Variability Reduction Through Operations Reversal

Hau L. Lee • Christopher S. Tang
Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, California 94305
Graduate School of Management, University of California at Los Angeles, Los Angeles, California 90024

Products with high product variety are often made in a manufacturing process (or a supply chain) consisting of multiple stages, with products taking certain features or "personalities" at each stage. The product may start as a common single engine. As the product moves along manufacturing process, more features are added, and the product assumes more identities of the final end product. When demands of the end products are variable from period to period, the production volumes of the intermediate stages in the manufacturing process are also variable. It is widely recognized that variabilities of production volumes may add cost to the process. This paper is motivated by our observations in industry, where some companies have reengineered the manufacturing process by reversing two consecutive stages of the process. Such changes could lead to variance reduction, thereby improving the performance of the process. We develop formalized models that characterize the impact of such changes: operations reversal. These models are used to derive insights on when such reversal would be advisable.
(*Restructuring; Process Design; Operations Sequencing*)

1. Introduction

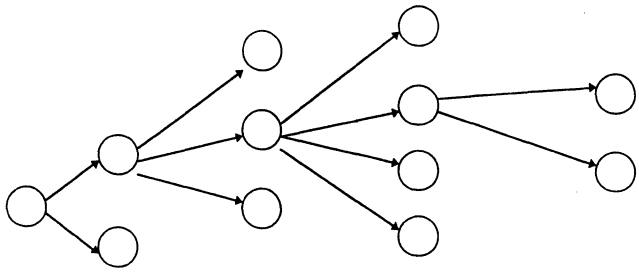
Today's market of industrial and consumer products is characterized by a proliferation of product variety. Product variety, however, can add significant manufacturing costs to the product (see, for example, Child et al. 1991). To compete successfully, companies need to develop manufacturing capabilities that would allow them to be able to offer a high degree of product variety but with low supply chain cost. Recently, we begin seeing concepts such as "mass customization," "agile manufacturing," and "flexible processes" as critical ways for companies to gain control of product variety (see Pine et al. 1993, and Fisher et al. 1993).

Product variety usually follows a product structure that is often arborescent in nature. Fisher et al. (1993) labeled this product structure as the "product hierarchy." In describing such product structure in automobiles, Fisher et al. (1993) started the hierarchy with a platform, which then branched out to multiple models and body styles, which in turn branched to multiple packaged options, followed by stand-alone options. Lee and Billington (1994) described how product fanout can occur for high technology products. Products are often made in a supply chain consisting of multiple stages, with products taking certain features or "personalities" at each stage. These stages can be described as (1) manufacturing, (2) integration, (3) customization, (4) localization, and (5) packaging (see Lee and Billington 1994). The product thus starts as a common single engine. As the product moves along the manufacturing process (or a supply chain), more features are added, and the product assumes more identities of the final end product. Figure 1 depicts the product structure or hierarchy.

Demands of the end products are highly variable from period to period. Consequently, production volumes of the intermediate stages in the manufacturing process are also variable. It is widely recognized that variabilities of production volumes may add cost to the process (Lee and Billington 1993). Such variabilities can be significantly reduced by standardization of the intermediate processes and components, a concept known as "form postponement" in industry (see Lee and Bil-

0025-1909/98/4402/0162$05.00

Figure 1 Product Hierarchy

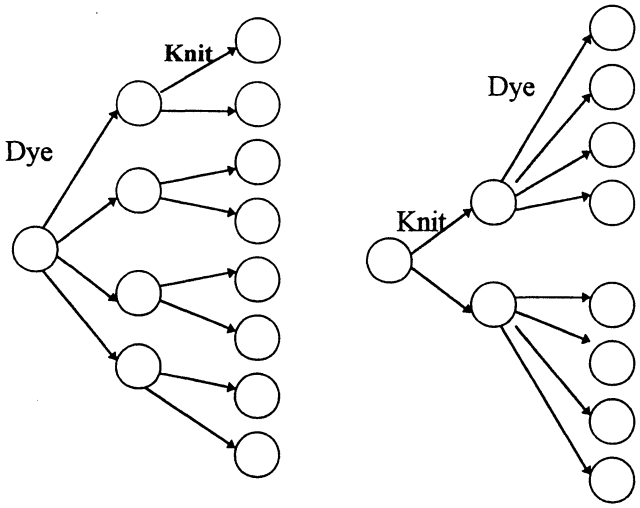


lington 1994, and Lee and Tang 1996). There are, however, potentially other ways in which variabilities can be controlled by reengineering the manufacturing process, leading to better operational performance in the form of lower inventories, higher customer service, and lower operating costs. One such reengineering effort is reversing the sequence of two consecutive stages in a manufacturing process or a supply chain. The classic, and by now well-known, Benetton case (Harvard Business School Case 1986, Dapiran 1992) is an example of such an effort.

As a major apparel manufacturer, Benetton used to manufacture its product by first dyeing yarns into different colors, and then knitting the colored yarns into different finished products (different styles and sizes). Mismatch of inventory of finished garments with different colors had resulted in costly end-of-season markdowns. Luciano Benetton, the chairman of Benetton, was credited with his innovative reengineering of the supply chain by reversing the "dyeing" and "knitting" stages (see Harvard Business School case 1986, and Bruce 1987). Hence, bleached yarns are knitted into the different styles and sizes, and then dyed into the different colored end products when the season's fashion preferences become more established. For this change, the yarns have to be treated in a strong chemical solution to increase their receptiveness to dye. Such a change was considered to be a major breakthrough for Benetton to significantly improve its operational performance. Inventory reduction, better customer service, increasing sales, and fewer write-downs were reported (Dapiran 1992). Figure 2 illustrates such operations reversal innovations at Benetton.

