THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Cross-Training with Imperfect Training Schemes

Citation for published version:
Buke, B, Araz, O & Fowler, J 2016, 'Cross-Training with Imperfect Training Schemes' Production and Operations Management, vol. 25, no. 7, pp. 1216-1231. DOI: 10.1111/poms.12543

Digital Object Identifier (DOI):
10.1111/poms.12543

Cross-Training with Imperfect Training Schemes

Burak Büke

School of Mathematics, The University of Edinburgh,

Edinburgh, Scotland, UK EH9 3FD

Phone:+44-131-650-5840

Fax:+44-131-650-6553

E-mail: B.Buke@ed.ac.uk

Özgür M. Araz

Department of Management, College of Business Administration,

University of Nebraska Lincoln,

Lincoln, NE 68588,

Phone:+1-402-472-2866

Fax:+1-402-472-5855

E-mail: oaraz2@unl.edu

John W. Fowler

W. P. Carey School of Business, Arizona State University,

Tempe, AZ 85287

Phone:+1-480-965-3727

Fax:+1-480-965-8629

E-mail: john.fowler@asu.edu

Cross-Training with Imperfect Training Schemes

Abstract

Cross-training workers is one of the most efficient ways of achieving flexibility in manufacturing and service systems for increasing responsiveness to demand variability. However, it is generally the case that cross-trained employees are not as productive on a specific task as employees who were originally trained for that task. Also, the productivity of the cross-trained workers depends on when they are cross-trained. In this work, we consider a two-stage model to analyze the effects of variations in productivity levels on cross-training policies. We define a new metric called *achievable capacity* and show that it plays a key role in determining the structure of the problem. If cross-training can be done in a consistent manner, the achievable capacity is not affected by when the training is done which implies that the cross-training decisions are independent of the opportunity cost of lost demand and are based on a trade-off between cross-training costs at different times. When the productivities of workers trained at different times differ, there is a three-way trade-off between cross-training costs at different times and the opportunity cost of lost demand due to lost achievable capacity. We analyze the effects of variability and show that if the productivity levels of workers trained at different times are consistent, the decision maker is inclined to defer the cross-training decisions as the variability of demand or productivity levels increases. However, when the productivities of workers trained at different times differ, an increase in the variability may make investing more in cross-training earlier more preferable.

Keywords: cross-training, flexibility, newsvendor networks, productivity factors

History: Received: January 2013; Accepted: November 2015 by Michael Pinedo, after 2 revisions.

1 Introduction

Designing flexible systems is one of the key strategies to increase responsiveness to variability in the market without sacrificing efficiency of the system. One way to achieve flexibility in a system is to cross-train workers on several processes. Cross-training has been proven to be highly beneficial in many different business environments, including but not limited to the semiconductor and automative industries, call centers and healthcare. For example, in the semiconductor industry, machine operators are often cross-trained to run more than one type of sophisticated equipment and technicians are often cross-trained to maintain more than one type of machine. In addition to increasing efficiency, cross-training can help keep budgets low, increase a company's ability to pay more to the employees, reduce turnover rate, and increase quality due to the workers' ability to react to unexpected changes (see e.g., Lyons (1992), McCune (1994), Iravani et al. (2007)).

Cross-trained workers can be shifted to work on new tasks when needed, which yields a more efficient usage of the resources. However, it is generally the case that cross-trained employees do not perform equally well on a specific task as employees who were originally trained for that task, i.e., the training schemes may be imperfect. Moreover, the productivity levels of the cross-trained workers may depend on when the cross-training is done. In this paper, our main goal is to analyze how these imperfections in training schemes affect cross-training decisions. To analyze the effect of imperfect schemes and timing of cross-training decisions, we consider a two-stage model and study the problem in a newsvendor network setting, introduced by Van Mieghem (1998) and Van Mieghem and Rudi (2002). Similar to the prior work, the decision maker decides on the number of employees to cross-train, i.e., the level of flexibility, before realizing the demand, which we refer to as offline cross-training. In the prior work, the first stage decisions are structural (design) decisions where the level of flexibility is fixed and does not change in the future. In practice, the decision maker may see that additional cross-training is beneficial after the demand is revealed and may wish to cross-train workers online as the demand is observed. However, the productivities of the employees who are cross-trained before and after the demand is observed may differ. If the demand exceeds the capacity even after the online cross-training, the excess demand is lost and an opportunity cost is incurred. Once the demand is revealed there is a capacity requirement for each task to fulfill the demand. With a slight abuse of terminology we refer to the capacity requirement for each task as the demand for the task.

