What Happens After the Experiment?

Donald J. Wheeler

The following appeared in ***Quality Engineering*** as an invited discussion of *"Must a Process be in Statistical Control Before Conducting Designed Experiments?"* by Søren Bisgaard.

It is both an honor and a pleasure to be asked to discuss Dr. Bisgaard's paper. With his usual clarity he has presented a straightforward explanation of why Randomized Complete Block Designs are used in agricultural and biomedical experiments. By including the historical background he has also demystified a topic that is often shrouded with confusion and obscurity. His paper shows how randomization and blocking use the power of averaging to extract useful information from a collection of experimental units that are clearly not the same. However, this does not quite answer the question posed in the title of his paper. Bisgaard makes the extension from the agricultural case to industrial experimentation by means of an analogy between the experimental units and points in a time series. Unfortunately this analogy overlooks the fundamental differences between agricultural and industrial experimentation. So while Dr. Bisgaard is absolutely correct in what he has written about agricultural experiments, his conclusions about industrial experiments, while technically correct, miss the point because they answer the wrong questions.

For over 400 years the basic confirmatory tool of the scientific method has been the replication of results. Virtually all scientific and engineering knowledge has been gained by virtue of the replication of results. While randomization is central in agricultural and biomedical research, it is virtually unheard of in physics, chemistry, and engineering. In these latter fields experiments are done sequentially, and the replication of results in successive iterations, rather than the alpha-level, is the confirmatory tool. In agriculture, where we typically get only one iteration of an experiment, the replication of results is built into the experiment by the use of multiple experimental units with each treatment. Thus, agricultural experiments tend to be broad and shallow (many experimental units, but one iteration) while industrial experiments tend to be narrow and deep (one experimental unit and multiple iterations). Dr. Bisgaard treats experimental units and iterations as if they are analogous. Unfortunately, in practice, they are not.

In agricultural practice the purpose of the experiment is to demonstrate, in spite of the differences in the experimental units used, that Treatment A will be superior to Treatment B when it is used in other fields and other conditions. The differences between the experimental units are, therefore, a nuisance factor that must be accommodated by the experiment. In fact, as noted by Dr. Bisgaard, it is desirable that the experimental units vary in order that the results will have the generality needed to justify the extrapolation to other fields and other conditions that is required in practice. Moreover, with only one iteration to use for this extrapolation, the emphasis is upon confirmation. You need to be reasonably sure that Treatment A will be better than Treatment B in practice before you will be willing to say so.

However, in industrial practice we are often confronted with a single operating unit and want to know if this unit is operating up to its full potential. Any results we obtain are not going to be generalized to other units (at least not without some additional trial runs). Moreover, if our results are to have any practical utility we have to be able to replicate them in practice. Replication over time will give us our confirmation, which will remove the need to build the confirmation into each experiment. As the emphasis changes from confirmation to discovery, our analysis becomes more exploratory in nature.

With this shift we avoid the problem raised by Dr. Bisgaard when he wrote that "any observational approach to process improvement is fraught with the possibility of unknown biases and confounding." Since we do not have to make a definite statement about the relationship between Treatment A and Treatment B at the end of the first iteration, we can use successive iterations to confirm our ideas or correct any misunderstandings. Thus, with this shift from confirmation to discovery, the whole focus of our studies change. The objective becomes one of learning from the process in order to operate the process up to its full potential. Knowing that Treatment A is better than Treatment B may allow us to operate closer to the full process potential, but it does not completely answer the main question in industrial operations.

To illustrate this last point, consider the data from the last column of Bisgaard's Table 6. Dr. Bisgaard suggests that these values could be used to represent an industrial experiment performed in a Randomized Complete Block Design. Considering the Treatments (A, B, C, D, and E) to be five levels of a single factor (Factor X), the ANOVA in Table 8 in his paper shows that Factor X has a detectable effect upon the response variable (Yield). This ANOVA also shows a detectable difference between the four blocks. As a time-series, this Block Effect is equivalent to a lack of statistical control. Using the expected mean squares for a random effects model, the values in Bisgaard's Table 8 result in the following variance components: for Error, Var(E) = 7869; for Factor X, Var(X) = 8999; and for Blocks, Var(B) = 8725. Adding these up we would estimate that, in the absence of any efforts to control either the Block Effect or the effect of Factor X, we would have process outcomes with a mean near 3286 and a standard deviation of approximately 160. A distribution with these parameter values is shown in Figure 1(a).