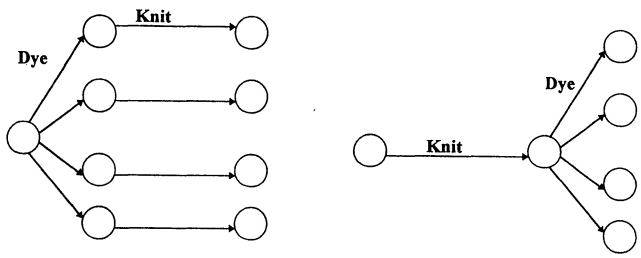
It is probably intuitive as to why Benetton's operations reversal could lead to improvements when there

Figure 2 Operations Reversal at Benetton



is only one single style and size, but multiple colors of the end product. Suppose the total aggregate demand is constant, but there exists uncertainties regarding the color of the end product demanded by the customers. While the demand variabilities of the end products remain unchanged by this reversal, the reversal of the two operations would then effectively reduce the production variability of the first operation, since we are essentially knitting a single product at that stage (instead of dyeing multiple colored yarn at the first stage). In this case, the reversal of the operations delays the point of product differentiation. Figure 3 illustrates this point. This is true even if knitting and dyeing take the same amount of time. Of course, if knitting is a longer operation, then the reversal of the two operations would yield even greater benefits.

Figure 3 Operations Reversal at Benetton: Single Product Style with Four Color Choices



One question that one might ask is: if the aggregate demand of the end products is not constant, but highly variable, would the same result hold? Moreover, when Benetton's product variety is due to features like style, size and color, and when there are multiple options for each of the features, it is also not at all clear why the operations reversal as shown in Figure 2 would give better operational performance. With the advances in automation in the knitting operation, it is also no longer an operation that takes a much longer time to complete, so that the time postponement effect may not be there anymore as well. Why, then, is "knit before dye" better than "dye then knit"? Well, the answer could very well be that it is sometimes better, and sometimes not. As Bruce (1987) reported, Benetton only made 20% of its woolen production using the reengineered process. A better question is then, under what circumstances should "knit before dye" be better than "dye before knit"?

There are other examples similar to the Benetton case where the answer to whether operations reversal should be carried out is not obvious. Consider the manufacturing of the hard drive for a personal computer. One feature of the hard drive could be the memory size and another feature of the hard drive could be the "preloaded" software. In this case, the options with respect to memory size could be 80 MB or 120 MB, and the options with respect to preloaded software could be software for WINDOWS applications or DOS applications. In this example, a hard drive as an end product has four options: 80 MB with WINDOWS applications, 80 MB with DOS applications, 120 MB with WINDOWS applications, and 120 MB with DOS applications. It is possible to insert the hard drive into the motherboard first, and then install the software afterward. However, it is also possible to install the software onto the hard drive by using a "dummy" motherboard before inserting the hard drive into the motherboard. The question here is: which is better? If both installations take negligible times and are inexpensive, then it is probably immaterial as to which should be installed first. However, if extensive testing is required for the installations, and with capacity constraints, then the variabilities of production matter, and the sequencing of the two operations would also be important.

Hewlett-Packard (HP) Company's Deskjet Printer Division also faces a problem similar to the one described above. The variety of the end product is a result of several key features: MAC versus DOS, color versus mono, and the country options. The sequencing of the manufacturing stages in a supply chain in which production differentiation by these features requires some engineering design efforts, but it is not at all clear as to which sequence would give superior operational performance.

Gupta and Krishnan (1995) also described how resequencing the steps in the assembly of a fountain pen can lead to process improvements. In one sequence, the nib is first assembled to the nib-head, followed by the assembly of the inner and outer bodies. In another sequence, the nib-head is first assembled to the inner and outer bodies, before adding the nib. Gupta and Krishnan showed how these two sequences can have very different process flexibility and efficiency in the manufacturing of fountain pens.

This paper is motivated by our observations in industry, where some companies have reengineered the manufacturing process by reversing two consecutive stages of the process. Under what circumstances would such changes lead to variance reduction, thereby improving the performance of the manufacturing process? What are the key drivers? To address these questions, we developed formalized models that characterize the impact of such changes: operations reversal. These models are used to derive insights on when such reversal would be advisable.

We have used the Benetton example to motivate this study. Although we do not have detailed data to perform a complete analysis of the Benetton case, we will be referring to this example merely as a means to illustrate the insights and results. Without a complete analysis, we are not in a position to make strong statements on the exact applicability of the results to the specific case of Benetton.

In the next section, we present the model formulation for the case of two stages, representing two distinct features of the product, and where there are two options for each feature. Section 3 gives results for the special case when the choices for the options of the two features are independent, which help our understanding of the questions posed before. Section 4 describes how the independent choice case can be extended to more general settings, as well as some corresponding results and their

insights. We conclude the paper with summary of results and a description of how the independence assumption of feature option choices can be relaxed.