We first consider the case where online cross-training is not profitable either because it is too expensive or the workers cannot be cross-trained online effectively. We provide a newsvendor-type equation which states the necessary and sufficient conditions that the optimal offline cross-training levels satisfy, and helps us to quantify when it is beneficial to invest in cross-training. Then we consider what happens if we are able to increase the effectiveness of training schemes. One possible argument is that as the training schemes become more effective, it creates an incentive to invest more in offline cross-training to exploit this positive change. On the other hand, one might also argue that with the increased effectiveness, the same or even slightly decreased level of cross-trained workers may be enough to hedge against possible excess demand. We show that both are reasonable outcomes of increasing the effectiveness and provide a condition to decide which argument is applicable. We also study how the optimal offline cross-training policies are affected by the demand variability. If cross-training costs and opportunity cost of lost demand are comparable, we show that offline cross-training gets less beneficial as variability increases, which is consistent with the literature.

Our main focus is the analysis of the trade-off between offline and online cross-training levels in the presence of imperfect training schemes. We define the metric *achievable capacity* for a task as the expected maximum demand that can be satisfied from the workforce with the skills available for this task and all the idle workers from the other task are cross-trained. Our main conclusion is that achievable capacity plays a key role in determining the structure of the cross-training problem. Our results can be summarized as follows:

(i) If the offline and online training schemes are consistent, i.e., their productivity factors are equal, then the achievable capacity does not depend on when the training is done, and the opportunity cost of lost demand becomes a fixed cost with respect to the cross-training levels. Hence, the cross-training decisions are independent of the opportunity cost of lost demand and the decision maker should only consider the trade off between the online and offline cross-training costs to decide on the cross-training levels. However, when the offline and online training schemes are not consistent, especially when the online cross-training is less effective than offline, then some achievable capacity is lost if cross-training is postponed. Then, the decision maker needs to consider a three-way trade off between cross-training costs of offline and online training schemes and the opportunity cost due to lost achievable capacity.

(ii) When only two tasks are under consideration, the cross-training problem is separable and the decision maker can set the offline cross-training levels separately for each task. This implies that the structure of the two task cross-training problem is very similar to the problem of dual sourcing under random yields where the decision maker aims to satisfy the demand from two different sources (offline and online cross-training) whose productivities are random (see e.g. Wang et al. (2010)). The main feature differentiating the two task cross-training problem is that the capacities of two sources are dependent through a total capacity constraint.

(iii) When the training schemes are consistent over time, the decision maker postpones cross-training, i.e., decreases the level of offline cross-training and the total cost increases, as the variability of random parameters increases. This result is consistent with the postponement literature (e.g. Feitzinger and Lee (1999)). However, when the productivity factors for offline and online cross-training schemes differ, then this result can be reversed. In this case, it may be beneficial to invest more in offline cross-training to increase achievable capacity as demand variability increases, in order to avoid the possibility of incurring high opportunity cost.

(iv) If the variability of the online productivity factor increases while making sure that the online cross-training is always profitable, we show that the total cost also increases. However, if the variability of the online productivity factor increases beyond a threshold such that for some scenarios online cross-training is not profitable, then the total cost might actually decrease as the variability of online productivity increases.

(v) We also study how any improvement on training schemes affects cross-training decisions. We show that if we can improve both training schemes in a way such that the difference between productivity factors decreases or at least is kept constant, then the decision maker postpones cross-training. However, if the offline training schemes are improved more than the online schemes, investing more on offline cross-training is better as it also increases the achievable capacity.

