cooperate, dignity, value of training, and (even more so) value of education.

(Return to Workbook page 50 / Day 4 page 5.]

ACTIVITY 4–d

(a) You are likely to talk about it in the case of Scenarios 1 or 3, with the obvious recommendations to avoid it or try it respectively. But with Scenario 2 there is little or nothing to talk about. A (merely) "satisfied customer" has nothing to complain about and nothing to rave about—so the experience is hardly much of a conversation-piece!

(b) With Scenario 1, obviously you won't be back there—either then or at any other time! With Scenario 3, you are very likely to be back there again: in fact, you will probably have been finding excuses to go and spend more of your money there long before another year has passed by. With Scenario 2 ("satisfied customer"), it is possible but unlikely you will be back there in a year's time. Particularly for such special events, you *want* something special; and, unless you are unlucky enough to have a life full of disappointments, merely being satisfied is not special.

Regarding the final part of the question, I suggest the answers are simply (1) harmed it; (2) nothing or worse (lost opportunity); and (3) benefited it.

This Activity reminds me of an e-mail I received from my friend Dave Young after he had read through parts of this course while on holiday early in 2015 (there is more from Dave elsewhere, especially on Day 12):

> "We come back to this hotel (the Copthorne Orchid in Penang) year after year, as do many more people from the four corners of the Earth. The hotel is old, run-down in parts, and badly in need of renovation. So why do we all keep returning here when we could easily afford to go wherever we wish in Penang? The staff, well-trained, respond immediately to our every request, and they provide each of us with the personal touch. They know us all by name and take a sincere interest in us and in our lives, our families, etc. Their simple customer service knocks the more "swanky" hotels (probably run by accountants) into a cocked hat. Your material could completely transform the hotel business here within a year."

(Return to Workbook page 54 / Day 4 page 10 or continue on Day 4 page 11.)

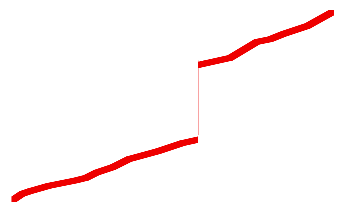
ADDENDA TO POINTS 1 AND 3

A couple of issues are discussed here which arise from the wording of Points 1 and 3: there was insufficient space to deal with these in the main text.

You might possibly regard the first issue as somewhat "quirky", but some people are puzzled by it. Both phrases "continuous improvement" and "continual improvement" clearly imply *unceasing* improvement—so does the difference in adjectives mean anything? Is it just a matter of semantics? You may be sure that Deming took great care here as elsewhere with his particular choice of words.

I imagine that his choice in this case was largely due to his background in Mathematics. For illustration, suppose it were possible to chart "quality" against time, and consider these two graphs:

Graph A



Graph B

We would all (including mathematicians) agree that Graph A shows continuous improvement. But the mathematician *wouldn't* describe Graph B as showing *continuous* improvement because it contains a *dis*-continuity—a jump or step-change. What could have caused that? Perhaps we have started using some new method or new material or other innovation which immediately and noticeably improves quality. You may be sure that Dr Deming wanted to include Graph B in his description! "Continual" simply implies that the improvement doesn't stop, *not* that it must improve *gradually*. So "continual" fits the bill without any ambiguity.

Part of our work on Day 8 will be to study Chapter 14 in *DemDim*: "Innovation—Not Just Improvement".

(Return to Workbook page 56 / Day 4 page 16.)

Secondly, the more general lesson to learn from the words "the need for" in Point 3 is as follows. Although it may be feasible to start work on adopting some of the 14 Points and on designing the cure of some of the Deadly Diseases without a great deal of delay, honest consideration must surely always be given as to *why* those practices that Deming advises us to get rid of, including the Diseases, are there right now. If there are some good reasons why they are there (rather than e.g. "We've always done it that way!"), clearly those reasons must be understood and addressed first. In the case of Point 3, if mass inspection is being used it is probably because the quality of our product or service is just not good enough. In other cases, the problem is there because, as Deming would say, "the management knows not how to manage". It becomes even more difficult if the reason is due to external influences or requirements. Whichever it is, if the practice or the Disease is dispensed with before the ground is adequately prepared, the obvious consequence is that a vacuum is left. I expect you know the old saying: "Nature abhors a vacuum". So then what happens? Something else floods in to fill the vacuum—and it could well be that that "something" is worse than what has just been thrown out! It is foolhardy to do anything just because "Deming said so" (as he readily agreed—see Day 1 page 22). Instead, you need good understanding of *why* "Deming said so" and also of *why* the problem he identifies is present. Your route forward needs to be mapped out with full attention to both.

(Return to Workbook page 60 / Day 4 page 20.)

DAY 5

Point 7. Institute leadership of people (Workbook pages 70–71 / Day 5 pages 2–3)

Obsession With Quality

This is leadership aimed at doing everything possible to *genuinely* improve the quality of what an organisation does and how it does it—as opposed to such prevalent approaches as trying to bribe or scare people into "doing their job properly". If you want to get the best out of people, I suggest that Dr Deming's "Attributes of a Leader" are likely to be rather more successful at achieving that aim.

All One Team

It is difficult to pick out anything very specific to say here: every item in the list of Attributes of a Leader contributes to the creation and development of All One Team.

Scientific Approach

A very strong emphasis on viewing an organisation as a *system* is obvious throughout the Attributes of a Leader. The relevance of the Scientific Approach to understanding people and how to manage them is particularly evident in Attribute 5. The use of understanding variation in this Attribute is a supreme example of how Shewhart's theory is applicable to far more than just shop-floor and manufacturing processes.

Point 8. Drive out fear (Workbook pages 72–73 / Day 5 pages 4–5)

Obsession With Quality

This is a case where what relates to the other two corners of the Triangle is also immediately relevant here. Regarding All One Team: although quality depends in part on individual effort, real quality is much more a joint effort, both across and up and down the organisation. People in fear do not work together: they are more concerned with personal survival—"What's in it for me?". Can you blame them? And, regarding the Scientific Approach: improvement and indeed innovation *need* valid, truthful information. Fear inhibits honesty. Also, we need to learn from mistakes. Understandably, those in fear do not even admit mistakes.

All One Team

All One Team implies working together for the benefit of all. But fear results in our working narrowly for what we see as personal survival, and "Devil take the hindmost". This is more "suboptimisation"—refer again to my All One Team comments for Point 5 (Day 4 page 25 *[WB 65]*).

Scientific Approach

Fear *increases* variation in so many obvious ways. People or groups "doing their own thing" for personal survival clearly increases variation. Another frequent observation from Dr Deming was: "**...** where there is fear, there will be wrong figures" (*Out of the Crisis* page 228[266]): people report what they can get away with and/or what they think management would *like* to see. Do you doubt it? Again, can you blame them? Where lies the fault?

Point 9. Break down barriers (Workbook pages 74–75 / Day 5 pages 6–7)

Obsession With Quality

As we have seen several times, the "Quality" being referred to at the top of the Joiner Triangle is quality on a very broad front. What happens in one part of the organisation can have dire effects on what happens elsewhere. As examples, suppose Purchasing buys "on the basis of price alone" (Point 4) or Sales promises delivery times that just cannot be met.

> My friend Alan Hodges observes that it would be valuable to add "Purchasing" to the list of areas in the statement of Point 9. Indeed so.

All One Team

The strong link here is wholly obvious. But note the implications concerning management style. Anything (and there is often much) that management does to *erect* or *strengthen* "barriers between departments or staff areas" harms and can indeed destroy any chance of All One Team. An obvious example is the setting of departmental numerical targets, particularly if reward and punishment are involved. It is often easy to improve your own figures by harming what goes on elsewhere (yes, more suboptimisation).

Scientific Approach

In addition to the emphasis yet again on improvement of the system, think of the additional variation that is caused if this Point is *not* put into effect. The different parts of the organisation pull in different directions to protect and advance what they see as their own local interests (suboptimisation again): all adding extra variation to overall performance. If the barriers have been erected, why should they even *try* anything else?

Point 10. Eliminate exhortations (Workbook pages 76–77 / Day 5 pages 8–9)

Obsession With Quality

The kind of poster suggested in that extract from *Out of the Crisis* would demonstrate that management is actively creating the environment and opportunity for improvement. Except for such an imaginary poster, use of posters, exhortations, slogans is surely relegating thoughts of quality to something entirely trivial—the direct opposite of *Obsession* With Quality.

All One Team

One of the slogans Dr Deming cites on *Out of the Crisis* page 57[66] is "Getting Better Together". He comments: "... workers have told me that this slogan makes them furious. Together! What is that when no one will listen to our problems and suggestions?".

Scientific Approach

Yet again, the thrust of this Point is that the bulk of quality, productivity, behaviour, performance—just about everything—comes from the *system* (common causes) not the *individual* (which might perhaps relate to some special causes). Yet at what or whom are the slogans, posters and exhortations that are referred to in this Point aimed? As my friend Mitch Beedie astutely pointed out to me: "Exhortations suggest that managers know what is wrong before they look".

