

# Chapter 3

## Flexibility Principles

Stephen C. Graves

Massachusetts Institute of Technology<sup>1</sup>

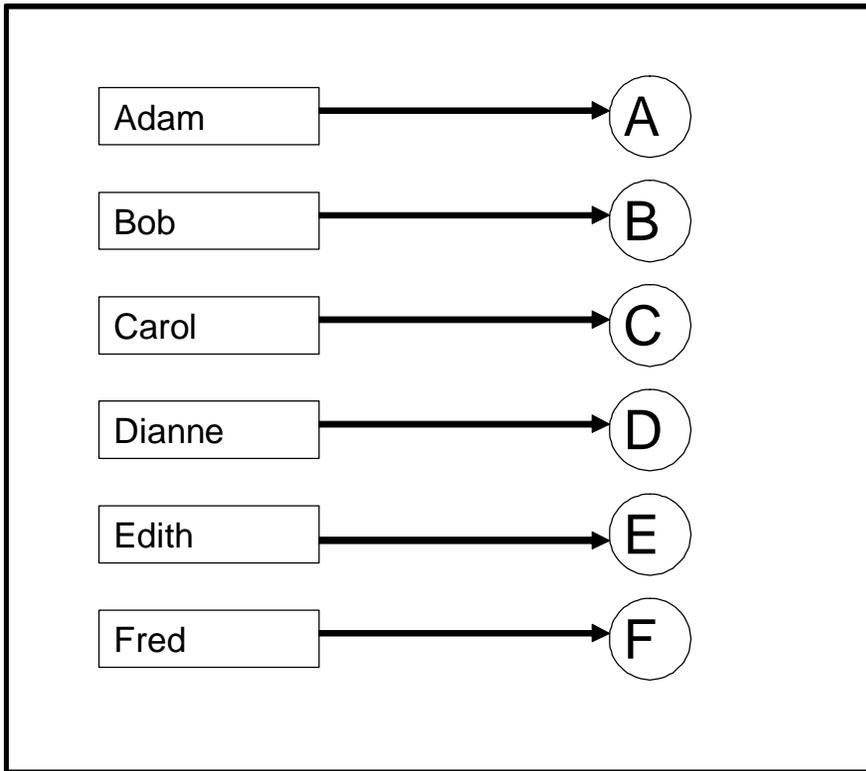
*Consider a setting with multiple demand classes that are served by a set of resources. When each resource is limited to serving only one demand class, we can often have a situation where some resources are under-utilized and idle, while others are over-utilized and not able to meet the demand. One tactic for dealing with this situation is to make each resource more flexible so that it can serve more than one demand class. But how much flexibility should each resource have and what is the best way to deploy flexibility across the resources? This chapter shows that when done right, limited flexibility can provide almost the same level of benefits as complete flexibility.*

### Introduction

FCM is a small manufacturing company that assembles a family of fuel controllers for jet engines and turbines. The assembly of each type of fuel controller is a complex, labor-intensive activity that requires specialized tools and skills and takes half of a work day; that is, the assembly time for each unit is four hours. FCM supplies six distinct products, labeled A, B, C, D, E and F, and has six certified technicians, Adam, Bob, Carol, Dianne, Edith and Fred. Each technician has been trained to assemble one type of fuel controller, as shown in Figure 1.

---

<sup>1</sup> This research has been supported in part by the MIT Leaders for Manufacturing Program, a partnership between MIT and major manufacturing firms; and by the Singapore-MIT Alliance, an engineering education and research collaboration among the National University of Singapore, Nanyang Technological University, and MIT.



**Figure 1: Base Case at FCM, a Dedicated System**

At the start of each week, FCM receives orders from its customers, which are Original Equipment Manufacturers (OEMs) for which the fuel controller is a subsystem required for their final products. The OEMs expect delivery of the completed orders at the start of the following week. Thus, FCM has exactly one week to meet the orders.

FCM follows a simple protocol for running its assembly operation. At the start of each week, it schedules each technician to assemble the number of units exactly equal to the demand for his/her product. For instance, if demand for product A is 8 units, then Adam will assemble 8 units, no more and no less, in the coming week so as to meet the demand for A.

Based on past order history, FCM has developed a forecast for demand for each product. For instance, FCM expects demand for product A to be 10 units on average. However, this is just a forecast and actual demand for A can deviate from the forecast. Indeed, weekly demand for product A has ranged between 8 and 12 units, and FCM

forecasts that the likelihood that demand is 8, 10 or 12 units is equally likely. [Customer orders are typically for pairs as each engine will require two fuel controllers.] That is, the probability is  $1/3$  for each of the possible outcomes, 8, 10 or 12 units.

Similarly, FCM has made a forecast for demand for the other five products, and each product has the same demand forecast. For each product the possible demand outcomes are 8, 10 or 12, with each occurring with equal likelihood, namely each with probability of  $1/3$ . Furthermore, the demand for each product is independent from week to week and independent of the demand of each of the other products. Thus, the fact that the demand for product A in the current week is 12 has no influence on the demand outcome for product B or on the demand outcome for A in the following week.

This is a quite stylized and simplified example, but is representative of many operational settings in which there are multiple demand categories or classes (products A, B, ..F) that are served by a set of resources (Adam, Bob, ... Fred). Many service systems operate with parallel servers and different demand classes. For example, the resources or servers might be physical therapists, or barbers, or tutors, and the categories of demand would correspond to customers with differing service requirements (e.g., therapy for different body parts; hair cuts for girls, boys and adults; help with math, science, or history). In a manufacturing operation, we might have a set of equipment that serves different types of jobs. For example, the photolithographic equipment in a semiconductor fabrication plant is responsible for the two to three dozen photolithographic process steps that are required for each lot of wafers that flow through the facility. Each step requires a different setup and specific tooling, e.g., a mask. As a consequence, a plant will often dedicate each process step to a limited subset of the available equipment.

A challenge in these settings is how best to match the resources to the demand. At FCM we have six technicians, each with a nominal work week of 40 hours. Thus, if demand for each product were 10 units each week, then we would have a perfect match between the available resources and the demand. Each unit requires four hours of assembly time, so each worker would complete his/her ten units in his/her normal work week of 40 hours. Unfortunately, when demand can vary from week to week, we lose this balance. When demand for product A is 12 units, Adam needs to work 48 hours, or six days, to complete the work within the week; Adam is willing to do this, as he gets

paid an overtime rate of 150% of his normal pay for the extra 8 hours he works. When demand for product A is 8 units, Adam is underutilized and needs to work only 32 hours. However, FCM still pays him for a full week, namely for 40 hours, and will assign him other tasks to fill out the remaining 8 hours of work time.