The opportunity to benefit from flexibility without too much investment has recently accelerated research in designing efficient flexible systems and we see this work as a part of this research stream. In their seminal paper, Jordan and Graves (1995) show that almost all the benefits of a fully flexible system, where all resources can perform all tasks, can be achieved by using a moderate level of flexibility. Their results demonstrate that using a special flexibility configuration referred to as "chaining" and under certain assumptions on demand, it is possible to obtain 98% of the throughput of a fully flexible system using resources that can perform only two different tasks. Using tools from queueing theory, Jordan et al. (2004) observe that cross-training can adversely affect performance if a poor control policy is used and demonstrate that a complete chain is robust with respect to the control policy and parameter uncertainty. In the literature, the ability of companies to achieve flexibility and efficiency while at the same time meeting customer needs is sometimes also refered as production agility (Gel et al. (2007); Hopp et al. (2004); Hopp and Van Oyen (2004)). Hopp and Van Oyen (2004) develop a framework for workforce cross-training, provide a comprehensive review of the recent literature and suggest some future research directions. Hopp et al. (2004) and Gel et al. (2007) analyze flexibility decisions for manufacturing systems operating under CONWIP or WIP-constrained policies and conclude that a cross-trained worker should perform her original task before helping on other tasks. Pinker and Shumsky (2000) perform numerical studies to analyze the trade off between efficiency and quality due to cross-training. Netessine et al. (2002) show how cross-training policies are affected by demand correlation. Davis et al. (2009) indicate that under high workload imbalances, an extensive level of cross training is required to significantly improve the overall production performance. In a service environment, Gnanlet and Wendell (2009) use a two-stage stochastic programming model to determine optimal resource levels and demonstrate the benefits of cross-training activities in a health care setting. In their recent papers, Bassamboo et al. (2012, 2010) define level-k resources to be the resources that are able to process k different tasks. In Bassamboo et al. (2012), they prove that for symmetric queueing systems one only need to use dedicated resources and level-2 resources. Similar to our analysis, Bassamboo et al. (2010) use the newsvendor network framework to prove that in the optimal flexibility configurations only two adjacent levels of flexibility are needed. Chou et al. (2010) discuss the effect of production efficiency comparing full flexibility with a chaining structure. Another paper which is closely related to our work is Chakravarthy and Agnihothri (2005), where they study the optimum fraction of flexible servers for a two task problem with perfect training schemes. They point to the fact that cross-trained workers may not be as efficient as dedicated workers. However, they do not provide an analysis of the problem. There have been several attempts to formulate the problem as a mathematical program (see e.g. Brusco and Johns (1998), Walsh et al. (2000) and Tanrisever et al. (2012)). In this paper, our primary focus is on getting managerial insights for aggregate planning of cross-training efforts.

2 A Two-Stage Model for Cross-Training

In this section, our goal is to provide a detailed analysis of the cross-training problem for two tasks when the cross-training schemes are imperfect. We analyze the problem of cross-training workers between two tasks, α and γ. We assume that the capacity is measured in time units and initially there is x_α^0 and x_γ^0 units of capacity dedicated to process tasks α and γ, respectively. The decision maker has to develop an aggregate workforce plan by cross-training some of the available workers offline before observing the demand $\tilde{d}_\alpha$ and $\tilde{d}_\gamma$. Before actual demand is realized, it costs c_α^1 to train one unit of dedicated capacity of γ to work on task α. The cross-trained workers can still work on their original task γ with full efficiency. However, they are not as efficient in their new skill α, and their capacity needs to be adjusted by a productivity factor $\tilde{\delta}_\alpha^1$, where $0 < \tilde{\delta}_\alpha^1 \leq 1$; i.e., if a cross-trained γ-worker spends one hour working on task α, it is equivalant to $\tilde{\delta}_\alpha^1$ hours of an original α-worker.

The productivity factor, $\tilde{\delta}^1_\alpha$, can also be perceived as an indicator of the effectiveness of the training program and is assumed to be random. The random parameters, offline and online productivity factors and the capacity requirements for each task, are revealed after the offline cross-training decisions are made. After the capacity requirements for tasks are revealed, we first use the dedicated workers and workers cross-trained offline in the first stage to satisfy the demand. If the available capacity for task α is not enough to satisfy the requirement and there is excess capacity for task γ, additional cross-training can be performed online at a unit cost of c^2_α. The productivity factor for the workers cross-trained in the second stage is $\tilde{\delta}^2_\alpha$, where $0 < \tilde{\delta}^2_\alpha \leq 1$. Unless otherwise stated we assume that $\tilde{\delta}^2_\alpha \leq \tilde{\delta}^1_\alpha$ with probability one (w. p. 1). If the workforce at hand cannot satisfy the capacity requirements even after the second stage cross-training, the demand is lost incurring a unit opportunity cost of h_α. A similar mechanism works to satisfy the demand for γ interchanging the subscripts.