Point 11. Eliminate arbitrary numerical targets (Workbook pages 78–79 / Day 5 pages 10–11)

Obsession With Quality

Again, this is quality "across the board". Those "arbitrary numerical targets" do the opposite—they focus attention on bits and pieces of the picture and, if associated with reward and punishment, encourage cheating even with those bits and pieces.

All One Team

Such targets (especially if associated with reward and punishment) that are set for individuals or groups or departments are one of the most effective destructors of teamwork at any level. Almost any number can be improved by causing harm elsewhere (one more time: suboptimisation).

Scientific Approach

Except for cheating, our understanding of variation teaches us that a number will be largely what the *system* provides. So, to reach a numerical target, one has to either (a) improve the system, or (b) bend the system in favour of that number (causing harm elsewhere), or (c) cheat. Come to think of it, doesn't that imply that cheating is also a product of such a system? So I guess I could have omitted the first three words of this paragraph ...

Point 12. Permit pride of workmanship (Workbook pages 80–81 / Day 5 pages 12–13)

Obsession With Quality

Anything which is a barrier to people taking pride in their work must surely, as an automatic consequence, be a barrier to quality. If there is quality, people are proud of their contribution to it. If they are denied the possibility of pride of workmanship, why should they contribute to what might otherwise have produced it?

All One Team

Most (if not all) of us know people who have particular skills, abilities, imagination and inspiration that provide invaluable contributions to what we and others can do. But do they not still need the help and cooperation of others to allow them the time and resources and environment to enable them to use their particular skills, etc effectively? *[The bottom half of Day 6 page 13 is worth reading on this.]* So don't the rest of us?

Scientific Approach

The Scientific Approach, and all the understanding and useful aids for improvement within it, is vital for improvement of the quality of what we do—which is precisely what enables us to have pride in our work.

Point 13. Encourage education (Workbook pages 82–83 / Day 5 pages 14–15)

Obsession With Quality

First, whereas training is very focused, "education" implies *breadth*—learning about other areas—and this often provides inspiration about what could be helpful for improvement in our own area. Further, an

Obsession With Quality is involved with innovation as well as improvement. Dr Deming's interpretation of "education" is therefore instrumental in encouraging and developing—not destroying—the creativity needed for innovation. Indeed, self-improvement (which is mentioned in the first sentence of the statement of this Point) is itself encouraged and stimulated by education. And who benefits? Win-Win.

All One Team

Another drawback of traditional "education" is its concentration on *individual* learning and either the sense or indeed the accusation of "cheating" if students work together. But most real learning in life and in the workplace happens *together* with others, combining individual and differing strengths, knowledge, expertise. C.f. the concept of *optimisation* of a system which produces far more than the mere sum of its parts.

> Looking far ahead, Dr Deming will have much to teach us about the concepts of optimisation and suboptimisation of a system in the material to be used on Day 10.

Scientific Approach

In Point 6 we observed that the Scientific Approach is certainly tied in with *training*, because of its tools content. But we know that this is the smaller part of the Approach. In the expanded Joiner Triangle are seen those two crucial, and by now very familiar, words: "Understand variation". Many people who have been taught the control chart just as a tool still do not understand variation: they merely obey rules. Understanding variation—and using control charts to *real* advantage—is much more a matter of *education* in Dr Deming's far-sighted sense of the word. Control charts do not simply give answers: far more importantly, they help us to *think* more intelligently and reliably.

Point 14. Top management commitment and action (Workbook pages 84–85 / Day 5 pages 16–17)

Obsession With Quality

Remember Dr Deming's familiar observation: "Quality is made at the top" (which was discussed on Day 1 page 20 *[WB 1]* in the very first Pause for Thought) or equivalently, on other occasions, "Quality is made in the Boardroom". For numerous reasons already discussed, Obsession With Quality is so different from conventional views and approaches to quality that what chance is there without Point 14?

All One Team

All *One* Team means what it says. Many approaches to quality improvement involve formation of teams, maybe lots of them—but, as previously observed (during Point 5 on Day 4 page 25 *[WB 65]*), the teams are sometimes set into competition with each other! So much of conventional management *obstructs* All One Team by doing precisely that: creating internal competition and, indeed, fear (of not winning). So, again—what chance is there without Point 14?

Scientific Approach

Here we go again! With the Scientific Approach's emphasis on the system (common causes), and with the system being management's responsibility: one more time—what chance is there without Point 14?

Disease 1. Lack of constancy (Workbook pages 86–87 / Day 5 pages 18–19)

Obsession With Quality

Recall the emphasis on *innovation* in the discussion of Point 1 in *Out of the Crisis*. There (on page 24[25] of his book) Dr Deming states unambiguously that "Innovation, the foundation of the future, can not thrive unless the top management have declared unshakable commitment to quality and productivity". I could equally well have included this in the discussion on Point 14. Yet again, "Quality is made at the top".

All One Team

There can only be All One Team if the team knows what it is working on and is allowed to work on. Frequent changes of direction and priorities are fractious in their effect. At the very least, those who are causing the lack of constancy of purpose will become estranged from those who are suffering from it.

Scientific Approach

Constancy of purpose automatically reduces variation. The opposite is also true: *lack of* constancy of purpose automatically *increases* variation. Also, as we know, the thinking and the methods of the Scientific Approach are there to aid the learning of how to do something better. All well and good. However, that's of but limited value if that "something" is regarded as important today but no longer so regarded tomorrow. Research and experimentation aimed at innovation is considerably enhanced by reduced variation: it thins the "fog" which otherwise obstructs the view and hence the understanding of what innovation may be able to produce.

Disease 2. Short-termism (Workbook pages 88–89 / Day 5 pages 20–21)

Obsession With Quality

When you think of the many contributors needed to develop real quality in whatever you are doing, there is one common requirement: *time*! Time to learn, time to experiment, time to discover whom to work with, who can best help. So, by definition, short-termism destroys any possibility.

All One Team

All One Team includes development of personal relationships, learning to work together, learning to help each other for mutual benefit and the benefit of the company and its customers. A short-termism culture forces you to "look out for yourself" rather than contributing to the whole. After all, a straightforward way for management to increase short-term profits is to let some people go.

And what a dishonest term that is—it sounds as if they *want* to go and you're just giving them permission.

Scientific Approach

Short-termism implies knee-jerk reactions: "Act now, pay later". Remember "tampering"? What could more effectively *increase* variation?

The destructive nature of short-termism on the Scientific Approach is compactly expressed by Dr Deming in that sentence: "Cut down on research, education, training". In other words: "Cut down on the Scientific Approach, both on what it needs and hence on what it can produce".

Disease 3. Appraisal of performance (Workbook pages 90–91 / Day 5 pages 22–23)

Obsession With Quality

Obsession With Quality *has* to imply obsession with improving the *system* which produces the quality—*not* imagining that quality can merely be obtained by "doing things" to people, especially of a "motivational" kind. It does not work. It will not work.

All One Team

Here are some words from the video *Management's Five Deadly Diseases*. If you watched it earlier, I think you will remember them:

> "Annual appraisal of performance ... annihilates teamwork: people can not work together ... you have to get ahead. By working with a team, you help other people. You may help yourself equally, but you don't get ahead by being equal: you get ahead by being ahead."

Scientific Approach

As covered in my discussion, the whole concept of judgmental performance appraisal flies in the face of the fundamental understanding of variation: where does most variation come from, and within what are the major opportunities for improvement? The system.

Disease 4. Management job-hopping (Workbook pages 92–93 / Day 5 pages 24–25)

Obsession With Quality

Obsession With Quality implies having the *time* to learn how to genuinely improve quality in the specific organisation and environment rather than e.g. just changing some figures which may well have only tenuous relevance to real quality.

All One Team

This is where Dr Deming saw this Deadly Disease as being its most damaging (*Out of the Crisis* page 103 [121]):

> "Mobility annihilates teamwork, so vital for continued existence. A new manager comes in. Everyone wonders what will happen. Unrest becomes rampant when the board of directors go outside the company to bring someone in for a rescue operation. Everyone takes to his life preserver."

Scientific Approach

It is evident that mobility of management increases variation; how could it do otherwise? In the case of a "job-hopping" culture as opposed to a learning culture, new managers must quickly justify their existence, make their mark. As we have seen, Dr Deming says it in typically few words: "Mobility of management causes instability". "Instability" is one of the ways in which we describe "out of statistical control".

Disease 5. Use of only visible figures (Workbook pages 94–95 / Day 5 pages 26–27)

Obsession With Quality

With that summary of invisible figures in mind, the obstructive effect to real quality of dependence on *only* visible figures becomes immediately evident—frighteningly so when realising how universal is this fifth Disease, even in those organisations that are making progress in curing the others.