Because of this imbalance, each technician works a day of overtime once every three weeks on average, and thus, FCM pays, on average, for two days of overtime each week. FCM naturally wonders about how to mitigate or avoid this overtime expense. In such settings, the first thought is to explore how the variability might be eliminated or reduced. If FCM might somehow entice the customers to order exactly 10 units of each product each week, then there would be a perfect match of demand and resources and no need for overtime. This might be possible if there were a few large customers, and if FCM were willing to provide an incentive, like a price discount for stable ordering. However, this could be more costly than the overtime, and/or impossible to achieve if there were lots of small customers.

If FCM cannot affect the demand variability, then it might consider the common operational tactics of creating a buffer of some form as a counter measure to the variability.

One form of buffer is a *capacity buffer*; FCM might hire and train a seventh technician, who would be capable of assembling any of the products (e.g., Jack, a jack-of-all trades). Then each week Jack would pitch in and help out with any demand surpluses; that is, he would assemble the 11<sup>th</sup> and 12<sup>th</sup> units for any product that has demand of 12. This would eliminate the need for overtime, with one exception. If the demand for each product were 12, then Jack would have 12 units to assemble and he would need to work an extra day of overtime; but this would occur very rarely, with probability  $\left(\frac{1}{3}\right)^6 = \frac{1}{729}$ , on average once every 729 weeks (about 14 years). Thus, adding capacity in the form of an additional technician effectively eliminates the need for overtime. The cost, though, might be quite high; we now need an extra person, fully trained to assemble every product. In most settings, this cost is likely to exceed the savings, namely saving two days of overtime each week, unless we can find other things for Jack to do.

A second way to buffer variability is to create a *time buffer*. Currently FCM commits to deliver orders within a week. By lengthening its delivery lead time from one week to, say, two or three weeks, FCM could create a time buffer that would allow it to smooth out the week to week demand variability. This could substantially reduce FCM's overtime, but only if its customers were willing to accept a longer delivery lead time.

The third type of buffer is to use an *inventory buffer*. FCM could assemble its fuel controllers to stock, and then serve demand from this inventory. Similar to a time buffer, the inventory buffer would allow FCM to smooth its production and reduce the amount of overtime. Making to stock presumes that each product is not customized to an order and that each product has a stable design specification. The cost for this tactic is the holding cost for the inventory.

Let's suppose that none of these options is readily viable for FCM. It's too expensive to hire and train a seventh technician; customers will not accept a longer lead time; and the fuel controllers cannot be made to stock as each has some customized content and thus must be made to order. Then another option to reduce overtime might be to increase the current capability of the work force by cross-training. Currently each technician is trained to assemble just one product, and each product can be assembled by just one technician. We term this a *dedicated system*, in which each resource is dedicated to one product and each product can be assembled by only one resource. Suppose we could train and certify the technicians to assemble more types of products. As we will see, this cross-training is a form of risk pooling, as the resource flexibility permits the pooling of the demand variability across multiple products that share the same set of resource; see Chapter 9 (Sobel 2006) in the volume for a discussion of risk pooling.

What is the benefit from having a *flexibility buffer* in the form of cross-trained labor? How is the best way to create this flexibility? We will explore these questions in the remainder of this chapter.

## **Total Flexibility**

To get started, let's consider an extreme case. Suppose we train each technician to be able to assemble all six products. Again, each unit requires four hours of assembly

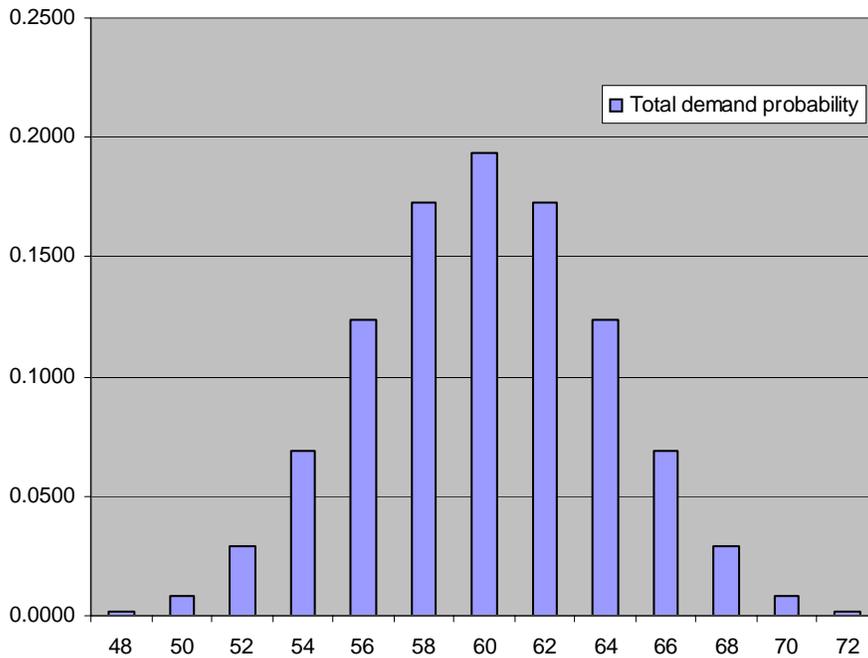
time, so each technician can assemble ten units in the normal work week. Thus, the work force can assemble 60 units each week without resorting to overtime, regardless of the product mix. This is because the technicians are fully interchangeable, with each of them capable of doing any mix of work – so each can do any 10 units per week. However, overtime will be needed to assemble any demand beyond 60 units.

Under the assumption that demand for each of the six products has outcomes of 8, 10 and 12 units with equal probability, we can find the probability distribution for the total aggregate demand. We report this in Table 1 and depict it as a histogram in Figure 2. We also report in Table 1 the cumulative probability and overtime required for each demand outcome. For instance, the weekly aggregate demand is 60 units or less with probability 0.5967; also, if demand were 66 units, then overtime is needed to assemble 6 units, requiring a total of 24 hours.