2. The Basic Model

Consider a manufacturing process with two stages. At each of these stages, a particular feature of the product is defined through an installation or customization process. Hence, there are two features under consideration for these two stages. For each feature, there are two choices. As a result, there are four distinct end products, each being characterized by the choice defined on the two features. Let A and B denote the two features. The question is: should we sequence the supply chain so that A is installed first, followed by B, or vice versa?

Suppose that the production control of the two stages is as follows. The process operates like a *pull* system. Hence, in each time period, the production volumes for the second stage are such that they equal the demands for the end products of the previous period. Similarly, the production volumes for the first stage correspond to the requirements of the previous period. Such a production control rule is actually used by Benetton, which calls such a system "consumer-pull," as opposed to the build-to-forecast alternative of "buyer-pull" (Zakon and Winger 1987). When the stages of the production system operate under order-up-to point type of inventory policies (see Clark and Scarf 1960), then indeed the production volume for every stage would correspond to the demand of the previous period, and so the assumption of pull operating policy is valid. Order-up-to point type of inventory policies are commonly used in practice, and indeed have been shown to be optimal for serial inventory systems under some general assumptions (Clark and Scarf). For this reason, we will concentrate on pull type systems here, and will not consider more complex production planning systems such as production smoothing.

For example, suppose we start with a process where the two stages correspond to feature A being installed before feature B. Let X_{ij} denote the demand of the end product with choice i of feature A, and choice j of feature B, $i, j = 1, 2$. The production volumes for the second stage (B) would correspond to the demands of the four end products in the previous period, i.e., the X_{ij}s. The production volumes of the first stage, which corresponds to the installation of feature A, would then be $X_{i1} + X_{i2}$, for $i = 1, 2$. The production volume for the basic engine, before the installation of features A and B, would then be $X_{11} + X_{12} + X_{21} + X_{22}$.

There are many dimensions in which operations reversal could be evaluated upon. For example, one needs to consider cost (engineering cost to retool, and variable processing cost), quality (e.g., process yield), and speed (e.g., manufacturing lead time) associated with the two alternatives. These basic factors relate to the processing technologies, the complexity of different processes, the labor cost, and other factors. In this paper, we assume that the cost consequence of operations reversal can be evaluated off-line, and focus on the value of variance reduction in operations reversal. Specifically, we focus on the variability of production volumes in the process.

Why are we concerned with the variability of production volumes in the manufacturing process? First, variability drives buffer inventory. If there are buffer inventories, then these buffers are often set as a function of the variability of downstream requirements. Our pull system means that the production volumes are the same as the requirements, and therefore the variabilities of production volumes are the same as the variabilities of requirements. Second, high variability of production volumes are often associated with degradation of quality, process yields, machine downtimes, and other efficiency measures. Third, high variability also makes staff planning more troublesome. Often associated with high variability are overtimes, use of subcontractors, temporary workers, as well as additional costs of other expediting services.

For the two-stage case under consideration, it is clear that the variabilities of the four end product options would not be affected by operations reversal. Similarly, the variability of the production volumes of the input base engine would also be unaffected. Hence, we can concentrate on the variabilities of the intermediate stage, namely, that of the first stage. With two choices, we have the variabilities of two production volumes, one for each choice. Here, we simplify our consideration by defining the sum of the variabilities, i.e., the total variability at the first stage, as our key variability measure.

The total variability as a key performance measure certainly has its limitations. Ideally, one would like to

model the exact consequences of variabilities in the form of inventory, overtimes, expeditions, capacity, and other costs. We use the total variability measure for simplicity. Moreover, it has often been used in the literature as a way to mimic the nonlinear nature of production costs. For example, Johnson and Montgomery's classic text (1974, p. 227, 231) uses the sum of quadratic costs on production level changes and inventory level changes to highlight the nonlinear nature of such cost impacts.

In most cases, the demands for the end product options are *negatively* correlated. To capture these correlations, we assume that the demands for the end product options (X_{11}, X_{12}, X_{21}, X_{22}) are random variables that are multinomially distributed with parameters (N; θ_{11}, θ_{12}, θ_{21}, θ_{22}), where N is the size of the demand, i.e., the number of customers who will buy the product in each period with one of the four options. Assume that N is a random variable with mean μ and standard deviation σ. We also assume that the total demands in different periods are independent.

Here, θ_{ij} represents the probability that the customer will purchase the product that has choice i of feature A and choice j of feature B. Such a demand model would result in:

$$E(X_{ij}|N) = N\theta_{ij}, \quad \mathrm{Var}(X_{ij}|N) = N\theta_{ij}(1-\theta_{ij}), \quad \text{and}$$

$$\mathrm{Cov}(X_{ij}, X_{mn}|N) = -N\theta_{ij}\theta_{mn}, \quad \text{for } ij \neq mn. \qquad (1)$$

Thus, the negative correlations among the options are captured.

Instead of trying to estimate θ_{ij} directly, we use the following intermediate parameters. Let:

p = probability that a customer will purchase a product with choice 1 of feature A, given that he/she will purchase a product. Hence, $1 - p$ is the conditional probability of a customer buying the product with choice 2 of feature A.

$f(p)$ = Prob($B1|A1$) = conditional probability that the customer buys the product with choice 1 of feature B, given that the customer has decided to purchase the product with choice 1 of feature A.