All One Team

The summary already includes the value of a specific team as amongst the important invisible figures. How much more massive and yet invisible (regarding actual numbers) is the value of *All One* Team? The truth is that just about every advantage of All One Team which one can identify (and surely there are many) has literally unmeasurable value.

Scientific Approach

A delegate at a four-day seminar ventured to quote the well-known saying: "If you can't measure it, you can't manage it". Deming's typically brief and pointed response was: "You'd darn well *better* manage it".

With regard to that response I have often noticed how, once people have become familiar with control charts and how to interpret them—particularly in terms of changing focus onto the system rather than scratching around for possible special causes—they would similarly change their behaviour even *without* visible figures.

When relevant visible figures are available then, of course, we need to learn how to use them appropriately. The appeal of the proliferation of tools and techniques, and the enthusiasm of consultants to sell them, is apparent. But they do not teach many tools and techniques suitable for analysing unknown and unknowable figures.

There is considerable wisdom in understanding the *limitations* of figures as a vital part of maturity in the Scientific Approach—or indeed of a statistician—but it is rare. For some exceptions, the quotations from wise statisticians at the start of *DemDim* Chapter 10 (pages 151–152) are worth much reflection.

DAY 7

ACTIVITY 7–a

Just a few comments here that I hope may be helpful. Regarding Category 1—"facts of life", i.e. matters that are unarguable and not a matter of opinion, remember that in Point 11 Deming used the word "arbitrary": "Eliminate *arbitrary* numerical targets." So that rules out "facts of life" since, by definition, facts are not arbitrary and therefore, most certainly, cannot be eliminated. Regarding Categories 2 and 3, I will just re-emphasise that many suggestions which can arise in this Activity *could* go into either category—it all depends on how they are used. Will the number be used for guidance and help, or will it be used for judgment and possibly punishment or reward?

On Day 7 pages 7–8, I include what I trust are some wise words on putting theory into practice. I advise doing things gradually rather than precipitously and I spend a little time there on performance appraisal in particular. But, regarding numerical targets, I suggest that the little I have just said provides some useful clues. As part of the move toward better management, those who set targets should consider well their "arbitrariness" and also shift the balance in how they are treated and used from Category 3 to Category 2.

(Return to Workbook page 105 / Day 7 page 4.)

TRANSCRIPT OF LETTER

Sir: I have to take issue with the assertion in your leading article that Ofsted inspections are a "rigorous procedure ... which has done so much to keep standards high". I went to a school that appears high in your tables and recently received an excellent Ofsted report, possibly in part due to the fact that, whenever an inspector was present, each question asked by the teacher would be greeted with a sea of raised hands, and that the pupil selected to answer would almost always give a perfect response.

How? We had all been coached to raise our right hands if we knew the answer, and our left hands if we didn't. Any child at our school could have told you that the quality most rewarded by Ofsted is not so much excellent teaching as low cunning.

M HOPWOOD
Queen's College
Oxford

(Return to Day 7 page 13.)

MORE ON TARGETS IN GOVERNMENT

In connection with what is often referred to as the UK Government's "hostile immigration environment", Home Secretary Amber Rudd told the Home Affairs Select Committee on 25 April 2018: "We don't have targets for removals". A little later, she said: "If you're asking me: 'Are there numbers of people we expect to be removed?', that's not how we operate." The next day, evidence of regional removal targets came to light: Ms Rudd apologised, saying that she had not been aware of their existence. On 28 April the *Guardian* revealed a six-page memo dated 21 June 2017 that was sent to Ms Rudd and others. It referred to the Home Office's "target of achieving 12,800 enforced removals in 2017–18"; it also said that: "We have exceeded our target of assisted returns". Ms Rudd insisted she had never seen it. On 29 April a letter, sent to Prime Minister Theresa May in January 2017, was leaked in which Ms Rudd outlined her plans to increase removals by 10%. In the evening of 29 April she resigned from her post as Home Secretary.

PS On 16 November 2018 she returned to the Government as Secretary of State for Work and Pensions (!!)

(Return to Day 7 page 18.)

MAJOR ACTIVITY 7–i

1. Hope for instant pudding.

Letters and telephone calls received by this author disclose prevalence of the supposition that one or two consultations with a competent statistician will set the company on the road to quality and productivity—instant pudding. "Come, spend a day with us, and do for us what you did for Japan; we too wish to be saved." And they hang up in sorrow. It is not so simple: it will be necessary to study and to go to work. One man actually wrote to me for my formula, and the bill therefore.

An example of expectation of quick results without effort and without sufficient education to the job is exemplified in a letter received by Dr Lloyd Nelson, statistician with the Nashua Corporation, which reads as follows:

> "The President of my company has appointed me to the same position that you hold in your company. He has given me full authority to proceed, and he wishes me to carry on my new job without bothering him. What ought I to do? How do I go about my new job?"

Appointment of someone to the same job that Dr Nelson has will not create another Dr Nelson. It would be difficult to convey in three lines so much misunderstanding. The President's supposition that he can resign from his obligation to lead improvement of quality is a glaring fallacy. And who would accept such a mandate from the head of an organisation? Only someone that is a complete novice on quality and improvement of quality.

(Return to Workbook page 115 / Day 7 page 28.)

2. The supposition that solving problems, automation, gadgets, and new machinery will transform industry.

A group of workers took pride in changes that saved $500 a year. Every net contribution to efficiency is important, however small.

The big gain is not the $500 per year that the men saved. What is important is that these men could now take pride in the improvement. They felt important to the job and to the company. The quality of their output improved along with outturn *[i.e. the quantity]*. Moreover, this improvement brought forth better quality, productivity, and morale all along the line. This improvement can not be quantified. It remains as one of the invisible figures, so important for management.

(Return to Workbook page 115 / Day 7 page 28.)

3. Search for examples.

A man just called on the telephone from Johannesburg with the proposal that he come to this country and visit with me six companies that are doing well. He needed examples, he said.

Too often this is the story. The management of a company, seized with a desire to improve quality and productivity, knowing not how to go about it, having not guidance from principles, seeking enlightenment, embark on excursions to other companies that are ostensibly doing well. They are received with open arms, and the exchange of ideas commences. They (visitors) learn what the host is doing, some of which may by accident be in accordance with the 14 Points. Devoid of guiding principles, they are both adrift. Neither company knows whether or why any procedure is right, nor whether or why another is wrong. The question is not whether a business is successful, but why? And why was it not more successful? One can only hope the visitors enjoy the ride. They are more to be pitied than censured.

It was related to me during a seminar that the management of a company that makes furniture, doing well, took it into their heads to expand their line into pianos. Why not make pianos? They bought a Steinway piano, took it apart, made or bought parts, and put a piano together exactly like the Steinway, only to discover that they could only get thuds out of their product. So they put the Steinway piano back together with the intention to get their money back on it, only to discover that it too would now only make thuds.

(Return to Workbook page 116 / Day 7 page 29.)

4. **"Our problems are different."**

A common disease that afflicts management and government administration the world over is the impression that "Our problems are different." They are different, to be sure, but the principles that will help to improve quality of product and of service are universal in nature.

And that is all Dr Deming wrote on this one.

(Return to Workbook page 116 / Day 7 page 29.)

5. **Obsolescence in Schools *[of Business]*.**

Most students in Schools of Business in America have had no experience in production or in sales. To work on the factory floor with pay equal to half what he hoped to get upon receipt of the MBA, just to get experience, is a horrible thought to an MBA, not the American way of life. As a consequence, he struggles on, unaware of his limitations, or unable to face the need to fill in the gaps. The results are obvious.

Faults in education were stated well by Edward A Reynolds in *Standardization News* (Philadelphia), April 1983, page 7:

> "There are many reasons why US quality/productivity (they go hand in hand) have not kept ahead. A few of the major ones are: the educational system that turns out math ignoramuses and emphasises the MBA (which teaches management how to take over companies, but not how to run them); the short-term goals of corporate heads (this year's profit for bonus or a better job elsewhere); the practice of moving about the country, and finally out of it, for cheaper labour (despite the fact that direct labour is a small minority of costs); the change from honest leadership and work ethic to 'get yours' and 'everyone does it' at all levels.
>
> Practically all of our major corporations were started by technical men—inventors, mechanics, engineers, and chemists, who had a sincere interest in quality of products. Now these companies are largely run by men interested in profit, not product. Their pride is in the Profit & Loss statement or stock report."

(Return to Workbook page 117 / Day 7 page 30.)

6. **Poor teaching of statistical methods in industry.**

Awakening to the need for quality, and with no idea what quality means nor how to achieve it, American management have resorted to mass assemblies for crash courses in statistical methods, employing hacks for teachers, being unable to discriminate between competence and ignorance. The result is that hundreds of people are learning what is wrong.