Demand Outcome	Probability	Cumulative Probability	Overtime Production (units)	Overtime (hours)
48	0.0014	0.0014		
50	0.0082	0.0096		
52	0.0288	0.0384		
54	0.0686	0.1070		
56	0.1235	0.2305		
58	0.1728	0.4033		
60	0.1934	0.5967		
62	0.1728	0.7695	2	8
64	0.1235	0.8930	4	16
66	0.0686	0.9616	6	24
68	0.0288	0.9904	8	32
70	0.0082	0.9986	10	40
72	0.0014	1.0000	12	48

**Table 1: Probability Distribution for Total Demand**

**Figure 2: Total demand probability**



From the table we see that overtime is needed for only about 40% of the weeks; in six out of ten weeks, we would not need any overtime. In contrast, in the dedicated system where each technician can assemble only one product type, then each technician works overtime with probability  $\frac{1}{3}$ ; thus, the probability that there is no overtime needed by any of the technicians would be  $\left(\frac{2}{3}\right)^6$  or only 9% of the weeks.

From Table 1 we find the average amount of weekly overtime to be 6.32 hours. We compare this to the base case in which each technician could assemble only one product; in this case the average overtime is 16 hours per week. Hence, for FCM the benefit of complete cross-training of the work force is a reduction in the average overtime from 16 hours to 6.32 hours per week. This is the absolute best that can be done from increasing the flexibility of the current work force. In the following sections we examine how close we can get to this best case with more selective cross-training.

### Limited Flexibility: Train Adam to Assemble B

As training is quite expensive, a natural question is what the benefits might be from a less ambitious schedule of cross-training. Again, we will explore this by starting with an extreme case. Suppose we train Adam so that he can assemble product B, in addition to product A. All of the other technicians remain dedicated to their products. What is the benefit to FCM from this investment?

With a little thought, one realizes that the only benefit from this investment is when demand for product A is low and demand for product B is high, that is when  $d_A = 8$  and  $d_B = 12$ , where  $d_x$  denotes demand for product x. Then Adam can assemble 8 units of A and 2 units of B within his 40-hour work week and Bob can assemble the remaining demand for B, namely 10 units, in his normal week. Thus, the cross-training of Adam to produce B would save one shift (8 hours) of overtime for Bob whenever  $d_A = 8$  and  $d_B = 12$ .

For all other demand outcomes there is no benefit. When  $d_A = 10$  or  $12$ , then Adam is of no help to Bob, as he must spend his entire time assembling product A. When  $d_A = 8$  and  $d_B = 8$  or  $10$ , then Adam has idle time but Bob does not need any help.

The probability that  $d_A = 8$  and  $d_B = 12$  is easily found to be:

$$\Pr[d_A = 8 \text{ and } d_B = 12] = \Pr[d_A = 8] \times \Pr[d_B = 12] = \left(\frac{1}{3}\right) \times \left(\frac{1}{3}\right) = \frac{1}{9}.$$

Thus, relative to the base case of the dedicated system, the benefit from training Adam to assemble B is a reduction in overtime per week, on average, of

$$\frac{1}{9} \times 8 \text{ hours} = 0.89 \text{ hours}.$$

For FCM the cost of an hour of overtime is approximately \$80. Thus, the annual benefit from training Adam to assemble B is given by

$$\frac{\$80}{\text{hour}} \times \frac{0.89 \text{ hours}}{\text{week}} \times \frac{52 \text{ weeks}}{\text{year}} \approx \frac{\$3700}{\text{year}}.$$

FCM needs to compare this to the cost of training; a technician requires between 100 and 150 hours of training in order to be certified to assemble a new product. FCM estimates the one-time cost of this to be on the order of \$10,000. Thus, there would be about a three-year pay-back from training Adam to assemble B, which is an unacceptable return for FCM. FCM requires a two-year pay back on this type of investment.

### **Limited Flexibility: Create Two-person Teams**

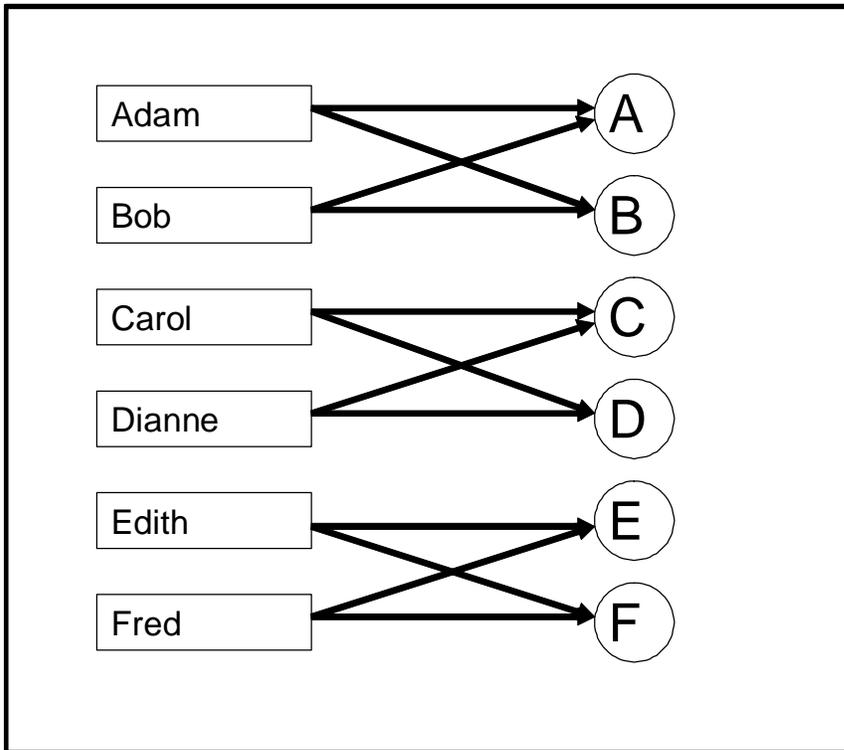
Suppose FCM also trains Bob to assemble product A, so that Adam and Bob can form a two-person team to handle both products A and B. From a similar analysis, we find that there are now two demand scenarios when this cross-training leads to a benefit:

$d_A = 8, d_B = 12$  and  $d_A = 12, d_B = 8$ . Each scenario occurs with probability  $\frac{1}{9}$  and saves 8 hours of overtime, relative to the base case in Figure 1. Thus, cross-training both Adam and Bob leads to an average reduction in overtime of  $\frac{2}{9} \times 8 \text{ hours} = 1.78 \text{ hours}$ .