$g(p)$ = Prob($B1|A2$) = conditional probability that the customer buys the product with choice 1 of feature B, given that the customer has decided to purchase the product with choice 2 of feature A.

Note that the conditional probabilities $f(p)$ and $g(p)$ enable us to model the interactions between product features and between choices. The probability of a customer purchasing product with different options is depicted in the following diagram (Figure 4).

Given such a probability structure, we can derive:

$$\theta_{11} = pf(p), \quad \theta_{12} = p[1 - f(p)], \quad \theta_{21} = (1-p)g(p),$$

$$\theta_{22} = (1-p)[1 - g(p)].$$

We will use the well-known probability relationship that, for two random variables X and N, then:

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|N)] + \mathrm{Var}[E(X|N)].$$

We refer to $E[\mathrm{Var}(X|N)]$ as the expectation of the conditional variance, and $\mathrm{Var}[E(X|N)]$ as the variance of the conditional expectation.

Denote A-B as the sequence of a process whereby the feature A is installed before feature B, and B-A as one with feature B installed before feature A. As discussed before, we focus on the variances of the intermediate stage. It can be easily verified that, for A-B, the total conditional variance is $2Np(1-p)$, so that the expectation of the conditional variance is $2\mu p(1-p)$. The sum of the variances of the conditional expectation is $\sigma^2[p^2 + (1-p)^2]$. The total variance measure is thus

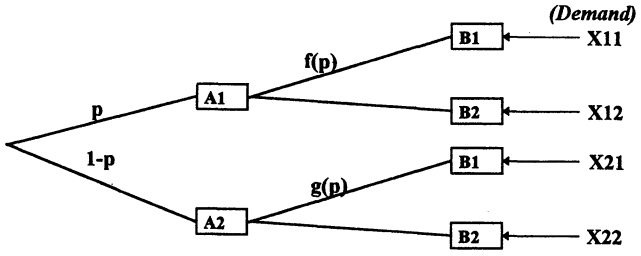
$$2\mu p(1-p) + \sigma^2[p^2 + (1-p)^2]$$
$$= 2(\mu - \sigma^2)p(1-p) + \sigma^2.$$

For B-A, the expectation of the total conditional variance is

$$2\mu[pf(p) + (1-p)g(p)]\{1 - [pf(p) + (1-p)g(p)]\}.$$

The sum of the variances of the conditional expectation is

Figure 4 A Two-Stage Model: 2 Features × 2 Choices System



$$\sigma^2\{[pf(p) + (1-p)g(p)]^2 + \{1 - [pf(p) + (1-p)g(p)]\}^2\}.$$

The total variance measure is now:

$$2(\mu - \sigma^2)[pf(p) + (1-p)g(p)] \cdot \{1 - [pf(p) + (1-p)g(p)]\} + \sigma^2.$$

Taking the difference, we observe that the *sequence A-B* has a smaller total variance than *B-A* if $C(p) < 0$, where:

$$C(p) = (\mu - \sigma^2)\{p(1-p) - [pf(p) + (1-p)g(p)] \cdot \{1 - [pf(p) + (1-p)g(p)]\}\}. \quad (2)$$

Note that, when $\mu = \sigma^2$, the sequencing of the features is immaterial. This would be the case of Poisson demands.

3. Independent Feature Choices

In §§3 and 4, we consider the case in which the choice selection of one feature is independent of that of the other feature, i.e., $f(p) = g(p) = q$, where q is the probability of a customer selecting choice 1 of feature B and is a constant that is independent of p. We will come back to the general case in §5.

3.1. Two Choices, Two Features

Here, $C(p) = (\mu - \sigma^2)[p(1-p) - q(1-q)]$. Hence, for stable total demand so that $\mu > \sigma^2$, then condition (2) is reduced to $p(1-p) < q(1-q)$. Observe that, for $0 \leq x \leq 1$, the function $x(1-x)$ is first increasing in x, attains its maximum at $x = 0.5$, and is decreasing in x thereafter. Hence, when $\mu > \sigma^2$, sequence *A-B* has a lower variance as long as the choice probabilities associated with feature A (i.e., p) are more "distinctive" than those of feature B (i.e., q). This implies that one can lower the total variance of the system if the feature with more "distinctive" or "imbalanced" choice probabilities is processed first. There is more "risk-pooling" by sequencing stages this way. Note that this result applies to the case with two distinct choices for each feature, and the more general case of multiple choices will be discussed in §5.

What kind of inference can we draw from this result? Let us use the garment example again. Suppose that the total demand of a certain type of garment is fairly stable, but the product mix in terms of style and color is variable. If the choice probabilities of style is more distinct than those of color, then our result says that by processing the style feature before the color feature, then the total variance would be reduced. This implies that the knitting operation should be performed before the dyeing operation.

The reverse, however, is true if $\mu < \sigma^2$. This is a surprising result. It states that the nature of total demand uncertainties is very critical in determining whether operations reversal is an effective means to reengineer the supply chain. When the total demand is highly variable, then by installing the feature with more imbalanced choice probabilities first, the production volume fluctuations could be magnified, leading to such a sequence being undesirable this time. In the Benetton example, it seems like the total demand for apparel items like sweaters is fairly stable, but it is the color, style, and size option mixes that are highly variable. On the other hand, there are other cases, like high technology products such as computer workstations, where the basic demand itself is highly variable. To copycat the Benetton-like operations reversal without understanding the nature of total demand variabilities could thus be dangerous.