If the process is operated with Factor X held at Level E, but the assignable cause that creates the Block Effect is not controlled, we might reasonably expect process outcomes like those shown in Figure 1(b). Here we would estimate the mean to be 3441 and the standard deviation to be 129. The increase in the average Yield and the reduced variation would both represent an improvement over the situation shown in Figure 1(a). Thus our experiment would show how to improve our process.

However, if the process is operated predictably, with Factor X at Level E and with the Assignable Cause controlled to result in Yields like those in Block One, then we might reasonably expect process outcomes like those in Figure 1(c). Here we would estimate the mean to be 3535 and the standard deviation to be 89. The increased average Yield and the reduced variation represent substantial improvements over both 1(a) and 1(b).

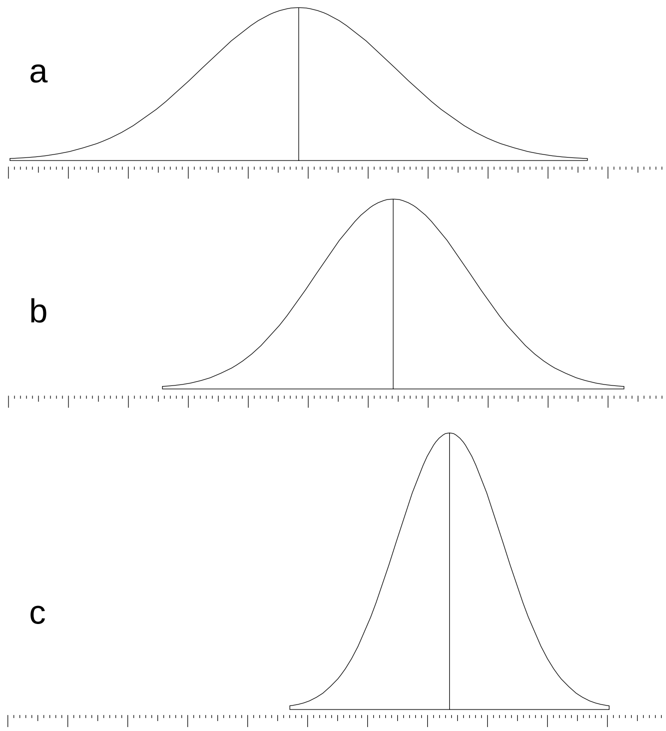

Figure 1 Process Outcomes when
(a) Neither Factor X nor Block Effects are Controlled,
(b) When Factor X Effects are Controlled, and
(c) When Both Factor X and Block Effects are Controlled

Therefore, as Bisgaard correctly notes, in an agricultural experiment, blocking allows us to remove the effects of nuisance components that we cannot do anything about, thereby giving us a more sensitive and confirmatory analysis. However, in industrial experiments, blocking to minimize the impact of an unpredictable process effectively turns an opportunity for learning about the process (discovery) into a nuisance component. While it can be done, it emphasizes the confirmation of one factor effect while missing the opportunity to learn about the effect of another factor.

What Happens After the Experiment?

Assume that we have performed our Randomized Complete Block Experiment and have identified Level E of Factor X as the appropriate level to use in production. This moves our process from Figure 1(a) to Figure 1(b), and is regarded as an improvement. However, without the operational discipline introduced by the use of process behavior charts (control charts), this knowledge is likely to have been gained in vain. Assume for purposes of this illustration that the lower specification for Figure 1 is a Yield of 3300. Figure 1(a) would have about 81 percent nonconforming. Figure 1(b) would have about 14 percent nonconforming. Thus, operating at Level E of Factor X represents a substantial reduction in nonconforming product. However, what is likely to happen whenever some nonconforming items do get made? Without the discipline introduced by a process behavior chart, the presence of nonconforming product is typically the stimulus for changes to be made to the process. Now that the operators know that Factor X has a pronounced impact upon the fraction of nonconforming product, what knob is likely to be adjusted? As the random walk begins, the knowledge gained from the experiment is lost. As one engineer aptly said, "I spend my life working on projects that have an average half-life of two-weeks, implementing solutions that have an average half-life of two-weeks." Or, as a manager once said to this author, "Our biggest problem is vibration [in the finished product]. We have a task force working on the problem of vibration. In fact, we have had to assemble a task force on the problem of vibration every year for the past 30 years." Without the discipline of consistent process operation provided by a process behavior chart, we often keep solving the same problem over and over again.