Based on these two examples, one might conclude that cross-training is not a viable option for FCM. Training each technician to be capable to assemble a second product entails a one-time cost of \$10,000 per technician. The annual benefit from each cross-trained technician would be a reduction in overtime leading to a savings of \$3700 per year. Thus, each investment would be recovered in three years, which is not acceptable for FCM. One cross-training plan is shown in Figure 3, in which FCM establishes three two-person teams. This flexibility or cross-training plan results in a reduction in the expected overtime from 16 hours per week for the base case to 10.67 hours per week.

We previously had found that training each technician to assemble all products reduces the expected overtime from 16 hours per week to 6.32 hours per week, which is the best we can do. The cross-training plan shown in Figure 3 closes more than half of the gap between the dedicated system (16 hours of overtime per week) and the total-flexibility configuration (6.32 hours of overtime per week). We can close this gap by

adding more flexibility to the configuration in Figure 3, by continuing to cross-train each technician; but one wonders whether there is a better way to proceed.



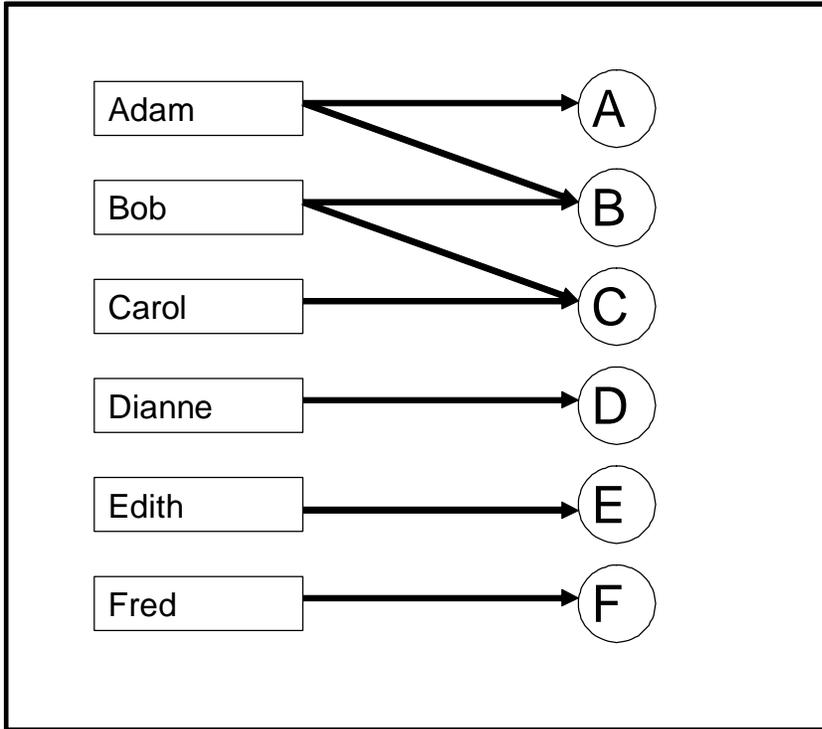
**Figure 3: Two-person Teams**

**Limited Flexibility: Build a Chain**

Suppose instead of creating two-person teams we follow a different strategy, as shown in Figure 4. That is, let’s train Adam to assemble B, and Bob to assemble C. As before, we see that there are benefits when  $d_A = 8$ ,  $d_B = 12$  and  $d_B = 8$ ,  $d_C = 12$ . Each scenario occurs with probability  $\frac{1}{9}$  and saves 8 hours of overtime, relative to the dedicated system with no cross-training.

But we have another demand scenario to consider here. Suppose demand for product A is low and demand for product C is high, namely  $d_A = 8$ ,  $d_B = 10$ ,  $d_C = 12$ . Then, from inspection, we see that we can accommodate the surplus demand for C without overtime: Adam assembles 8 units of A and 2 units of B; Bob assembles 8 units

of B and 2 units of C; and Carol assembles the remaining 10 units of C. Thus, we have been able to use Adam's excess capacity to meet the demand surplus for product C, even though Adam is not capable of assembling product C!



**Figure 4: Chaining Adam, Bob and Carol**

This demand scenario occurs with probability  $\left(\frac{1}{3}\right)^3 = \frac{1}{27}$  and saves 8 hours of overtime, relative to the base case. Thus, cross-training Adam to product B and Bob to product C leads to an expected reduction in overtime of

$$\left(\frac{2}{9} + \frac{1}{27}\right) \times 8 \text{ hours} = 2.07 \text{ hours} .$$

Recall that cross-training Adam to product B and Bob to product A led to a reduction of 1.78 hours. Thus, the benefit from cross-training Adam and Bob depends on how it is done. The plan in Figure 4 is better than a two-person team focused on products A and B. The reason is clear from consideration of the demand scenario  $d_A = 8, d_B = 10, d_C = 12$ . When demand is low for A but high for C, the configuration

in Figure 4 permits Adam's excess capacity to be applied *indirectly* to satisfy the excess demand for C. This is possible because Adam is linked to product C by way of Bob: since Adam can produce B, he can off load work from Bob, who can then use this capacity to help Carol with C.

Furthermore, we observe that the incremental benefit from the second investment in cross-training (training Bob to assemble product C) is *greater* than the benefit for the initial investment (training Adam to assemble B). When FCM trains Adam to assemble B, it obtains a reduction of 0.89 hours of overtime per week, for an annual savings of \$3700. Training Bob to assemble C increases the overall benefit to a reduction of 2.07 hours per week, for an annual savings in excess of \$8600. Thus, the incremental annual savings from training Bob is more than \$4900! In most contexts, the rule of thumb is that we get decreasing incremental returns from additional investment (for an example, see Chapter 2 (Bartholdi, 2006) in this volume); this is a somewhat surprising counter example to this rule.