Amazingly, it turns out that this condition $\mu > \sigma^2$ on the nature of total demand variability is a very important one, as will be shown below. To gain some more understanding as to why such a condition would lead to such contrary result, we consider the following special example. Suppose we have one style and two colors: red and blue. Let the probability of red being chosen be p. If we knit before we dye, then the total variance after the knitting operation is simply σ^2. Now, if we dye before we knit, then the variance of red yarns after dyeing is $\mu p(1-p) + \sigma^2 p^2$, whereas that of blue yarns is $\mu p(1-p) + \sigma^2(1-p)^2$. The covariance of blue and red yarns is given by [Var(red and blue yarns) − Var(red yarns) − Var(blue yarns)]/2 = $(\sigma^2 - \mu)p(1-p)$. If $\mu > \sigma^2$, then the covariance is negative, and as expected, the variance of the aggregate demand is smaller than the sum of the individuals. On the other hand, if the reverse is true, then the covariance is positive, leading to the variance of the aggregate demand being greater than the sum of the individuals!

3.2. Multiple Choices, Two Features

The above analysis can be extended to the case when the number of choices for each feature/stage exceeds two, as long as the number of stages in the process (and the

corresponding number of features of the product) remains at two. Let p_i be the probability that choice i of feature A would be chosen by the customer, and let q_j be the probability that choice j of feature B would be chosen by the customer, where $i = 1, \ldots, a$; and $j = 1, \ldots, b$.

We can easily show that total variability measure for sequence A-B is given by:

$$\mu \sum_{i=1}^{a} p_i(1 - p_i) + \sigma^2 \sum_{i=1}^{a} p_i^2 = \mu - (\mu - \sigma^2) \sum_{i=1}^{a} p_i^2.$$

The corresponding result for the sequence B-A can be easily derived. For sequence A-B to have lower total variance, we need:

$$(\mu - \sigma^2)\left(\sum_{i=1}^{a} p_i^2 - \sum_{j=1}^{b} q_j^2\right) > 0. \tag{3}$$

If $\mu > \sigma^2$, then essentially the same conclusion as before holds, i.e., sequence A-B has lower total variability when $\Sigma_{i=1}^{a} p_i^2 > \Sigma_{j=1}^{b} q_j^2$. The reverse is true otherwise.

Note that the p_is and the q_js sum to one, and that $\Sigma_{i=1}^{a} p_i^2$ attains its minimum when all the p_is are equal. If $a = b$, then we can obtain a qualitative interpretation of the result here. When we have stable total demand, then the feature with less "balanced" probabilities of the choices should be sequenced first to reduce system variabilities. When we have highly variable total demand, then the reverse is true. This result is certainly in line with that of §3.1.

Suppose further that the choice probabilities of the choices for each feature are equal. Hence, the choice probability of a particular choice of feature A is $1/a$, and the corresponding probability for B is $1/b$.

Using the same definition of θ_{ij} and X_{ij} as before, note that $\theta_{ij} = 1/ab$ for all $i = 1, \ldots, a$ and $j = 1, \ldots, b$. Using (1), we get: $E(X_{ij}|N) = N/ab$, $\mathrm{Var}(X_{ij}|N) = N(1/ab)(1 - 1/ab)$, and $\mathrm{Cov}(X_{ij}, X_{mn}|N) = -N(1/ab)^2$ for $ij \neq mn$.

Under the sequence A-B, the conditional production volume variance for each choice of A is given by $N(1/a)(1 - 1/a)$. This result can be obtained from standard probability theory, or can be obtained as:

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{j=1}^{b} X_{ij}|N\right) \\
&= b\,\mathrm{Var}(X_{ij}) + b(b-1)\,\mathrm{Cov}(X_{ij}, X_{in}) \\
&= bN(1/ab)(1 - 1/ab) + b(b-1)[-N(1/ab)^2] \\
&= N(1/a)(1 - 1/a).
\end{aligned}$$

Hence, the expectation of the total conditional variance for the first stage production is $a\mu(1/a)(1 - 1/a) = \mu(1 - 1/a)$. For each choice of feature A, the expected production volume, for a given N, is N/a. Hence, the variance of this expectation is σ^2/a^2. Since there are a such choices, the total is σ^2/a. The total variance measure for the first stage is now: $\mu(1 - 1/a) + \sigma^2/a = \mu - (\mu - \sigma^2)/a$. The corresponding variability measure for the sequence B-A is $\mu - (\mu - \sigma^2)/b$. If $\mu > \sigma^2$, i.e., stable total demand, then the sequence A-B has lower total variance if $a < b$.

Let us revisit our garment case. For illustrative purposes, suppose that the total demand is fairly stable, and suppose the features of style and color have choices that are roughly equal in attractiveness to customers. If the number of style choices is smaller than the number of color choices, then knitting before dyeing would give lower variabilities to the supply chain. On the other hand, if the total demand is highly variable, then the "knitting before dyeing" strategy would not be effective.