With the discipline of the feedback provided by a process behavior chart, it is possible to consolidate the knowledge gained from the experiment and also to discover the assignable cause behind the Block Effect, resulting in a process that operates like Figure 1(c). Here the fraction nonconforming would be about 0.4 percent. However, even when a rare nonconforming item was made, the process behavior chart would provide the needed information so that the process inputs would not be changed unnecessarily.

Using Experiments to Bring a Process into Statistical Control

Dr. Bisgaard makes it clear in the first paragraph of his paper that he sees designed experimentation as a "tool for bringing industrial processes into a state of statistical control." Yet his concluding section begins with the statement: "The state of statistical control is a fiction…." Leaving aside this contradiction, Dr. Bisgaard says three paragraphs later that "Design of experiments is the most potent tool available for the quality engineer to get a process in control." Unfortunately, Dr. Bisgaard does not say how to do this. To see how this might play out in practice, consider the problem of producing a product.

At the design phase each characteristic will be assigned a target value. In order to produce products where the characteristics are close to the target values, we have to identify those design and process inputs that have large effects upon the characteristics. To this end we decide to conduct some experiments.

When we conduct an experiment there are only three things we can do with an input variable: (1) we can study that factor by changing the levels of the factor and observing the different responses; (2) we can hold the factor constant during the experiment; or (3) we can ignore the factor altogether. (Blocking is a way of holding a factor locally constant while ignoring the variation between the blocks.) When we hold a factor constant (or block on that factor) we are assuming that it does not interact with the factors that we are studying. When we ignore a factor we are assuming that it has minimal impact upon the response and minimal interactions with the factors we have held constant or studied. When we randomize over one or more of the ignored factors we are simply buying some insurance that if the prior assumptions are not correct, then perhaps the contamination will be averaged out within each treatment.

Given the fact that any one product characteristic will have dozens, if not hundreds, of possible cause-and-effect relationships, we will never have enough time or money to study all of the cause-and-effect relationships. Some factors will be studied, some will be held constant, and others will be ignored. Our experimental results will tell us about the factors we have studied, but if we do not choose the right factors, then we will not get the answers we need. R&D will only discover the relationships shown in black in Figure 2. Moreover, the results shown in black are only valid for those levels of Factors 6 through 10 that were used as constants for the experiment. However, Production has to work in an environment where all 21 factors can, and eventually will, vary.

Based on the results of the experiment in Figure 2, Production would be told to control Factors 5, 1, and 4. The remaining 18 factors would be left uncontrolled. In production, virtually all of the product variation will come from these 18 uncontrolled factors. Moreover, more than 50 percent of this variation will be attributable to the three dominant factors. Such a process fits the profile of an unpredictable process (one that displays a lack of statistical control). On page 151 of his *Economic Control of Quality* Shewhart wrote that "a state of statistical control" exists when "the chance fluctuations in a phenomenon are produced by a constant system of a large number of chance causes in which no cause produces a predominating effect." The dominant nature of Factors 7, 14, and 16 guarantees that the process shown in Figure 2 will be problematic, and unpredictable, until these dominant assignable causes are identified and added to the set of control factors. Therefore, until we have a way to learn from our process, to discover these previously unknown dominant cause-and-effect relationships, we will continue to be at the mercy of Factors 7, 14, and 16.

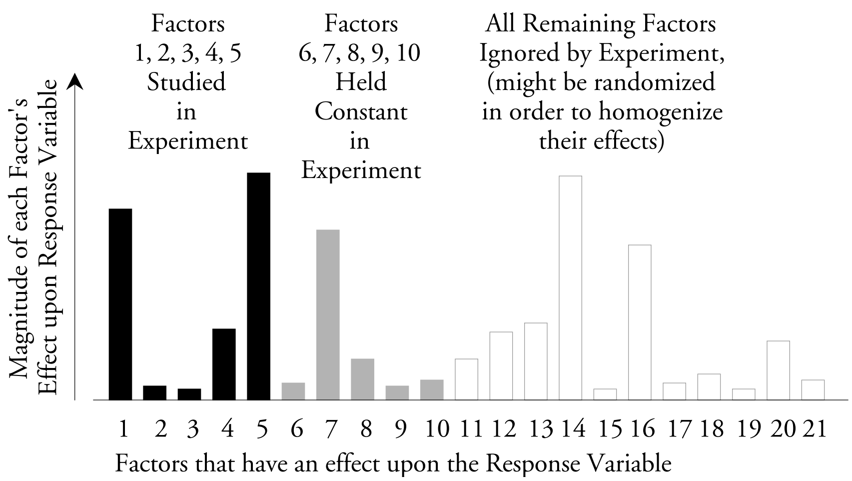

Figure 2: The Amounts of Variation in the Response Variable Due to Each of 21 Factors