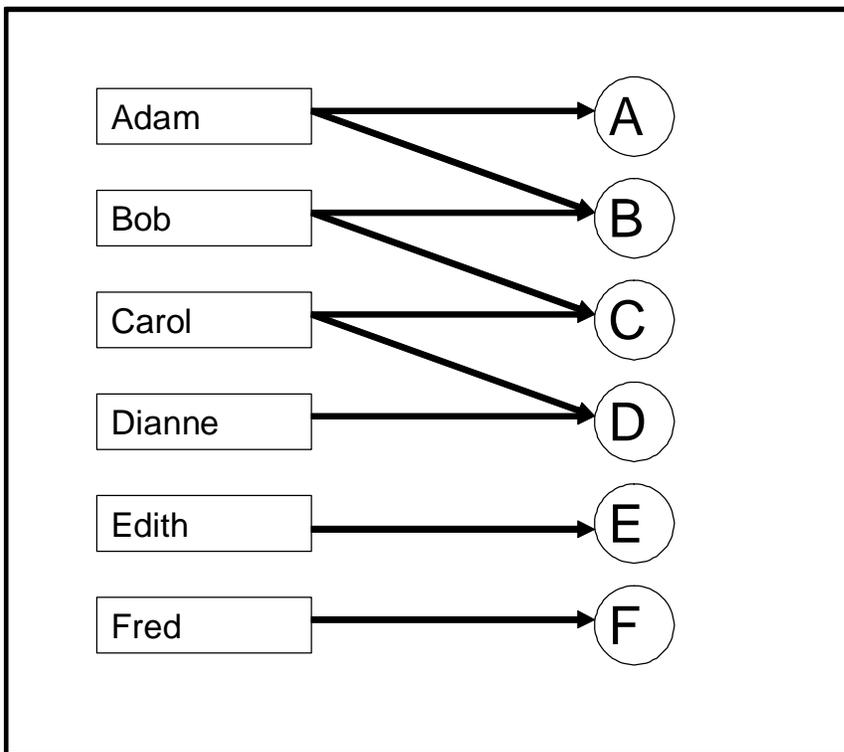
The configuration in Figure 4 is the best way to train two technicians to each do an additional product. In network terminology, this is the best way to add two links or arcs to the dedicated system, shown in Figure 1, where each link signifies the training of one technician to assemble one product. We say that there is a *chain* that connects products A, B and C to the resources Adam, Bob and Carol, in that these products and technicians are connected by the links. Within the chain, we can trace a path between any product or technician to any other product or technician using the links.<sup>2</sup> The existence of this chain is what provides the extra benefit, or dividend, relative to the two person-teams in Figure 3: the chain permits Adam's excess capacity to be deployed so as to handle the excess demand at C.

There are many equivalent configurations to that shown in Figure 4, in which we add two links and form a chain. For instance, we could train Adam to assemble C and train Carol to assemble E. This will create a chain from A to Adam, Adam to C, C to Carol, and Carol to E. The benefit is the same as that for Figure 4. When adding two links, the key idea is to coordinate the training so as to create the longest chain.

---

<sup>2</sup> In network terminology, a chain is a connected sub-graph of the network.

These ideas extend as we add more links to the network. Consider the cross-training plan shown in Figure 5. To see the benefits, we enumerate all of the relevant demand scenarios in Table 2, where blank cells signify that the demand can assume any outcome. From the table we can compute the expected reduction in weekly overtime to be 3.46 hours, for an annual savings of nearly \$14,400. Thus, the incremental benefit from training Carol to assemble product D is more than \$5700 per year. Again we have an increasing return from the incremental investment in cross-training that is attributable to chaining; we now have a chain connecting product A to product D and technicians Adam to Dianne, which permits the excess capacity of Adam and Bob to be deployed to meet surplus demand for products C and D. Yet, we still don't quite have a viable business case for FCM: the required investment is \$30,000 to train Adam, Bob and Carol and the annual savings are less than \$15,000, leading to a more than two-year payback.



**Figure 5: Chain Connecting Adam, Bob, Carol, Dianne with Products A, B, C, D**

$d_A$	$d_B$	$d_C$	$d_D$	probability	Overtime savings (hrs)
8	12			0.1111	8
	8	12		0.1111	8
		8	12	0.1111	8
8	10	12		0.0370	8
	8	10	12	0.0370	8
8	10	10	12	0.0123	8
8	8	12	12	0.0123	8

**Table 2: Demand Cases and Incremental Benefits from Configuration in Figure 5<sup>3</sup>**

### Limited Flexibility: Complete the Chain

We can continue in this fashion and add additional flexibility – train Dianne to assemble E and train Edith to assemble F. We then have a configuration as shown in Figure 6, with a chain that connects all products and all technicians. We find that the incremental savings from training Dianne to assemble product E is \$6300 per year; the incremental savings from then training Edith to assemble F is \$6711 per year. Thus, we find that there continues to be an increasing return from additional cross-training – as long as we keep building a longer chain. Furthermore, from the standpoint of FCM, the payback for the \$10,000 investment in cross-training is now about 18 months for both Dianne and Edith; this is now a quite attractive return on investment for FCM.

However, there is still a big gap between the performance of the chain in Figure 6 (with all products and technicians connected) and that for a total flexibility configuration in which each technician can assemble all products. The expected amount of overtime for the configuration in Figure 6 is 9.42 hours per week, whereas it is 6.32 hours per week if

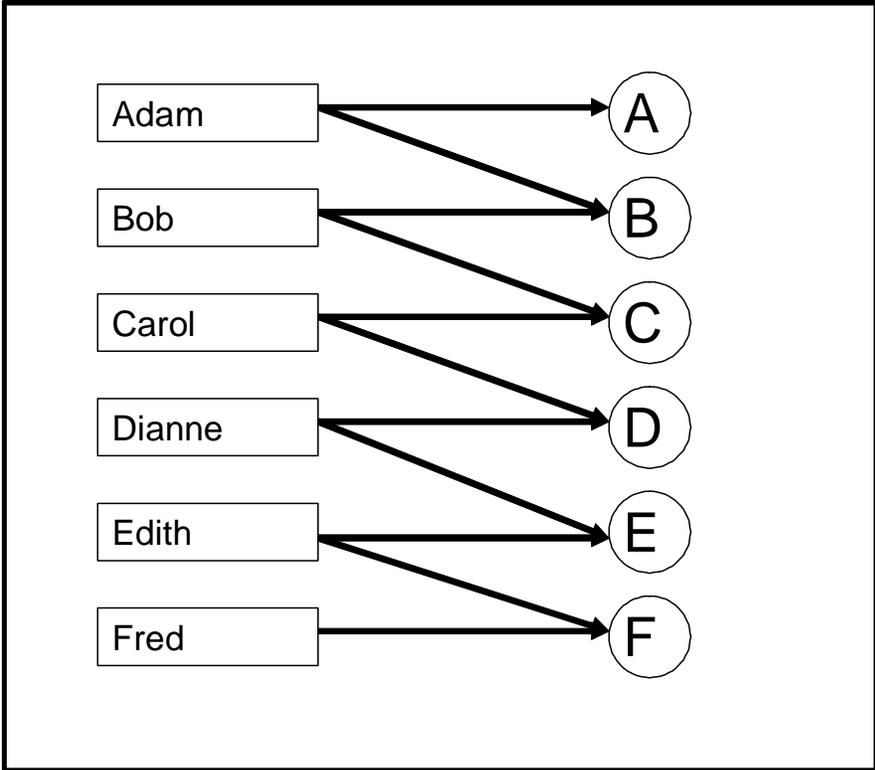
<sup>3</sup> The probabilities in the table are just the product of the probabilities for the specified demand outcomes, which are assumed to be independent; for instance, the probability for the demand case in the fifth row is found as .