People with Master's degrees in statistical theory accept jobs in industry and government to work with computers. Most if not all computer packages for analysis of data, as they are called, provide flagrant examples of inefficiency. It is a vicious cycle. Statisticians do not know what statistical work is, and are satisfied to work with computers. People that hire statisticians likewise have no knowledge about statistical work, and somehow suppose that computers are the answer. Statisticians and management thus misguide each other and keep the vicious cycle rolling.

(Return to Workbook page 117 / Day 7 page 30.)

7. Use of *[*Military Standard 105D and other*]* tables for acceptance.

Many thousands of dollars worth of product changes hands hourly, lots *[i.e. batches]* subjected to acceptance or rejection, depending on tests of samples drawn from the lots. Such plans can only increase costs, as will be clear from Chapter 15 *[in* Out of the Crisis*]*. If used for quality audit of final product as it goes out the door, they guarantee that some customers will get defective product. *[Recall I have just pointed out that "an essential requirement for deciding on the details of such a plan is the specification of an acceptable proportion of defectives".]* The day of such plans is finished. American industry can not afford the losses that they cause.

Incredibly, courses and books in statistical methods still devote time and pages to acceptance sampling.

(Return to Workbook page 118 / Day 7 page 31.)

8. "Our Quality Control Department takes care of all our problems of quality."

Unfortunately, Quality Control Departments have taken the job of quality away from the people that can contribute most to quality—management, supervisors, managers of purchasing, and production workers. They have failed to explain to management the importance of good management, including the evils of (e.g.) purchase of materials on the basis of price tag, evils of multiple vendors, of work standards, awkward and costly arrangements of plant. Management, mystified by control charts and statistical thinking, are glad to leave quality to people that mystify them.

The function of quality assurance in many companies is too often to provide hindsight, to keep the management informed about the amount of defective product produced week by week, or comparisons month by month on levels of quality, costs of warranty, etc. What management needs are charts to show whether the system has reached a stable state (in which case management must take on the chief role for improvement), or it is still infested with special causes.

(Return to Workbook page 118 / Day 7 page 31.)

9. "Our troubles lie entirely in the workforce."

The supposition is prevalent the world over that there would be no problems in production or in service if only our workers would do their jobs in the way that they were taught. Pleasant dreams.

Only recently, the entire management of a large manufacturing concern supposed, by their own declaration, that if all 2,700 operations in the plant were carried off without blemish, there would be no problems. I listened for three hours to their exciting achievements with statistical methods on the shop floor. Their engineers, I found, were treating every problem as a special cause—find it and remove it—not working on the system itself. At the same time, costs of warranty were soaring upward, and business was on the decline. The management seemed to be totally unaware of the need for better design of their main product, and more attention to incoming materials. Why did they put so much faith in statistical methods on the shop floor? Answer: What else is there? Quality is for other people, not for us.

Presumably the "statistical methods" alluded to here were those relating to Obstacle 6.

(Return to Workbook page 119 / Day 7 page 32.)

10. False starts.

False starts are deceptive. They give satisfaction, something to show for effort, but they lead to frustration, despair, disappointment, and delay.

One kind of false start arises from the supposition that wholesale teaching of statistical methods to enough people in production will turn things around. Understanding of variation, special causes and common causes, and the necessity to reduce constantly the variation from common causes, is vital. It is a fact, though, with a clean record, that a company whose management abrogates their responsibility for quality and depends wholly on statistical methods on the shop floor, and forces these methods on to suppliers, will within three years toss those methods out, along with the people engaged on them.

I refer you yet again to the final "bare bone" of the ten in Shewhart's Breakthrough (Day 1 page 33).

(Return to Workbook page 119 / Day 7 page 32.)

11. "We installed quality control."

No. You can install a new desk, or a new carpet, or a new dean, but not quality control. Anyone that proposes to "install quality control" unfortunately has little knowledge about quality control.

Improvement of quality and productivity, to be successful in any company, must be a learning process, year by year, top management leading the whole company.

As with Obstacle 4, this is all that Deming wrote here: brief but very much to the point.

(Return to Workbook page 120 / Day 7 page 33.)

12. The unmanned computer.

Some people make good use of computers. Few people are aware, however, of the negative input of computers. Time and time again, in my experience, when I ask for data on inspection, to learn whether they indicate that the process is in control, or out of control, and at what time of day it went out, and why, or ask about differences between inspectors and between production workers, or between production workers and inspectors, in an attempt to find sources of trouble and to improve efficiency, the answer is "The data are in the computer". And there they sit.

People are intimidated by the computer. They can not tell it what data or charts they need: instead, they take whatever the computer turns out.

A computer can be a blessing. It can also be a curse.

(Return to Workbook page 120 / Day 7 page 33.)

13. The supposition that it is only necessary to meet specifications.

Specifications can not tell the whole story. The supplier must know what the material is to be used for. If the supplier knows that the steel will be used for the inside door panel of an automobile, he may be able to supply steel that will do the job. Steel that merely meets the specifications can cause a lot of trouble. A vice-president in charge of manufacturing told me that half his problems arise from materials that met the specifications.

A programmer has a similar problem. She learns, after she finishes the job, that she has programmed very well the specifications as delivered to her, but that they were deficient. If she had only known the purpose of the program, she could have done it right for the purpose, even though the specifications were deficient.

My friend Robert Piketty of Paris put it this way: listen to the Royal Philharmonic Orchestra of London play Beethoven's Fifth Symphony. Now listen to some amateur orchestra play it. Of course, you like both performances: you enjoy home-grown talent. Both orchestras met the specifications: not a mistake. But listen to the difference. Just listen to the difference!

(Return to Workbook page 121 / Day 7 page 34.)

14. The fallacy of zero defects.

There is obviously something wrong when a measured characteristic barely inside a specification is declared to be conforming; outside it is declared to be nonconforming. A better description of the world is the Taguchi loss function in which there is minimum loss at the nominal value, and an ever-increasing loss with departure either way from the nominal value.

Yes, Dr Deming did write this under Obstacle 14, not Obstacle 13.

It will not suffice to have customers that are merely satisfied. An unhappy customer will switch. Unfortunately, a satisfied customer may also switch, on the theory that he could not lose much, and might gain. Profit in business comes from repeat customers, customers that boast about your product and service, and that bring friends with them. Fully allocated costs may well show that the profit in a transaction with a customer that comes back voluntarily may be 10 times the profit realised from a customer that responds to advertising and other persuasion.

(Return to Workbook page 121 / Day 7 page 34.)

15. Inadequate testing of prototypes.

A common practice among engineers is to put together a prototype of an assembly with every part close to the nominal or intended measured characteristic. The test may go off well. The problem is that when the assembly goes into production, all characteristics will vary. The fact is that volume production may turn out only one part in 100,000 that will perform like the prototype.

Anyone engaged in testing should ask himself the following questions:

1. What will the results refer to?
2. Will they refer to tomorrow's run or to next year's crop?
3. Under what conditions will these results predict results of tomorrow's run or of next year's crop?

Failure to understand variation in tests delayed some years the science of genetics. Ratios of (e.g.) tall and dwarf peas varied wildly below and above nature's average value 1:4. This variation bothered everyone, including Gregor Mendel, discoverer of the simple dominant gene.

(Return to Workbook page 122 / Day 7 page 35.)

16. **"Anyone that comes to try to help us must understand all about our business."**

All evidence points to the fallacy of this supposition. Competent men in every position, if they are doing their best, know all there is to know about their work except how to improve it. Help toward improvement can come only from some other kind of knowledge. Help may come from outside the company, combined with knowledge already possessed by people within the company but not being utilised.

And for the third time, here I have quoted all that Dr Deming wrote for a particular Obstacle.

This Obstacle struck a loud chord for me. In my very early days of trying to help a company learn about the Deming philosophy, I would spend much time before my visit trying to learn as much as I could about that company: I did not want to seem ignorant about what they did or how they did it. It was an utter waste of effort. However much time I spent on this fruitless task, I would still only know the merest fraction of what was known to everybody I spoke to—and that would be painfully obvious to them! When I changed tack, told the truth upfront that I knew nothing about what they all knew about—but that perhaps I knew something that *they* knew nothing about and thought it might be helpful to them, the reception I received was much more positive. What is needed is the combination, indeed the coalescing, of the two volumes of knowledge. Clearly I did not have the time or ability for that: but they did, with my help.

(Return to Workbook page 122 / Day 7 page 35.)