Operating a production process in a state of statistical control requires a certain amount of operational discipline. Upsets can occur at any time. Therefore, a state of statistical control is not a condition that can be established once and for all by identifying a few cause-and-effect relationships. This is why it *cannot possibly* be the result of a series of experiments alone. While experiments may be required at the design phase of production, the operational definition of a state of statistical control requires an observational study known as a process behavior chart. If the chart shows evidence of unpredictable behavior, then you need to look for the assignable causes of that behavior. As you find out about assignable causes, and take steps to remove their effects from your process, you will be removing the dominant sources of exceptional variation from your process. As a result, you will be significantly reducing the process variation. Typical reductions reported in practice range from 50 to 80 percent. (A comparison of Figure 1(c) with Figure 1(a) shows a 50% reduction.)

Conclusions

As Dr. Bisgaard has shown, when you are performing an experiment in which each treatment will be applied to a large number of experimental units, and when you will not have the opportunity to conduct subsequent, confirmatory experiments, then randomization is both an insurance policy and a necessary part of good scientific experimentation. When you are doing basic research, where you are not concerned with the identification of *all* the relationships that exist, but merely with confirming the existence of certain selected relationships, this approach makes sense. It is sound and it is proven.

Therefore, Dr. Bisgaard has shown that randomization is useful when: (1) experiments are conducted in circumstances that do not demonstrate statistical control; (2) there are multiple observations per treatment combination; and (3) the analysis is confirmatory (rather than exploratory) in nature. If these three conditions are not present, then randomization loses much of its usefulness.

However, if you are concerned with optimizing a system, or getting a process to work, you cannot depend solely upon experimental results that are always obtained in a limited context. You have to deal with the response variable in the presence of all of the factors that have an impact. You cannot simply study some factors and ignore the others. But, of necessity, every experiment will choose some factors and exclude other factors. So while you may begin with a set of experiments, you need to remember that limited results and conditional relationships do not tell the whole story. Eventually you will need a holistic approach, and this is what observational studies provide. When a factor makes its presence known in an observational study, you can be sure that it is a dominant factor.

So here we have the conundrum. As Dr. Bisgaard correctly shows, you can indeed perform experiments on an unpredictable process, but why would you want to do so? If the process is being operated unpredictably, then there are dominant assignable causes present which you have failed to identify in the past. How is conducting an experiment on the known cause-and-effect relationships going to help? It will generally be more profitable to use an observational study to discover and verify these assignable causes of exceptional variation than it will be to study the known factors you are already controlling.

On the other hand, if the process is being operated predictably, then you will have the process behavior chart to give you feedback about any upsets that may occur from time to time, and in the meantime you will be operating the process reasonably close to its full potential (given the current process configuration). Here you may conduct experiments to optimize the process, but your proven ability to operate the process predictably will remove the need for randomization and blocking.

Finally, if the process is unpredictable and no process behavior chart is being used, then the knowledge gained from the experiment is likely to be lost within two weeks due to the confusion and chaos that typically surround production lines. Without the operational discipline required to operate your process predictably, any knowledge gained is gained in vain.

Statistically designed experiments are essential in R&D. Randomization and blocking are powerful tools for increasing the sensitivity of the analysis in the face of sources of variation that you cannot do anything about. But in the industrial context, where we want to identify assignable causes of exceptional variation in order to remove their effects from our process, and where you have the luxury of successive iterations to confirm your ideas, an observational approach will complement and complete any program of experimentation. Thus, it is not a matter of whether experimental studies or observational studies are the only way to learn. Both are needed. While naturalists can do experiments to study lions at the zoo, there are some things that naturalists can only learn by observing lions in the wild. Observation is the mother of discovery. There is a time and place for experimental studies, and there is a time and place for observational studies. Learn the difference and use both.

This article appeared in ***Quality Engineering, Vol. 20, No. 2***, April-June 2008, pp. 160-164.