$$\Pr[d_B = 8, d_C = 10, d_D = 12] = \Pr[d_B = 8] \times \Pr[d_C = 10] \times \Pr[d_D = 12] = \left(\frac{1}{3}\right)^3 = \frac{1}{27} = 0.0370$$

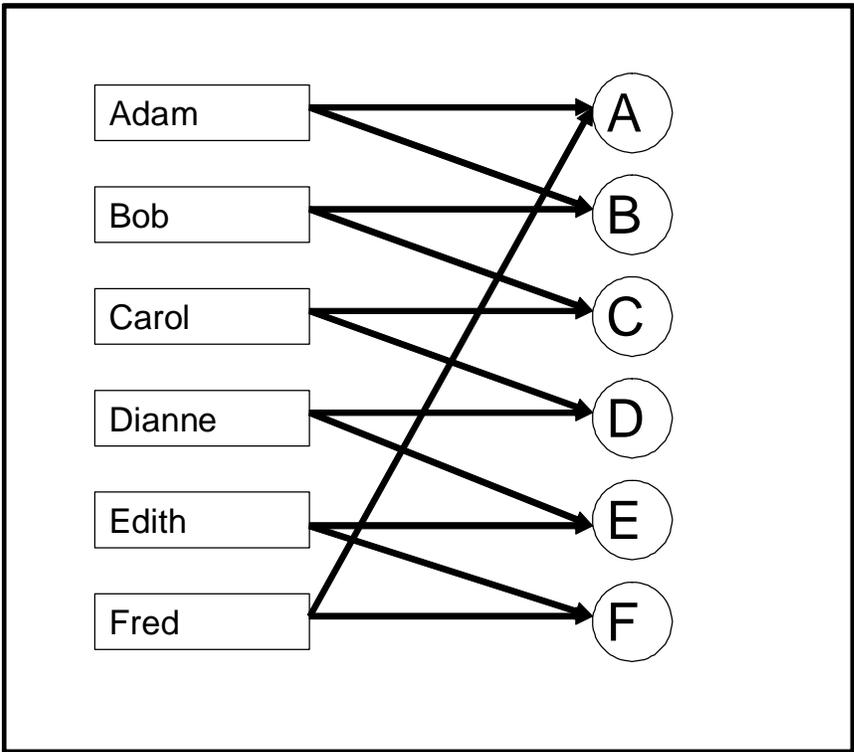
each technician is fully flexible, capable of assembling any product. Again, we ask whether we can close this gap with limited flexibility, and if so, how?

We note that whereas we have connected all products and technicians in Figure 6, there are some asymmetries in the configuration. There are two technicians that can assemble each product, except for product A that can only be made by Adam; and we have cross-trained all technicians to assemble two products except for Fred. What if we now train Fred to assemble A, as shown in Figure 7?

We find that this last investment in cross-training is extremely beneficial. As shown in Table 3 and Figure 8, adding this link, namely training Fred to assemble A, results in an incremental annual savings of over \$12,800. Furthermore, we find that the configuration in Figure 7 performs equivalent to the total flexibility configuration. Each configuration averages 6.32 hours of overtime per week, for an annual overtime cost of \$26,300. Yet, these two systems differ dramatically in terms of their investment in cross-training. On the one hand, the closed chain configuration requires an investment of \$60,000, to train each technician to assemble a second product; the savings relative to the base case of a dedicated system is \$40,000 per year, leading to payback in 1.5 years. On the other hand, the total flexibility configuration entails training each technician to produce all five other products, say, for a one-time training cost of \$50,000 per technician for a total of \$300,000 for FCM; the payback for the investment in total flexibility would be 7.5 years.



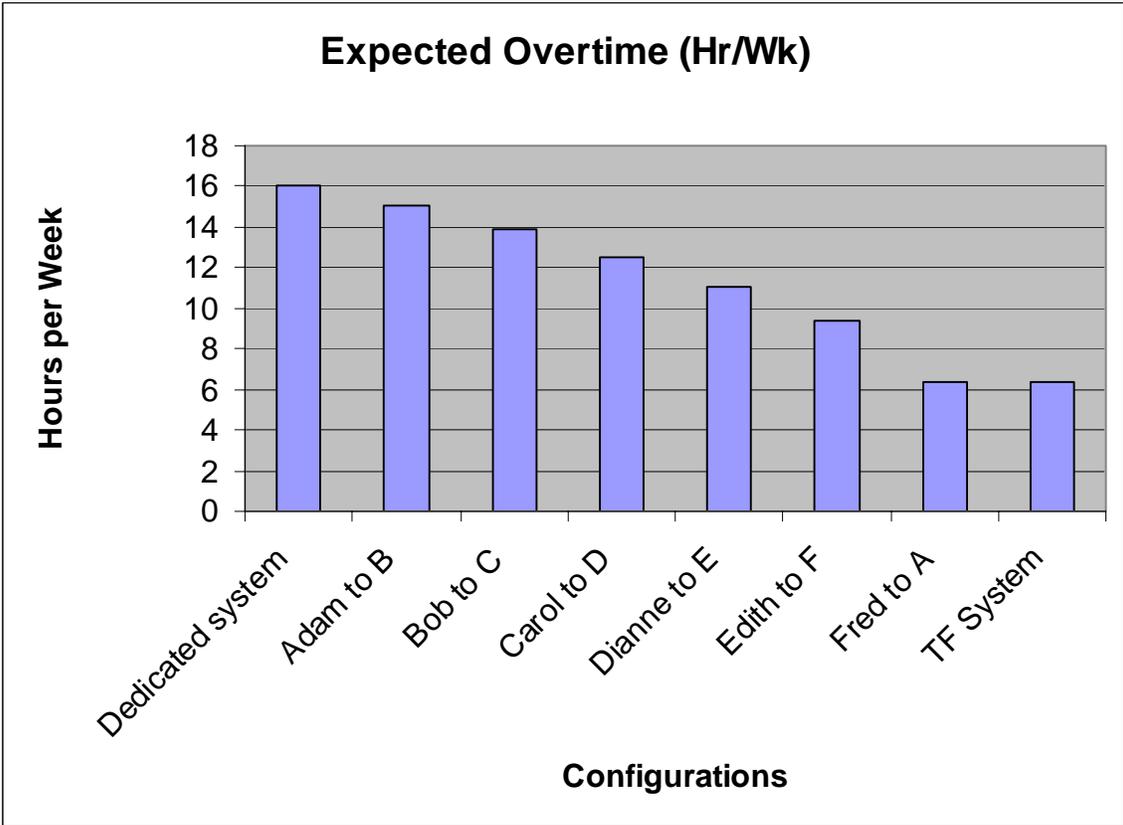
**Figure 6: Chain Connecting All Products and Technicians**