DAY 9

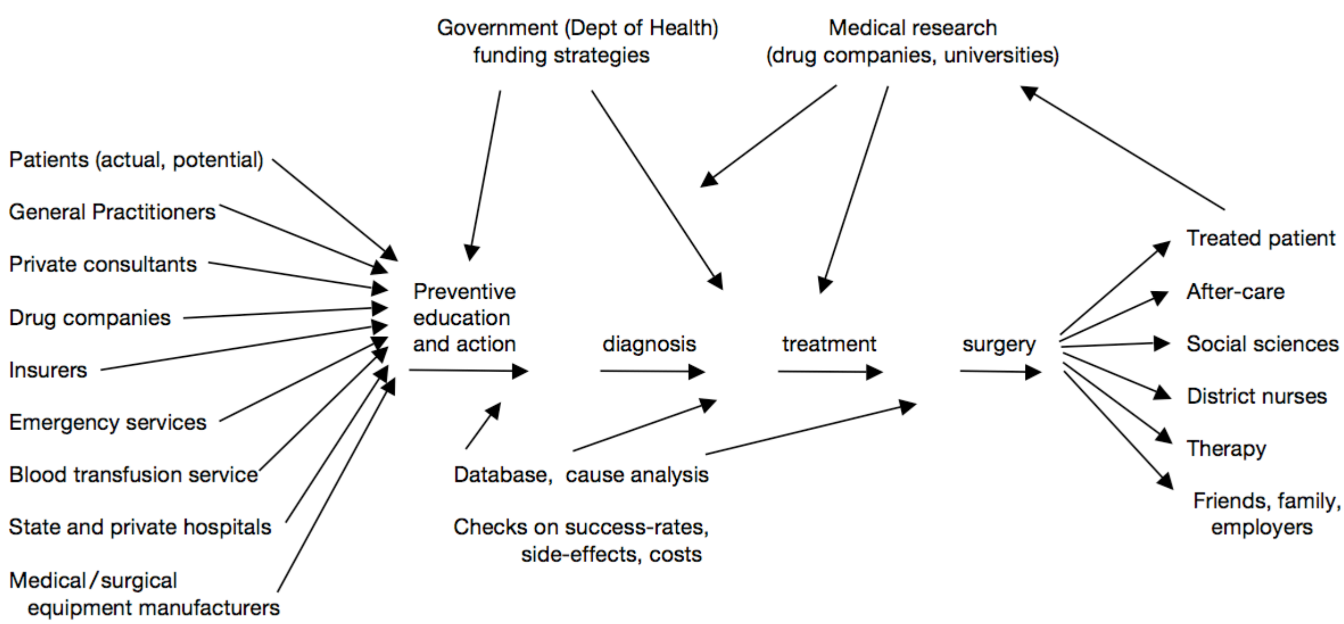
ACTIVITY 9–c

A hint here is first to ask yourself what prevents you (and others) from doing your job as well as you would prefer to. If you cannot do your job as well as you would like then that will presumably cause difficulties to your customers, be they internal or external; and those difficulties will be very likely to result in additional cost or indeed lost customers (which is, of course, also very costly). Frequent candidates as answers to this question come from administration systems and computer systems, with poor or nonexistent training lurking not far behind.

And now that you've identified some matters which increase cost, you can move on to consider related improvements that would therefore reduce cost.

(Return to Workbook page 142 / Day 9 page 7.)

MAJOR ACTIVITY 9–e



(Return to Workbook page 145 / Day 9 page 12.)

DAY 10

SECOND PROJECT

ACTIVITY 10–a

Let's recall Deming's very first words today: he defined a system as: "a network of ... components ... that work together for the aim ... ". I then immediately emphasised "work together" and "aim" as being the most crucial words of this definition. These emphases were further strengthened later on, e.g. around his illustration of a good orchestra being a well-optimised system.

As already observed, there cannot be much that was written earlier under "All One Team" that does not identify directly here with "work together". We have also noted Deming's observation that "the aim proposed here … is for everybody to gain"—i.e. an aim which it is in everybody's interest to achieve, thus strengthening the relevance of "All One Team" to a yet greater extent. Note also that one of the entries under "All One Team" in the detailed version of the Joiner Triangle on Day 4 was "Optimise the system as a whole". And further, that seemingly ambitious aim does not seem likely to be achievable without one of the other constituents of the Joiner Triangle: the "Obsession With Quality". So that is yet a further route for accessing your work in the First Project in order to contribute to your answers here.

And, having got that far, we may as well complete the job! As we know, a large part of the "Scientific Approach" is based on the understanding of Shewhart's founding principles about variation. And we are also very familiar with one of the greatest consequences of his work: that the large majority of problems and opportunities are to be found in the *system*—the subject of this first part of the System of Profound Knowledge.

So the inevitable conclusion is that just about everything said on Days 4 and 5 as regards links between the Joiner Triangle to the 14 Points and Deadly Diseases is relevant here! You will appreciate much of that relevance without difficulty. So, for a little extra learning, think about the items where you *cannot* recognise the connections so immediately, and see if you can then build them.

(Return to Workbook pages 167–170 / Day 10 pages 16–19.)

ACTIVITY 10–b

The style of what I have to say here is much the same as for Activity 10–a. In this case, the eye is most immediately drawn to the "Scientific Approach" corner of the Joiner Triangle. Indeed, virtually everything that I have written in the Scientific Approach sections of the First Project has focused on the understanding of variation and its consequences.

But the links are stronger still. In one of my seminars I often used a slide which simply shows the Joiner Triangle with two words inscribed in the middle: "Reduce variation". For that is intimately connected with the other two corners as well. First, what could be a major part of "Obsession With Quality" if not an obsession for reducing variation? Secondly, it is worth repeating some of what I wrote in the "Interaction of Forces" item on Day 10 page 24 *[WB 175]*. Working together, *pulling* together, as "All One Team" with common aims and purpose must reduce variation. The reverse statement is perhaps even more obvious: *failure* to work as All One Team is bound to *increase* variation. And then there was the reference to a "tug of war". What happens there, with the forces pulling in diametrically opposite directions? First, one side gains a bit of advantage so that everything moves one way; then everything moves the other way; then back again. So the opposing forces actually *produce instability.* Rather obviously, that instability includes something of a zig-zag pattern!

Thus, yet again (as promised!), there may often have been no need for you to add to those matters which you had already raised back on Days 4 and 5—though, of course, if you have done so then I'm not complaining!

(Return to Workbook pages 180–183 / Day 10 pages 29–32.)

DAY 11

SECOND PROJECT
CONTINUED

ACTIVITY 11–a

Links to the Joiner Triangle are less direct in the case of Theory of Knowledge, but nonetheless they are still there. Consider my grouping of the issues into "Prediction", "Theory and Learning" and "Operational Definitions". "Obsession with Quality" is Obsession with *improving* Quality, and that is wholly concerned with *predicting* the consequences of changes that might be made. Next, there is surely a clear major link between "Prediction" and the "Scientific Approach". Indeed, as we have seen, Deming spells that out in Item 3 of Part C when he reminds us that the practical difference between the states of statistical control and non-control is the difference between predictability and non-predictability. The three items in the "Theory and Learning" group are all to do with genuine improvement—so back we go again to "Obsession with Quality". And the use of "Operational Definitions" is undoubtedly an important contributor to reducing variation (the "Scientific Approach" and "Obsession with Quality" yet again) and is *necessary* to help make "All One Team" effective (see Item 7 of Part C in particular).

But, since the links with the Joiner Triangle are not quite so all-embracing here as in the first two parts of the System of Profound Knowledge, I'll now make some references to particular Points and Diseases. An immediate focus is on Point 13: "Institute ... education". Remember the way that I introduced Theory of Knowledge when we first heard of it on the opening day of the course: " … how do we know things, how do we learn things, how do we *improve* that learning and knowledge?" (Day 1 page 39). So we are talking about real education here, not the type of "education" to which (as we saw when discussing Point 13 on Day 5 page 14 *[WB 82]*) Deming ruefully referred near the end of *The Deming of America*:

> "Our education is failing. We just don't educate people, youngsters. We grade them, but don't educate them, don't teach them to think."

As we know, Points 1 and 5 and the first Deadly Disease are concerned with genuine, sustained and sustainable improvement, and that most certainly is what the Deming Cycle is for. Although the Plan–Do–Study–Act cycle is not specifically referred to in Deming's words in Part C, you still do not have to look very hard to find it—e.g. consider Item 6 alongside his sketch which we have already looked at on *DemDim* page 143. Indeed, the purpose of the whole of that second group, "Theory and Learning", is to make improvement genuine and sustainable rather than ill-founded and illusory.

The three items under "Operational Definitions" also have some interesting contributions to make. Item 7 is particularly important with reference to Point 9: "Break down barriers". Barriers between groups, departments, or indeed between companies, which harm customer-supplier relationships, are toughened by the very differences which the use of operational definitions avoids—differences in ways of measuring, of testing, of classifying. Operational definitions help to create a *common language*—and what could be a greater aid to communication? And those initially mysterious Items 8 and 9 ("no true value" and "no such thing as a fact") have interesting reflections in the final Deadly Disease: that of dependence on only visible figures. The more that one thinks about these items from Part C, the more that dependence on only visible figures is seen to be dangerously insecure.

(Return to Workbook pages 197–200 / Day 11 pages 18–21.)

ACTIVITY 11–b

At the very start of the Psychology section, we observed a considerable concentration on system thinking, concerning interactions not only between people and other people but between people and circumstances, etc. So, as a vital constituent of that concentration, again we could head for the Joiner Triangle, particularly the "All One Team" corner.