Without loss of generality, we assume that $h_\alpha \leq h_\gamma$. When workers are cross-trained to work on both tasks, a natural question is on which task they are allocated when they are needed for both. In this paper, we assume the following:

Assumption 2.1. *It is always preferable to use cross-trained workers on their original tasks rather than their new tasks when they are needed for both, i.e.,* $h_\alpha \geq \tilde{\delta}^1_\gamma h_\gamma$ *holds w. p. 1.*

To simplify our analysis and notation, we also assume that random variables $\tilde{d}_\alpha, \tilde{d}_\gamma, \tilde{\delta}^1 = (\tilde{\delta}^1_\alpha, \tilde{\delta}^1_\gamma)$ and $\tilde{\delta}^2 = (\tilde{\delta}^2_\alpha, \tilde{\delta}^2_\gamma)$ are continuous with joint density function $f(\tilde{d}_\alpha, \tilde{d}_\gamma, \tilde{\delta}^1, \tilde{\delta}^2)$ and use $\tilde{\xi} = (\tilde{d}_\alpha, \tilde{d}_\gamma, \tilde{\delta}^1, \tilde{\delta}^2)$, whenever we do not need to address specific random variables. Throughout this work, a random variable x is denoted $\tilde{x}$, $\mathbb{E}[\tilde{x}]$ denotes the expected value of the random variable. Similarly, we use $\mathbb{E}[\tilde{x};\Omega] = \int_\Omega x d\mathbb{P}(x)$ to denote the expectation over a scenario region Ω. We present proofs of our propositions given below in Appendix A to allow for a better readability.

We need to write out the objective function explicitly based on the mechanism described above. The decision maker initially decides on x^1_α and x^1_γ, which are the amount of workforce cross-trained offline to work on α and γ from the dedicated capacity of γ and α, respectively. We can decompose the cost function $g(x^1_\alpha, x^1_\gamma, \tilde{\xi})$ into the first stage cost which is incurred

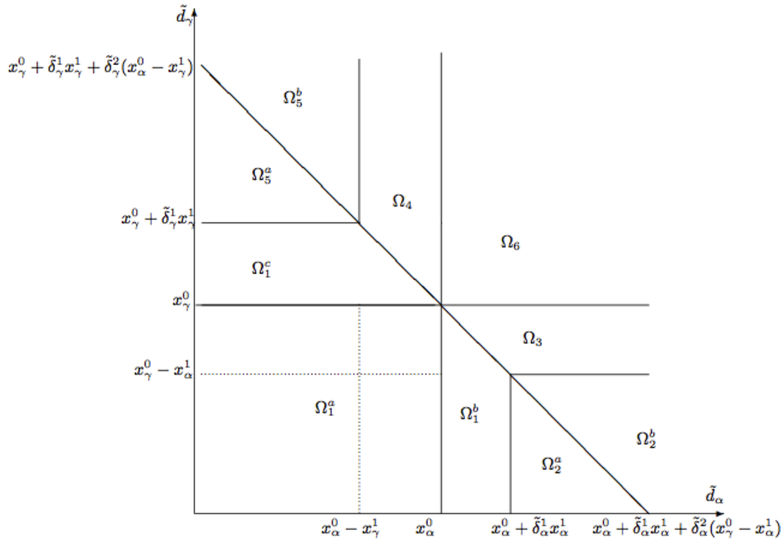


Figure 1: Partitioning the support of the demand vector

due to offline cross-training and the second stage cost $v(x^1_\alpha, x^1_\gamma, \tilde{\xi})$ which is revealed after the realization of random parameters as

$$\min_{x^1_\alpha, x^1_\gamma} \mathbb{E}[g(x^1_\alpha, x^1_\gamma, \tilde{\xi})] = \min_{x^1_\alpha, x^1_\gamma} c^1_\alpha x^1_\alpha + c^1_\gamma x^1_\gamma + \mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi})]. \tag{1}$$

The second stage cost, $v(x^1_\alpha, x^1_\gamma, \tilde{\xi})$, depends on whether new cross-training is needed after the demand is observed. Hence, $v(x^1_\alpha, x^1_\gamma, \tilde{\xi})$ takes different forms depending on the realization of random parameters. To analyze this function further and calculate the expected value for given x^1_α, x^1_γ, we first use the tower property $\mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi})] = \mathbb{E}[\mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]]$. To calculate the conditional expectation $\mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]$, we first partition the support of $(\tilde{\delta}^1, \tilde{\delta}^2)$ into subsets, where in each subset the nature of the second stage decision is different. Then, we partition the support of $\tilde{d}_\alpha$ and $\tilde{d}_\gamma$ so that the function $v(x^1_\alpha, x^1_\gamma, \tilde{\xi})$ has a single form within a partition. This partitioning scheme is explained below and the most general graphical representation is given in Figure 1.