**Figure 7: Closed Chain**

Configuration	Expected overtime per week (hours)	Expected cost per year (\$/yr)	Incremental reduction from link (\$/yr)
Dedicated system	16.00	66,560	
Train Adam on B	15.11	62,862	3,698
Train Bob on C	13.93	57,932	4,930
Train Carol on D	12.54	52,180	5,752
Train Dianne on E	11.03	45,880	6,300
Train Edith on F	9.42	39,169	6,711
Train Fred on A	6.32	26,295	12,874
Total Flexibility	6.32	26,295	

**Table 3: Performance of Chains, Adding One Link at a Time**



**Figure 8: Results from Adding Flexibility to Base Case**

Why is this last link so valuable? This last link *closes the chain* in that we now have a complete circuit that connects all products and technicians. This circuit or closed chain permits us to move excess capacity from any part of the network to meet surplus demand for any of the products. In particular, if we consider the configuration in Figure 6, even though we have chained all of the products and resources together, there is no way to move excess capacity from the bottom of the network (i.e., Edith or Fred) to meet surplus demand at the top of the network (products A and B); adding the last link from Fred to product A now permits this to happen. Indeed, with the closed chain we now have a remarkable ability to match the available capacity to the variable demand.

Furthermore, in this example any additional cross-training provides no benefit. The closed chain configuration is literally equivalent to the system with total flexibility, for the given demand outcomes. Having more cross-training does not lead to any reductions in the overtime requirements. Thus, this is the best we can do.

As another way to see the value from chaining, let's compare the plan in Figure 3 (two-person teams) with the plan in Figure 7 (closed chain). In both instances we train each technician to be able to assemble a second product. In both instances, each product can be assembled by two technicians. These two systems require the same investment in cross-training, \$60,000, and would seem to be very similar in structure. But the return on this investment is much different, as these two systems perform much differently. In comparison with the dedicated system (Figure 1), the plan with two-person teams reduces the weekly overtime, on average, from 16 hours to 10.67 hours. In contrast, the plan with the closed chain reduces the weekly overtime, on average, from 16 hours to 6.32 hours, yielding almost twice the benefit from the two-person teams.

### **General Insights on Flexibility**

We have developed and used the FCM example as a way to examine flexibility in an operating setting. In this case, we have a relatively simple production system with uncertainty in the demand requirements and with production inflexibility due to constrained resources. We have an opportunity to increase the flexibility of the system

by cross-training; however, there is a tradeoff between the cost of cross-training and the possible benefits from reduced overtime. From this example we observe the following:

- The benefits from increased flexibility depend on how it is deployed; in particular, we get the greatest benefits by building longer and longer chains that connect more and more products and resources together.
- As we build a chain, we get increasing returns from our investments in flexibility; that is, the incremental benefit from an investment in flexibility grows as we increase the size of the chain.
- Once we have created a chain that connects all of the products and resources together, we get the largest incremental benefit by closing the chain; that is, we get the largest return by adding a link that creates a circuit connecting all of the products and resources.
- In our example the performance of the closed chain configuration is equivalent to that for a system with total flexibility, that is, a system in which each resource can produce any of the products. Thus, there is no benefit from adding any flexibility beyond the closed chain.

Although we have developed these observations for a quite stylized setting, we have found them to be quite robust. Indeed, we contend that (1) limited flexibility, when deployed in the right way, yields most of the benefits of total flexibility, and (2) limited flexibility provides the greatest benefits when configured to chain products and resources together to the greatest extent possible. In particular, as we increase the demand variability for the products, we do find that the performance of a totally flexible system can be better than that for a system with a closed chain configuration. However, for realistic ranges of the demand variability, this performance gap, albeit not zero as in the case of the example, is quite small. As a consequence, we observe that there is usually a very poor return from adding any flexibility beyond the closed chain.

The FCM example has a great amount of symmetry – same number of products and resources, same production rates for each resource, same demand characteristics for each product. This is not the case for more realistic scenarios, where each product has

distinct demand characteristics, the production rates vary across the resources and the number of products differs from the number of resources. Nevertheless, we have found that the principles cited above are most helpful in providing guidelines for exploring flexibility options. We still want to build as long a chain as possible and to create a circuit that connects the primary products and resources together. In addition we try to equalize the load or expected demand that is assigned to each resource, and try to connect each product to its share of capacity. With these additional considerations, limited flexibility deployed to create chains provides most of the benefits of total flexibility.

The principles that we uncovered with the FCM example apply to many other contexts. The general conditions are that we have multiple resources and multiple tasks. A resource is required to perform each task. However, some form of investment, qualification or training is required in order for a specific resource to have the capability to perform a specific task. As a consequence, each resource has limited capability in that it is qualified to perform only a subset of all of the tasks. Furthermore there is some uncertainty in demand requirements, namely how much work is required for each type of task. Key issues for such an operation are to decide what capability each resource should have and to predict the performance of a given configuration of resources.

## **Applications**

The FCM example is one context. The multiple resources are highly skilled technicians. The tasks are the assembly of a complex product, e.g., fuel controllers. There is uncertainty in the demand requirements for these products, and there is a non-trivial expense to train a technician to be certified to assemble each of the products.

Other examples include

- Maintenance personnel and repair tasks. Equipment in a plant fails at random intervals. The repair task is specific to the type of equipment and requires a qualified maintenance engineer. The flexibility questions are to decide the training for each maintenance engineer, and then how to dynamically assign to repair tasks.
- Service representatives and/or technicians in a call center. A call center receives a mix of types of calls: questions about a product warranty, questions

about how to use a product or how to install software, questions about how to return a product for repair, requests to place an order or make a reservation, questions about order status or to cancel/change an order, etc. Service agents handle these calls but require specialized skills and/or access to on-line information, depending upon the product and the nature of the call. Most call center operating systems will automatically screen calls to determine the type of call so as to direct it to an appropriate service agent. The flexibility questions are to decide how many service agents and their training, and how to route calls as they arrive to the system.