But instead, let's go more directly to the Points and Deadly Diseases. In Part D you will have seen strong emphasis on matters such as dignity, self-esteem, respect, intrinsic motivation. Here we have some of the most apparent diametrically-opposite differences between much of the "conventional" management (and government) that surrounds us in the 21st Century on the one hand and the Deming approach on the other. Much conventional management and government *destroys* rather than enhances those reflections of what we might call the "inner worth" of people. Is conventional management and government concerned more with intrinsic motivation or with extrinsic motivation? I doubt whether you'll have much trouble answering *that* question. It is in such human matters that Deming's teaching is often particularly striking to the newcomer. However, it would be a mistake to deduce that the Deming approach is "humanitarian" for the sake of being humanitarian: the truth is that those "humanitarian" aspects are *necessary ingredients* for creating that better future which he wanted so much to help us toward.

Nevertheless, it could well be possible to see strong links between just this one aspect of Part D (let alone all the rest of it) and the need to adopt all the 14 Points and eliminate all the Deadly Diseases. E.g., the link with Point 12 ("Pride of Workmanship") is obvious. But remember the "*inter alia*": abolition of performance appraisal, dealt with explicitly in the third Deadly Disease. Everything to do with enabling people to contribute better must also provide strong links, e.g. to Points 1 and 5 and the first Deadly Disease. It is worth noting the emphasis in all three of these on the longer rather than the shorter term: genuine, substantial improvement and indeed innovation need both opportunity and time—which takes us to both the second and the fourth of the Deadly Diseases. Then the matter of invisible versus visible figures strongly emerges: how can you *measure* dignity, self-esteem, respect, intrinsic motivation? The *effects* are clear to see, but how could you *measure* them in any significant way? As Deming says in his discussion of the fifth Deadly Disease on *Out of the Crisis* page 103[121], a manager who ignores these unmeasurable figures "will in time have neither company nor figures". We're also close to Point 8: "Drive out fear". One of the greatest fears in organisations today is that of being measured and judged on inappropriate figures. In fact, those judgment (appraisal) schemes almost invariably push people toward getting figures which will be regarded as acceptable, and often the only way to get them involves the *reduction* or *destruction* of those "inner worth" properties. Which brings us to Point 11 on arbitrary targets ... .

I could go on, but there is probably no need!

(Return to Workbook pages 211–214 / Day 11 pages 32–35.)

OPTIONAL EXTRAS

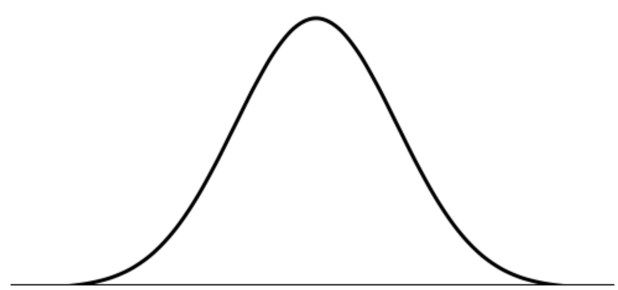
(BY REQUEST OF BALAJI REDDIE ...)

Introduction

This article that Balaji Reddie has been very keen for me to show you (see his "Contributions" page 33) was the seventh in a series of eight short articles that I drafted around 20 years ago. My title for the series was *SPC—Back to the Future*. I did not submit them for publication to any of the main journals but was content to simply distribute them to members of the British Deming Association. My aim was to help a keen but general audience to develop understanding and confident use of control charts; that audience, of course, included many who had no previous background in Statistics. It will not surprise you therefore to know that there is quite a lot of overlap between the content of some of those articles and parts of the "Optional Extras" section in *12 Days to Deming* including, of course, Part D: the "crash-course in conventional Statistics!".

Part E of the Optional Extras: "Is There *Anything* Normal about Control Charts?" is effectively a rewrite of the sixth of those short articles. The article was titled *Two Superstitions*; quite simply, those superstitions, both of which were involved with control charts, just do not relate to the "real world". The seventh of the short articles was titled *More Superstitions.* These further superstitions were related not to control charts but, as Balaji has already told you, to the concept of "six-sigma" quality.

So the original readers, like Balaji, of the seventh article were already familiar with the previous six articles—whereas you, of course, are not. Alternatively, had you reached this Appendix section after already reading Parts D and E of the Optional Extras then again you would have covered the relevant material from those earlier articles. However, since instead you might have arrived here as a result of what Balaji wrote in his "Lessons from History", you may not have that background. In particular, you may know nothing about the conventional statisticians' favourite toy, the normal distribution, with which they can endlessly entertain themselves. In fact (as you might perhaps guess from the title of Part E as stated above), inappropriate use of the normal distribution is actually the prime source of all these various superstitions. If at some time you decide you'd like to learn something about the normal distribution then Part D of the Optional Extras, the "crash-course in conventional Statistics!", is ready and waiting for you; there is also more about it in Part F. But, for now, you'll just have to accept what I tell you about it!



A useful start is a sentence from Day 7 page 26: "If you are uncomfortable with the use of the concept of 'distribution' ... , you can simply think in terms of a histogram having roughly the shape illustrated". And alongside is a typical "shape illustrated" in the case of the normal distribution; for obvious reasons, it is often described as a "bell-shaped curve". As you can imagine even from your brief work on histograms on Day 3, it's not at all unusual for a histogram formed from some process-data to have a shape roughly approximating the shape of a normal distribution. So it is not surprising that conventional statisticians often use the normal distribution in their mathematical derivations with the hope that their results will have some relevance to real data.

But beware! If you read Part E of the Optional Extras you will find me referring on page 60 to the famous British statistician, the late Professor George Box, saying: "All models are wrong—but some are useful". One very clear reason why the normal distribution model is wrong regarding real-world process-data is that absolutely *any* number, never mind *how* large (positive or negative) it is, may appear in the data. Now, at least as far as I'm concerned, I can't imagine any real-world process for which that's true! Referring to my diagram of the normal distribution, the implication is that that bell-shaped curve never *quite* gets right down to the horizontal line (usually called the horizontal *axis*)—although, of course, you could never use a scale large enough to show that on a diagram. The likelihood of numbers occurring far out in the "tails" of the distribution is naturally very tiny, but—in the normal distribution—that likelihood is *never zero*, i.e. enormous values are always *possible*. Why am I telling you all this? Because it is this very matter which lies at the heart of both the first superstition tackled in Part E of the Optional Extras and also that claim of "3.4 parts per million" which so puzzled Balaji regarding "six-sigma" quality.

Nevertheless, the normal distribution has such appealing mathematical properties for the conventional statistician that I need to tell you a little more about it so that you'll be able to see where those impressive claims come from. For they *do* depend on the unrealistic assumption that one is dealing with processes that produce normally-distributed data.

Actually, it's a little misleading to talk of *"the"* normal distribution: there's a whole *family* of normal distributions. In particular, the "bell-shape" can be either narrower or wider than I've illustrated; also, of course, it needs to be centred on the process-average. The width of the bell-shape depends on the distribution's *standard deviation* (you've seen this mentioned several times) which is traditionally denoted by the Greek letter σ (sigma)—and yes, there is some distant connection with the σ with which you're familiar from your work on control charts. The centre of the normal distribution is also traditionally denoted by a Greek letter: that letter is μ (which is pronounced like a cat's "mew"). If you're interested in seeing the content of this paragraph in pictorial form then take a look at page 47 of the Optional Extras.

Finally in this preamble to my revision of the seventh *SPC—Back to the Future* article, "six-sigma" quality is involved with judging quality in terms of conformance to specifications. This is a topic which is first seen early on Day 3 in *12 Days to Deming* but is subsequently studied more fully on Day 7. Also, this current article ends with a brief but very useful mathematical footnote involving another topic which arises on Day 7. However, in this attempt to revise the article so as to become appropriate to a more general readership than originally, I shall not assume you have yet reached Day 7.

"SIX-SIGMA" SUPERSTITIONS

I was delighted when, several years ago, I first heard the phrase "six-sigma quality", made famous through the quality programme bearing that name at Motorola, the American electronics company. I interpreted "six-sigma quality" as implying that, rather than being content to merely meet specifications, the aim was to make the natural variability in a process cover only *half* of the specification range *[see the chart at the top of the next page]*—which I presumed implied the *middle* half. That's not the Deming ideal of "continual improvement", but it seemed a fine step in that direction.

Sadly, my enthusiasm was short-lived. First, I soon came across some advisors who were extolling the virtue of "six-sigma" quality as being the greater freedom to let the process "wander" rather than trying to keep it properly centred. And then I started hearing the remarkable—and unbelievable—claim about "3.4 parts per million" … .