3.3. Two Choices, Multiple Features

For this case, let p_k be the probability of a customer choosing choice 1 of feature k, where $k = 1, \ldots, m$ and that m is the number of the stages / features. By using a simple interchange argument, it can be easily shown that, if $\mu > \sigma^2$, then the total variance of the production volume at the intermediate stages is minimized by sequencing the stages / features in ascending order of $p_i(1 - p_i)$. This is equivalent to sequencing the stages / features in descending order of $|0.5 - p_i|$. We omit the details for this proof.

3.4. Multiple Choices, Multiple Features

The result of §3.3 can be easily generalized to the case of multiple choices and multiple features. In this case, suppose p_{ij} denotes the probability that choice j will be chosen for feature i, then, when $\mu > \sigma^2$, the total variance of the production volumes at all the intermediate stages is minimized by sequencing the stages / features in descending order of $\Sigma_j p_{ij}^2$, where the summation is over j, the possible choices of feature i. The proof is straightforward, and is omitted here.

4. Incorporating Leadtimes

So far, we have considered the variabilities of production volumes of each period. Implicitly, we have as-

sumed that the production in each period is completed within the same period. When this is not the case, and in particular, when the leadtimes for the two stages under consideration are not equal, then we need to expand our definition of variabilities. This is especially important when we are considering operations reversal in a supply chain, where the two stages can be located at different sites and therefore the leadtimes would include the transportation times into the sites, which could often be very different.

Consider a stage. In each period, there is of course the production initiated in this period. When production leadtime is greater than one, than we also have volumes that are work in process. Suppose that demands over time are independent. Then, since we initiated a production volume in each period equal to the demand of the previous period, the total work in process should be equal to the total demand of the past periods that is of exactly the duration of the leadtime. Hence, the variance of the total work in process is the variance of demand in leadtime. With unequal leadtimes for the two stages, we will thus concentrate on this measure of variability in our analysis in this section. Such a measure is also useful if we are concerned with having to use inventories at the end of the stages to buffer against uncertainties. It is well known that the key driver to buffer inventories is leadtime demand variabilities.

This time, we cannot just focus on the variabilities of the first stage. The variances at both stages have to be accounted for. Moreover, it is likely that the impacts of variability at the two stages of the supply chain are different. In other words, variabilities at one stage may be more harmful than the other. For example, if safety stocks at the end of the two stages are of key management concern, and since inventory value for stocks at the end of the second stage would presumably be greater than that of the first stage, then variability, a driver of safety stock, at the second stage may be of greater significance than that of the first one. In this section, we consider the case when the measure of total variability is obtained by applying some stage-specific weights to the variabilities of two stages. Let w_{ij} be the weight associated with the variability of stage i, $i = 1, 2$, if feature j is the first feature in the sequence, $j = A, B$. For example, w_{2A} corresponds to the weight applied to the variability observed at the second stage for the sequence A-B. In this case, the second stage corresponds to the stage where feature B is installed (and that feature A would have already been installed prior to the second stage.)

Let T_A and T_B be the respective leadtimes (in periods) for the two stages for feature A and B.

Consider the sequence A-B first. We refer to the derivation in §3.1, and the total conditional variance for the first stage is given by $2Np(1-p)T_A$. The variance of the conditional expectation for the first stage is given by $\sigma^2[p^2 + (1-p)^2]T_A$. Using (1), the total conditional variability for the second stage, $\Sigma_{i,j}\,\mathrm{Var}(X_{ij}|N)$, can be simplified to

$$N\{1 - [p^2 + (1-p)^2][q^2 + (1-q)^2]\}T_B.$$

The corresponding variance of the conditional expectation at the second stage is

$$\sigma^2[p^2 + (1-p)^2][q^2 + (1-q)^2]T_B.$$

Hence, combining these results, we see that the total variability measure for sequence A-B is given by:

$$\begin{aligned}&(\mu - \sigma^2)\{2p(1-p)T_A w_{1A}\\ &\quad + \{1 - [p^2 + (1-p)^2][q^2 + (1-q)^2]T_B w_{2A}\}\}\\ &\quad + \sigma^2(T_A w_{1A} + T_B w_{2A}).\end{aligned} \tag{4}$$

Similarly, for the sequence B-A, the corresponding measure is:

$$\begin{aligned}&(\mu - \sigma^2)\{2q(1-q)T_B w_{1B}\\ &\quad + \{1 - [p^2 + (1-p)^2][q^2 + (1-q)^2]T_A w_{2B}\}\}\\ &\quad + \sigma^2(T_B w_{1B} + T_A w_{2B}).\end{aligned} \tag{5}$$

Consequently, sequence A-B would have a lower total variance if (5) is greater than (4). To isolate the impact of leadtime and the relative weights applied to the two stages, we make the assumption that $p = q$ so as to gain more insights.

Suppose $w_{ij} = w$, i.e., equal weights are applied to the variabilities at the first and second stages. Then the condition for (5) to be greater than (4) becomes:

$$\begin{aligned}&(\mu - \sigma^2)(T_A - T_B)\\ &\qquad \{1 - [p^2 + (1-p)^2]^2 - 2p(1-p)\} > 0.\end{aligned} \tag{6}$$

In the appendix, we show that the expression

$1 - [p^2 + (1-p)^2]^2 - 2p(1-p)$ is always nonnegative, for $p \in [0, 1]$. Consequently, in a stable demand situation, i.e., $\mu > \sigma^2$, condition (6) is always satisfied when $T_A > T_B$. This is an intuitive result, i.e., when the choice probabilities of the two features are similar and when the total demand is fairly stable, then sequencing the stage with the longer leadtime first would result in lower total variability.