Under Assumption 2.1, task α demand will be lost only when the demand $\tilde{d}_\alpha$ cannot be supplied by using the initial capacity x^0_α and cross-trained workers who are not allocated to task γ. Since $h_\alpha \leq h_\gamma$, the same claim is always true for task γ. If we choose to cross-train workers online in the second stage, we need to spend $c^2_\alpha/\tilde{\delta}^2_\alpha$ to satisfy unit demand of task α, or we lose the demand and incur a cost of h_α. Hence, for task α, we choose to resort to online cross-training first if $c_\alpha^2 \leq \tilde{\delta}_\alpha^2 h_\alpha$ and we never cross-train online in the second stage if $c_\alpha^2 > \tilde{\delta}_\alpha^2 h_\alpha$. Similar logic applies to task γ.

2.1 Case 1: Online cross-training is not profitable

We first consider the situation where losing the excess demand is preferable over online cross-training in the second stage for both tasks, i.e., $c_\alpha^2 > \tilde{\delta}_\alpha^2 h_\alpha$ and $c_\gamma^2 > \tilde{\delta}_\gamma^2 h_\gamma$ w. p. 1. Under this assumption we can use the following partitioning to explicitly state $v(x_\alpha^1, x_\gamma^1, \tilde{\xi})$ for any given x_α^0, x_γ^0 and $\tilde{\xi}$.

1. $\Omega_1^a = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : \tilde{d}_\alpha \leq x_\alpha^0, \tilde{d}_\gamma \leq x_\gamma^0\}$. For the scenarios in Ω_1^a, the initial workforce is enough to satisfy the capacity requirements. Hence, $v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = 0$ on Ω_1^a.

2. $\Omega_1^b = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 < \tilde{d}_\alpha \leq \min\{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1, x_\alpha^0 + \tilde{\delta}_\alpha^1 (x_\gamma^0 - \tilde{d}_\gamma)\}, \tilde{d}_\gamma \leq x_\gamma^0\}$. On Ω_1^b, the initial workforce x_α^0 cannot satisfy $\tilde{d}_\alpha$. Task γ may need to use some workers who are cross-trained offline to work on α, but the remaining cross-trained workforce is enough to satisfy the excess demand for α. Hence, again $v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = 0$ on Ω_1^b.

3. Ω_1^c is defined similar to Ω_1^b with α and γ interchanged, and $v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = 0$ on Ω_1^c.

   We define $\Omega_1 = \Omega_1^a \cup \Omega_1^b \cup \Omega_1^c$. When the demand falls in this region, no recourse action is needed in the second stage and hence no cost is incurred.

4. $\Omega_2 = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 < \tilde{d}_\alpha, 0 \leq \tilde{d}_\gamma \leq x_\gamma^0 - x_\alpha^1\}(= \Omega_2^a \cup \Omega_2^b$ in Figure 1). The demand for γ is low so that all the cross-trained workers can be used to work on task α. However, even this is not enough to satisfy $\tilde{d}_\alpha$ and the excess demand is lost. Hence, $v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = h_\alpha(\tilde{d}_\alpha - x_\alpha^0 - \tilde{\delta}_\alpha^1 x_\alpha^1)$ on Ω_2.

5. $\Omega_3 = \{(\tilde{d}_\alpha, \tilde{d}_\gamma), x_\alpha^0 + \tilde{\delta}_\alpha^1 (x_\gamma^0 - \tilde{d}_\gamma) < \tilde{d}_\alpha, x_\gamma^0 - x_\alpha^1 < \tilde{d}_\gamma \leq x_\gamma^0\}$. For scenarios in this subset of the support, some but not all of the workers who are cross-trained to work on task α can be shifted to α, and this is not enough to satisfy the capacity requirement. The excess demand is lost and hence, $v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = h_\alpha(\tilde{d}_\alpha - (x_\alpha^0 + \tilde{\delta}_\alpha^1 (x_\gamma^0 - \tilde{d}_\gamma)))$ on Ω_3.

6. $\Omega_4 = \{(\tilde{d}_\alpha, \tilde{d}_\gamma), x^0_\alpha - x^1_\gamma < \tilde{d}_\alpha \leq x^0_\alpha, x^0_\gamma + \tilde{\delta}^1_\gamma(x^0_\alpha - \tilde{d}_\alpha) < \tilde{d}_\gamma\}(= \Omega^a_4 \cup \Omega^b_4$ in Figure 1). This region is defined in the same way as Ω_3 with α and γ interchanged. The second stage cost function is given by $v(x^1_\alpha, x^1_\gamma, \tilde{\xi}) = h_\gamma(\tilde{d}_\gamma - (x^0_\gamma + \tilde{\delta}^1_\gamma(x^0_\alpha - \tilde{d}_\alpha)))$.