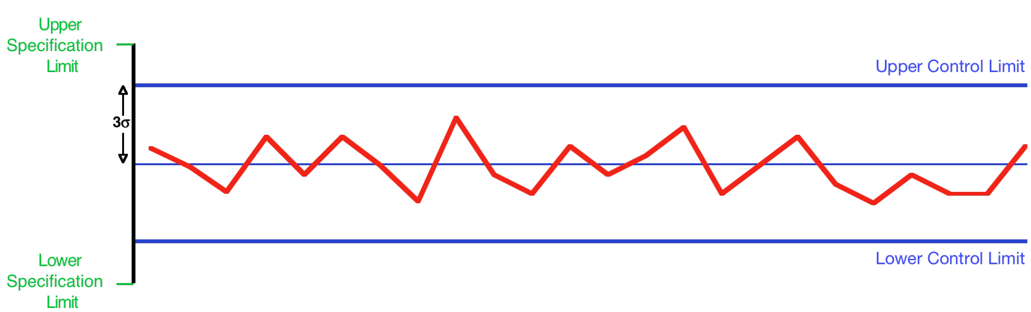
"Six-Sigma" Quality

At the time of first hearing of "six-sigma" quality, one of my frequent battles had been (and continued to be) that of getting people to think seriously of improving quality well beyond what is regarded as officially necessary. The notion that quality is just about satisfying requirements, meeting standards, achieving targets—i.e. conforming to specifications by any variety of names—is in fact a severe obstacle to genuine quality improvement, so much so that it is included in Dr Deming's collection of "Obstacles to the Transformation" which you will examine in Day 7's Major Activity.

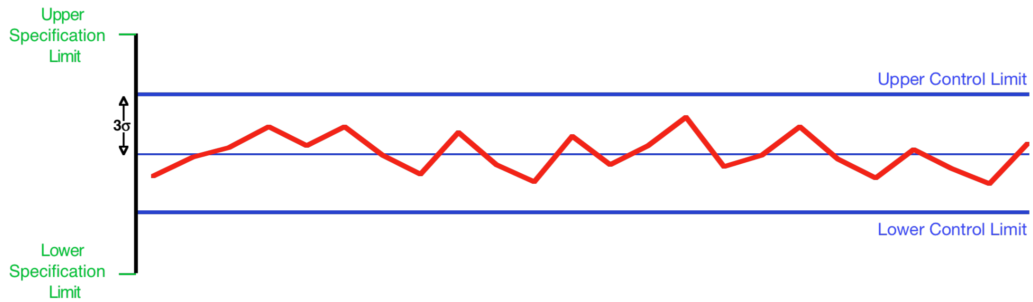
We'll restrict attention for now to situations where specifications are provided, e.g. by the customer (internal or external)—or by the boss! That is, there is a zone outside which the value of whatever we're recording is supposed never to fall. The edges of that zone are the Lower and Upper Specification Limits. The usual situation is that there is an "ideal" or "perfect" value of the measurement which lies in the middle of the zone and the Specification Limits show how far from the ideal the measurement is allowed to stray before being deemed unacceptable—that distance is sometimes referred to as the "tolerance".

Stated concisely, "six-sigma" quality is achieved when that tolerance reaches or exceeds 6σ. But that is likely to be easier said than done! Note that, in theory and when assuming data come from a normal distribution, etc, the value of σ might be presumed "known". In practice it will need to be computed from data in the same or a similar way to that used when constructing a control chart. But, rather than bothering just now about that kind of fine detail, let's see pictorially what sort of situation is implied by "six-sigma".

First, here is a situation which is definitely *not* of "six-sigma" quality:



The process appears to be in statistical control, but the distance from the Central Line to the Specification Limits is only around 4.5σ rather than the desired 6σ. So what can be done? It is unlikely that whoever has set the specifications will be kind enough to relax them, i.e. move them further out, in view of the fact that our process is not good enough to satisfy their "six-sigma" criterion! So it's σ that has to change, i.e. we have to improve the process so that, in due course, we'll get a chart which looks more like this:



Spot the difference! Yes, we have now improved the process, i.e. reduced σ, sufficiently enough for the distance between the Central Line and either Specification Limit to now be about 6σ. (Note that the Specification Limits are where they were before.) Of course, this will not have been achieved overnight!

So I'm all in favour of "six-sigma" quality—or preferably even better than that! Certainly, for those who are still thinking and acting in terms of mere conformance to specifications (which certainly looks as if it was already being achieved in the first scenario at the bottom of the previous page), "six-sigma" quality is a major advance! But my initial excitement was soon tempered, for two reasons:

First, just a few weeks after initially hearing about "six-sigma", I read papers in two journals (which I shall not reference because I don't want *you* to read them!), both of which advertised the big advantage of "six-sigma" quality to be that you no longer have to bother about keeping the process centred. (I'll return to that later.) And second, I discovered people making claims that, with "six-sigma" quality, the proportion of items falling outside the specifications was reduced to that mere *3.4 parts per million*—yes, exactly as Balaji was told. Wow, such precision!

Superstition no. 3: 3.4 parts per million

Whenever I hear a statement like that, I *know* we're back in the mathematical world, no longer in the real world. And so it proved.

In my introduction, I've already told you one of the extraordinary things about the normal distribution: i.e. that, although the curve gives the visual *impression* that it has disappeared from sight by the time we reach the edges of a diagram such as that on page 43, in the mathematical world that curve *never* quite gets down to the horizontal axis. There is little need to tell you that that could hardly be true in our real world. However, as "3.4 parts per million" comes from the mathematical world, we'd better stay there for a little while longer to see *where* in the (mathematical) world the 3.4 parts per million comes from. Then we'll explore a little further ... .

I'll ask you to look one more time at the Optional Extras, this time on page 49. Here we have the same pictures as you saw two pages earlier (page 47) but now with those pictures divided into sections, each being of width σ and with percentages stated of the total area under the curve in each section. Rather similar to the way that the area under a section of a histogram corresponds to the proportion of the data that lie in that section, the percentages printed on page 49 of the Optional Extras indicate the *probabilities* that an observed value from the normal distribution lies in each of the different sections. Notice that those probabilities are exactly the same in each diagram, never mind how narrow or wide the bell-shape happens to be. This is actually an almost unique property of the normal distribution, and is one of the reasons that Mathematical Statisticians are so fond of it.

Note in the diagrams that the area in either tail of the normal distribution beyond 3σ from the mean μ is 0.135%. (This is what led to the existence of the first superstition dealt with in the sixth article and similarly in Part E of the Optional Extras.) Since (in the mathematical world) the normal curve never gets down to the axis, there will *always* be some area in a tail, never mind how far out that tail starts. Since we couldn't hope to indicate any further such detail in pictures, the following table shows a variety of such "tail areas", i.e. probabilities.

Tail Probabilities in the Normal Distribution		
Left-hand tail starts at	**Right-hand tail starts at**	**Area (probability) in tail**
μ – 2.5σ	μ + 2.5σ	0.00621
μ – 3.0σ	μ + 3.0σ	0.00135
μ – 3.5σ	μ + 3.5σ	0.000233
μ – 4.0σ	μ + 4.0σ	0.0000317
μ – 4.5σ	μ + 4.5σ	0.00000340
μ – 5.0σ	μ + 5.0σ	0.000000287
μ – 5.5σ	μ + 5.5σ	0.0000000190
μ – 6.0σ	μ + 6.0σ	0.000000000987

Can you see it? Do you see the "3.4 parts per million"? It's the 0.00000340 for a tail area starting at 4.5σ from μ. *Why there*? That's a very good question.

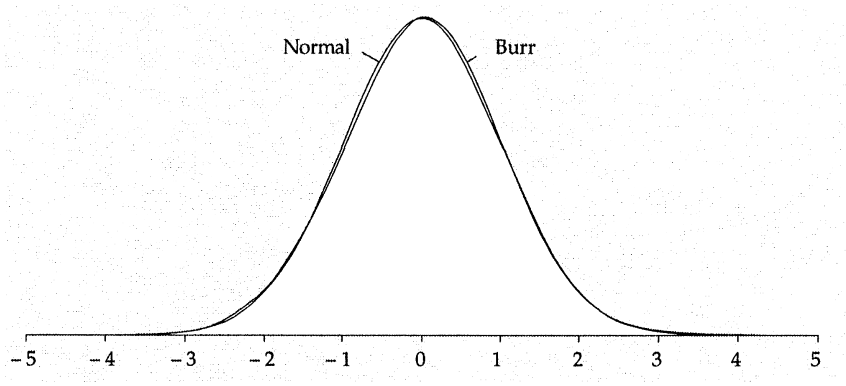
Now, if a process having "six-sigma" quality process is properly centred, the relevant figure from the table is the final one, the 0.000000000987—except that there are two such tails, thus giving a total probability of 0.000000001974. That's 1.974 parts per *billion* outside specifications. Perhaps that's too embarrassing a figure to claim even in the mathematical world!