Back to our garment example. If style and color have similar choice probabilities from customers, then performing the knitting operation first is desirable, as knitting takes relatively longer time than dyeing.

When $\mu < \sigma^2$, then interestingly, the reverse is true.

Earlier in this section, we argued that a motivation for having differential weights for the two stages of the supply chain could be that the value of buffer inventories used to hedge against the variabilities at these two stages are different. Because of this reason, one way to define these differential weights is to apply the inventory values at the end of the stages. Let v_A and v_B be the value added of feature A and B, respectively, and v be the inventory value of the item entering the two stages under consideration. Then, under the sequence A-B, the inventory value at the end of the first stage after feature A is added, is $v + v_A$. The inventory value at the end of the second stage, after feature B is added, is $v + v_A + v_B$. The corresponding values for sequence B-A are $v + v_B$ and $v + v_A + v_B$ respectively. Hence, we can define: $w_{1A} = v + v_A$; $w_{1B} = v + v_B$, and $w_{2A} = w_{2B} = v + v_A + v_B$.

Again, consider $p = q$. Taking the difference of (5) and (6) gives the following condition for sequence A-B to have a lower total variability measure:

$$(\mu - \sigma^2)(T_A - T_B)(v + v_A + v_B) \cdot \{1 - [p^2 + (1-p)^2]^2 - 2p(1-p)\} + [2(\mu - \sigma^2)p(1-p) + \sigma^2](T_A v_B - T_B v_A) > 0. \quad (7)$$

Note that $2(\mu - \sigma^2)p(1-p) + \sigma^2$ is always positive. Hence, if $T_A = T_B$, then (7) holds as long as $\mu - \sigma^2 > 0$ and $v_A < v_B$. This is probably a well-known result: sequence the stage with the smaller value-added first. Similarly, if $v_A = v_B$, then (7) also holds as long as $T_A > T_B$.

In addition, we have a simple sufficient condition for (7) to hold. Suppose we have stable demand, i.e., $\mu > \sigma^2$. The first term on the left-hand side of (7) has been established to be positive if $T_A > T_B$. Now, if we also have $v_B / T_B > v_A / T_A$, then (7) would hold. Hence, if stage A takes longer time to process, and if the value-added per leadtime for stage A is smaller than that of stage B, then stage A should be performed before stage B. Thus, an operation that takes longer and one that builds up value of the product at a smaller pace is the one that should be sequenced first. In a special case, if the value-added of a stage is proportional to the leadtime, then we have $v_B / T_B = v_A / T_A$, and (7) is always satisfied if $T_A > T_B$.

5. Extensions

Sections 3 and 4 describe insights from analysis of the case when the choice probabilities of the two features are independent, i.e., $f(p) = g(p) = q$. More general cases can be analyzed, although at the expense of increased complexity. Several such examples are illustrated.

(i) *Choice Interaction.* To capture some interactions of the choice probabilities of feature B as a function of the choice of A, we can define $f(p) = q$, $g(p) = rq$, where $r \in [0, 1/q]$. The parameter r represents the impact of the choice selection of feature A on the choice selection of feature B. To see that, suppose $r > 1$. Then the customer would have a higher probability to choose choice 1 of feature B had he selected choice 2 of feature A.

A special case here is when $r = 0$. Here, $f(p) = q$ and $g(p) = 0$. This situation occurs when there is a technological constraint on the choice selection. For example, suppose feature A corresponds to the size of RAM in a PC and feature B corresponds to the "preloaded" software in the hard disk. These two features can be added to the CPU in any sequence. Feature A has 2 choices: 2 MB or 1 MB, and feature B has 2 choices: WINDOWS 3.1 applications or DOS applications. Since WINDOWS 3.1 requires at least 2 MB of RAM, the customer is limited to only DOS applications if 1 MB RAM is selected. This limitation is called "bundling," where the 1 MB RAM and the DOS applications have to be purchased as a "bundle." In this example, $r = 0$.

When $r = 0$, the choice probabilities associated with feature B are perfectly distinctive for choice 2 of feature A. Under this situation, we can show from (2) that $C(p) > 0$ so that when the total demand is stable, it is more

effective to sequence the stage corresponding to feature B first in the process.

(ii) *Choice Reversal*. Consider the case where $f(p) = q$ and $g(p) = 1 - q$. Thus, the customer's choice selection of feature B is completely reversed, depending on the choice selection of feature A. In this case, the reversal nature of the choice probabilities associated with feature B causes the choice probabilities associated with feature B less distinctive. This implies that, with stable total demand, it is more beneficial to sequence feature A first.

(iii) *Feature Interactions*. Suppose the features are dependent, i.e., both $f(p)$ and $g(p)$ are dependent on p. A simple way to model this dependence is to let $f(p) = g(p) = sp$, where $s \in (0, 1/p]$. The parameter s represents the impact of the choice selection of feature A on the choice selection of feature B. If $s > 1$, then a customer would have a higher probability to choose choice 1 of feature B had he selected choice 1 of feature A. When s is close to zero, we have a situation that is similar to the "bundling" case.