- An assembly cell. An assembly cell is set up to assemble a product or product family. The assembly of a product is broken down into a sequence of specific tasks. The assembly cell consists of a sequence of workstations, at each of which a subset of the tasks are to be performed. A trained technician is required to perform the tasks at each workstation, but often a cell will have fewer technicians than work stations. In this case, technicians will float between workstations. The flexibility questions are to decide how many technicians and what amount of cross-training is needed to assure an efficient operation.
- Semiconductor fabrication. The production of a semiconductor device entails several hundred process steps that are necessary to build twenty to thirty layers of circuitry. For many of the process steps, a device returns to the same set of process tools for each layer. Each process tool is capable of performing the process task for any layer, but usually requires a setup and/or specialized tooling to switch from one layer to another layer. The flexibility questions are to decide which tools to dedicate to which layers, which tools to flex between multiple layers, and then how to schedule the work flow through the set of process tools.
- Assembly plants. In the automobile industry, the assembly plants for final vehicle assembly can produce multiple vehicle models or name plates. But such flexibility requires extensive planning and investment in tooling and

training. The flexibility questions are to decide for a set of assembly plants and a set of vehicle models, which models are to be produced in which plants.

## **Historical Background**

The key reference for this chapter is Jordan and Graves (1995) who examine the benefits of process flexibility in the context of a set of plants producing a set of products. They introduce the concept of chaining and establish that (i) limited flexibility will yield nearly all of the benefits of total flexibility, if configured in the right way, and that (ii) limited flexibility provides the greatest benefits when configured to chain as many plants and products together as possible.

There have been a number of important subsequent developments. We mention a few here.

Hopp et al. (2004) examine cross-training in a serial production or assembly line and establish the value of chaining; they find that a cross-training strategy that attempts to chain together workers and workstations performs best.

Gurumurthi and Benjaafar (2004) consider a queuing system with multiple customer classes and heterogeneous servers. Each server can have the capability to process more than one customer class. They develop a framework and computational algorithm for evaluating the performance of different flexibility configurations and control policies. They find that limited flexibility, in the form of chaining, works extremely well in many settings.

Graves and Tomlin (2003) extend the model from Jordan and Graves to a multistage supply chain in which each product requires processing at each stage of the supply chain. They show how the single-stage guidelines from Jordan and Graves adapt and apply to a multistage setting.

Jordan et al. (2004) and Inman et al. (2004) examine two contexts in which cross-training can arise in the automobile industry. Jordan et al. (2004) consider the value of cross-training a workforce to perform maintenance and repair tasks. Inman et al. (2004) consider how best to do cross-training for an assembly line to counteract the impact from absenteeism. In both case, they find that a chaining strategy is effective.

Iravani et al. (2005) consider a general setting with multiple parallel resources and multiple parallel tasks. They define a configuration by a specification of the capabilities of each resource, i.e., which tasks each resource can perform. They develop and examine various measures for predicting the flexibility of a configuration, and provide additional evidence for the value of chaining.

Wallace and Whitt (2005) consider cross-training of agents in a call center with several types of calls. They discover that when each agent has only two skills, the system performance can be comparable to that when each agent has every skill, provided that the cross-training is done the right way. Again, the concept of chaining together as many agents and skills as possible is the key insight.

### **Selected Bibliography**

Bartholdi, J. J. (2006), "The Knapsack Problem," Chapter 2 in When Intuition Fails: Insights from Basic Operations Management Models and Principles, edited by D. Chhajed and T. Lowe, Springer, New York.

Graves, S. C. and B. T. Tomlin (2003), "Process Flexibility in Supply Chains," *Management Science*, July 2003, Vol. 49, No. 7, pp. 907 – 919

Gurumurthi, S. and S. Benjaafar (2004), "Modeling and Analysis of Flexible Queueing Systems," *Naval Research Logistics*, Vol. 51, No. 5, pp. 755 – 782.

Hopp, W. J., E. Tekin, and M. P. Van Oyen, (2004), "Benefits of Skill Chaining in Serial Production Lines with Cross-Trained Workers," *Management Science*, January 2004, Vol. 50, No. 1, pp. 83 – 98.

Inman, R. R., W. C. Jordan and D. E. Blumenfeld, (2004), "Chained Cross-Training of Assembly Line Workers," *Int. J. Prod. Research*, May 2004, Vol. 42, No. 10, pp. 1899 – 1910.

Iravani, S. M., M. P. Van Oyen and K. T. Sims (2005), "Structural Flexibility: A New Perspective on the Design of Manufacturing and Service Operations," *Management Science*, February 2005, Vol. 51, No. 2, pp. 151 – 166.

Jordan, W. C. and S. C. Graves (1995), "Principles on the Benefits of Manufacturing Process Flexibility," *Management Science*, April 1995, Vol. 41, No. 4, pp. 577 - 594.

Jordan, W. C., R. R. Inman, and D. E. Blumenfeld, (2004), "Chained Cross-Training of Workers for Robust Performance," *IIE Transactions*, October 2004, Vol. 36, No. 10, pp. 953 – 967.

Sobel, M. J. (2006), "A Risk Pooling Fable," Chapter 9 in When Intuition Fails: Insights from Basic Operations Management Models and Principles, edited by D. Chhajed and T. Lowe, Springer, New York.

Wallace, R. B. and W. Whitt, (2005), "A Staffing Algorithm for Call Centers with Skill-Based Routing," *Manufacturing & Service Operations Management*, Fall 2005, Vol. 7, No. 4, pp. 276 – 294.

### **Building Intuition**

The performance of a service or manufacturing system with multiple products being served by multiple resources depends upon the flexibility of the resources. A resource is flexible if it is capable of serving several products, but making a resource flexible can be quite expensive. Hence there is always a tradeoff between the costs and benefits from increasing the flexibility in these systems.

We show by example that limited flexibility, when deployed in the right way, yields most of the benefits of total flexibility. Furthermore, limited flexibility provides the greatest benefits when configured to chain products and resources together to the greatest extent possible.

For realistic settings, these principles provide helpful guidelines for exploring flexibility options. We still want to build as long a chain as possible and to create a circuit that connects the primary products and resources together. In addition we try to equalize the load or expected demand that is assigned to each resource, and try to connect each product to its share of capacity. With these considerations, limited flexibility deployed to create chains provides most of the benefits of total flexibility.