So instead, the mathematical "six-sigma" enthusiasts let the process mean (and thus the whole normal distribution) wander from its ideal central position. By how much? By 1.5σ. Why 1.5σ? I don't know—ask them. I'm just reporting where the "3.4 parts per million" comes from. *If* the mean wanders that far, it is now 4.5σ from the nearer Specification Limit and 7.5σ from the further one. The chance of reaching the further Specification Limit is obviously *really* negligible (OK: if you insist, it's 0.0000000000000319 which amounts to 31.9 errors in a quadrillion!), and so the *total* probability of falling outside specifications is then, to three significant figures, still the 0.00000340, i.e. the 3.4 parts per million.

The truth of all this (if "truth" can mean anything in the midst of such fantasy) depends, of course, on the process obeying the normal curve—right out into those far tails! How could you ever know? You'd never be able to even *see* it out there! You can compare a histogram of some data with the normal curve to see if it roughly fits in the region where you *can* see it (say, out to somewhere like 2.5σ or, being ambitious, 3σ from the mean), but no further. Does it matter? Does confirming that the process data might fit a normal curve in the region in which you *can* see it imply you may then assume the process follows the normal curve where you *can't* see it?

An Alternative "Truth"

My great friend Don Wheeler (from whom I've have learned so much over the years) told me about a family of probability distributions that was studied by Irving W Burr. Some of them look remarkably like normal distributions (where you can see them). One such is shown below, with a normal distribution drawn in for comparison. Don has kindly allowed me to reproduce this figure from page 196 of his *Advanced Topics in Statistical Process Control*. (He chose there to illustrate the "standardised" versions of the distributions, i.e. those having $\mu = 0$ and $\sigma = 1$, which explains the numbers on his horizontal axis.)



But what happens out in the tails—tails starting at 4.5σ or 6σ away from the mean, for example?

I'll just quote a few figures. First, suppose we have a "six-sigma" process *properly centred*. Recall that the normal distribution tells us that then there are 1.974 parts per billion outside specifications. We can carry out similar calculations with the illustrated Burr distribution (which, remember, *looks* virtually the same). What do we get? Well, actually, 312 parts per billion—over 150 times as many!

Let's try something else. What about if we let the process mean slide out by 1.5σ (where the normal distribution produces the famous "3.4 parts per million")? What do the Burr calculations give? It actually now matters *which way* the mean slides, for that Burr distribution *isn't* symmetric. (Had you spotted that?!) If it slides to the right, we get 22.2 parts per million—about seven times more than with the normal distribution. But, if it slides to the left, we instead get 6 parts per *billion*—about 500 times *less*.

And this is all from two distributions which *look* virtually the same where we can see them and which, in practice, we could *never* tell apart. And we haven't even mentioned here the additional problems caused by the inconvenience of not knowing the "true" values of μ and σ! Such numbers as I have just quoted have no relevance in the real world, be they computed from the normal distribution, the Burr distribution, or anything else. Don refers to the use of any such numbers as "computations winning out over common sense"—and he tells you much more about the Burr distribution in his *Advanced Topics* book if you are really interested!

I'll say something much simpler (for the real world, not the mathematical world):

If you have a "six-sigma" quality process, pretty much properly centred and in statistical control, you'll never get anything outside specifications.

(But that's still not good enough ... .)

Superstitions

Damaging superstitions. There are more. For example, to even *mention* "3.4 parts per million" (even if it were true) is evidence of another very serious superstition: that quality can sensibly be measured in terms of conformance to specifications. A few words later about that. Before then, some reminders of why, in those two articles, I focused on the three superstitions that hopefully have now all been laid to rest.

Regarding the two that are tackled in Part E of the Optional Extras section, there is a message to all of you who have been worrying about normality in connection with control charts: you can *stop* worrying!

Attempting to judge quality as conformance to specifications has one advantage for those doing the judging: it's *easy*! Set a specification (i.e. choose one or two numbers) and then reward or punish depending on whether or not specifications are met. (For "specification" read "requirement" or "target" or "standard" or "tolerance", etc. as desired.)

The Government judges by conformance to specifications. Think of the various Charters, league tables, etc, etc, … . A train meets the specification if it is no more than either five or ten minutes late (according to the kind of journey). A school has met the specification for training (educating?) a child for an examination if the child obtains a Grade C. The National Health Service meets the specification if a patient waits no longer than 18 weeks for an operation. So to wait 127 days is *bad* but 125 days is *good*? Or for a train to be 10 minutes 5 seconds late is *bad* but 9 minutes 55 seconds late is *good*? And there might be only one mark difference between a Grade C and a Grade D.

Tightening the specifications doesn't solve the problem. Perhaps being punctual now means no more than one minute late. So 59 seconds late is *good* but 61 seconds late is *bad*? It makes no sense.

Wise Words from Dr Don Wheeler

During Day 3 you were introduced to the ancient but valuable study by Don Wheeler of "A Japanese Control Chart": it is examined in some detail in Part B of the Optional Extras. There is a yet more detailed account in Chapter 7 of *Understanding Statistical Process Control* by Don Wheeler and David Chambers, published by SPC Press. Long ago, SPC Press also issued a video: *A Japanese Control Chart* of Don describing the chart throughout the period August 1980 to March 1982. I would always show this video in my seminars on Understanding Variation. One of Don's concluding remarks on the video was the following:

> ***"****The total conformance to specifications is no longer enough. Remember that specifications are a compromise created when we could do no better. Now that we can do better, those who are content to live with the compromises of the past will only become increasingly noncompetitive.****"***

A Better Way

"A better description of the world is the Taguchi loss function" said Deming on *Out of the Crisis* page 120 [141]. What's *that* when it's at home?! Basically it's the rather reasonable hypothesis that the best is ideal and that, the further away from the ideal you get, the worse off you are and, in fact, the more rapidly things get worse. This is covered during Day 7, particularly in Activity 7–e (where you'll be guided on how to discover the Taguchi loss function for yourself) and Pause for Thought 7–f (see Day 7 pages 20–23 *[WB 109–112]*). This "better description of the world" was described by Genichi Taguchi at a famous meeting in Tokyo in 1960 at which Dr Deming was present.

So Deming's statement is guidance for us to focus instead on what would be *best*, and work steadily to get closer and closer to it. It's called "continual improvement". "Create constancy of purpose for continual improvement" was the first of Deming's 14 Points (Day 4 page 16 *[WB 56]*). Subsequently he referred to "*lack* of constancy of purpose" as "the crippling disease" (Day 5 page 18 *[WB 86]*).

A (useful) Mathematical Footnote

Actually, if you want to measure "quality", a little mathematics shows that Taguchi's study of "loss" leads to a very neat criterion. (For those who are interested, the "little mathematics" needed is very similar to the "useful trick" used on Optional Extras page 81.) That criterion expresses the Average Taguchi Loss as simply being proportional to

$$\sigma^2 + (\text{non-centredness})^2 .$$

If you have a control chart then you can immediately compute a value for this expression. You will (directly or indirectly) have used a value of σ to find the control limits, and "non-centredness" is the distance between the Central Line and the "ideal". So a good way of regarding improvement is that of keeping the process in statistical control and properly centred while reducing the Average Taguchi Loss.

If you are mathematically inclined, you might like to use the above criterion to verify the following alarming argument. Let's consider starting with the process which gives rise to the chart near the bottom of page 45. It appears to be something like "4.5-sigma" quality, by which I mean that the distance from the Central Line to either Specification Limit is about 4.5σ. We set about improving the process in order to sufficiently reduce σ in order to achieve "six-sigma" quality, i.e. so that the distance from the Central Line to either Specification Limit is 6σ. This brings us to the chart near the top of page 46. Without doubt, to improve the process by that amount would be likely to involve a considerable amount of hard work. I was presuming there that, in both cases, the process was properly centred. However, now imagine letting the process mean wander out to that "3.4 parts per million" point. Would you believe: the Average Taguchi Loss for that *non-centred* "six-sigma" process is now considerably *greater* than it was for the *correctly-centred* "4.5-sigma" process at the bottom of page 45!

Then recall the simple example on Day 3 page 1. I observed that changing the *average* of my process there would be very easy. And, in my experience, it is indeed very often the case that changing a process average *is* a lot easier than reducing its σ. After all (also from near the start of Day 3), Ford's automatic compensation device was doing it after *each* item of data was recorded (I am, of course, not recommending *that*!). So not only are we supposed to believe claims from the "six-sigma" enthusiasts such as that 3.4 parts per million, we are now supposed to believe that a great advantage of "six-sigma" quality is that we are free to let the process mean wander around rather than keeping the process properly centred. What a waste that would be of all the hard work to reduce σ!

I have but a single word for it: absurd.

(Return to Contributions from Balaji Reddie page 33.)

Approvals, Acknowledgments and Information

[a] (page 10) This quotation from the Deming A5 Booklet W2 (formally BDA Booklet W2) has been reproduced with the approval of the Deming Transformation Forum.