(iv) *Joint Interactions of Feature and Choice*. Suppose $f(p)$ and $g(p)$ depend on p, and the choice selection of one feature depends on the choice selection of the other feature. A simple way to model these dependencies is to let $f(p) = rp$ and $g(p) = sp$ where $r, s \in (0, 1/p]$. The parameters r and s represent the impact of the choice selection of feature A on the choice selection of feature B.

6. Conclusion

In summary, operations reversal has been found to be a powerful means to reengineer a manufacturing process or a supply chain. In this paper, we develop simple models to gain insights as to when would operations reversal be desirable. Using total variability as a measure, we note that there are several situations for which operations reversal would be desirable. When the major source of demand uncertainty lies in the option mix and the total demand for all options is fairly stable, then the resulting sequence should essentially have one or more of the following properties:

(1) the resulting initial stage has more distinct choice probabilities for the options of the feature installed by that stage;

(2) the choice probabilities are roughly equal for both features installed by the two stages, but the number of choices for the feature in the first stage is smaller than the subsequent one;

(3) the first stage has longer leadtime than the subsequent one, if the value-added of the two stages / features are roughly the same;

(4) the value-added of the first stage is smaller than the subsequent one, if the leadtimes for the two stages are the same; and

(5) the first stage takes longer to perform, but the value-added per unit leadtime for the first stage is smaller than or equal to the corresponding value-added for the second stage.

When the total demand itself is highly variable, then interestingly, the reverse of the above observations is true, i.e., the properties of the sequence of the two stages indicated above should be reversed.

A key limitation of the current paper is our use of the total variability as a performance measure. Future studies can include the explicit modeling of the cost consequences of variability directly. For example, if we are concerned with safety stocks, then it may be the standard deviation instead of the variance that matters. In that case, the performance measure could be the sum of (weighted) standard deviations. It would be of interest to see how, or under what conditions, the current results would carry over in that case.

Design for supply chain management has been viewed as a major design principle that could lead to great savings in logistics costs and improvement in customer service (see Lee 1996). This paper shows that the reengineering efforts require careful planning, and understanding the nature of demand variabilities, the option choice probabilities, and the other characteristics of the manufacturing system or the supply chain are necessary. Our analyses have shed some light on when such efforts would be worthwhile.

Appendix

PROOF OF $1 - [p^2 + (1 - p)^2]^2 - 2p(1 - p) > 0$. Let $\phi(p) = 1 - [p^2 + (1 - p)^2]^2 - 2p(1 - p)$. It is clear that $\phi(0) = \phi(1) = 0$. Moreover, $\phi'(p) = 2(1 - 2p)^3$. Hence, $\phi(p)$ is a function that is increasing in p for $p < 0.5$, and decreasing in $p \in (0.5, 1)$. This shows that $\phi(p) > 0$ for $p \in (0, 1)$. □

References

Bruce, L., "The Bright New Worlds of Benetton," *International Management*, November (1987), 24–35.

Child, P., R. Diederichs, F-H. Sanders, and S. Wisniowski, "The Management of Complexity," *Sloan Management Rev.* (1991), 73–79.

Clark, A. and H. Scarf, "Optimal Policies for a Multi-Echelon Inventory Problem," *Management Sci.*, 6 (1960), 474–490.

Dapiran, P., "Benetton—Global Logistics in Action," *Asian Pacific International J. Business Logistics*, 5 (1992), 7–11.

Gupta, S. and V. Krishnan, "Product Family-Based Assembly Sequence Design to Advance the Responsiveness-Customization Frontier," Working Paper, University of Texas at Austin, Austin, TX, 1995.

Fisher, M., A. Jain, and J. P. MacDuffie, "Strategies for Product Variety: Lessons From the Auto Industry," Working Paper, The Wharton School, University of Pennsylvania, Philadelphia, PA, 1993.

Johnson, L. A. and D. C. Montgomery, *Operations Research in Production Planning, Scheduling, and Inventory Control*, John Wiley & Sons, New York, 1974.

Harvard Business School, "Benetton (A) and (B)," *Harvard Teaching Case* 9-685-014, Cambridge, MA, 1986.

Lee, H. L., "Design for Supply Chain Management: Concepts and Examples," in R. Sarin (Ed.), *Perspectives in Operations Management*, Kluwer Academic Publishers, Boston, MA, 1993, 43–65.

—— and C. Billington, "Material Management in Decentralized Supply Chains," *Operations Res.*, 41 (1993), 835–847.

——, "Effective Inventory and Service Management Through Product and Process Redesign," *Operations Res.*, 44 (1996), 151–159.

—— and C. Billington, "Designing Products and Processes for Postponement," in S. Dasu and C. Eastman (Eds.), *Management of Design: Engineering and Management Perspectives*, Kluwer Academic Publishers, Boston, MA, 1994, 105–122.

—— and C. S. Tang, "Modelling the Costs and Benefits of Delayed Product Differentiation," *Management Sci.*, 43 (1996), 40–53.

Pine, B. J., B. Victor, and A. C. Boynton, "Making Mass Customization Work," *Harvard Business Rev.*, September–October (1993), 108–119.

Zakon, A. and R. W. Winger, "Consumer Draw: From Mass Markets to Variety," *Management Review*, April (1987), 20–27.

Accepted by Gabriel Bitran; received October 1, 1996. *This paper was with the authors* 9 *months for* 2 *revisions.*