7. $\Omega_5 = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : 0 \leq \tilde{d}_\alpha \leq x^0_\alpha - x^1_\gamma, x^0_\gamma + \tilde{\delta}^1_\gamma x^1_\gamma\}(= \Omega^a_5 \cup \Omega^b_5$ in Figure 1). Ω_5 is the analog of Ω_2 with α and γ interchanged. Hence, $v(x^1_\alpha, x^1_\gamma, \tilde{\xi}) = h_\gamma(\tilde{d}_\gamma - x^0_\gamma - \tilde{\delta}^1_\gamma x^1_\gamma)$ on Ω_5.

8. $\Omega_6 = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x^0_\alpha < \tilde{d}_\alpha, x^0_\gamma < \tilde{d}_\gamma\}(= \Omega^a_6 \cup \Omega^b_6$ in Figure 1). In this case, the initial capacity is not enough to satisfy the demand for either task. Hence, excess demand is lost for both tasks. The second stage cost is $v(x^1_\alpha, x^1_\gamma, \tilde{\xi}) = h_\alpha(\tilde{d}_\alpha - x^0_\alpha) + h_\gamma(\tilde{d}_\gamma - x^0_\gamma)$.

After characterizing the cost function completely, we now focus on analyzing the optimal offline cross-training levels. Our first result states a necessary and sufficient condition that should be satisfied by the optimal cross-training levels. The sets, Ω_is, depend on the cross-training levels x^1_α or x^1_γ and to emphasize the dependency we adopt the notation $\Omega_i(x^1_\alpha)$ or $\Omega_i(x^1_\gamma)$ as appropriate in the equations below.

Proposition 2.1. *Any ${x^1_\alpha}^* \in [0, x^0_\gamma]$ that satisfies the equation*

$$\mathbb{E}[\tilde{\delta}^1_\alpha; \Omega_2(x^{1*}_\alpha)] = \frac{c^1_\alpha}{h_\alpha}, \tag{2}$$

is an optimal offline cross-training level for α. If there does not exist any ${x^1_\alpha}^* \in [0, x^0_\gamma]$ that satisfies (2), then either $\mathbb{E}[\tilde{\delta}^1_\alpha; \Omega_2(x^0_\gamma)] \geq \frac{c^1_\alpha}{h_\alpha}$ or $\mathbb{E}[\tilde{\delta}^1_\alpha; \Omega_2(0)] \leq \frac{c^1_\alpha}{h_\alpha}$ and the optimal offline cross-training level for α is ${x^1_\alpha}^* = x^0_\gamma$ or ${x^1_\alpha}^* = 0$, respectively. The same result holds for ${x^1_\gamma}^*$ when α and γ are interchanged above.

Condition (2) states that under the optimal cross-training policy, the expectation of the first-stage productivity factor over the scenarios where the cross-trained workers are utilized for their new tasks should be equal to the ratio of the first-stage cross-training and the opportunity costs. An important feature of Proposition 2.1 is that it indicates the problem is separable, i.e., the optimal cross-training levels for tasks α and γ can be decided separately. Hence, we only state results relating to task α below and similar results hold for

γ by interchanging the subscripts. Also, even though Proposition 2.1 assumes continuous distributions, the results can be extended to a discrete setting using a similar methodology to the newsvendor problem. An immediate consequence of Proposition 2.1 is as follows.

Corollary 2.1. *When the first-stage productivity factor $\tilde{\delta}^1_\alpha$ is deterministically equal to δ^1_α, then any ${x^1_\alpha}^* \in [0, x^0_\gamma]$ that satisfies the newsvendor-type equation*

$$\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^1_\alpha \leq \tilde{d}_\alpha, x^0_\gamma - x^1_\alpha \geq \tilde{d}_\gamma) = \frac{c^1_\alpha}{\delta^1_\alpha h_\alpha} \tag{3}$$

solves the cross-training problem. If there does not exist any ${x^1_\alpha}^* \in [0, x^0_\gamma]$ that satisfies (3), then, either $\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^1_\alpha \leq \tilde{d}_\alpha, x^0_\gamma - x^1_\alpha \geq \tilde{d}_\gamma) \geq \frac{c^1_\alpha}{\delta^1_\alpha h_\alpha}$ or $\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^1_\alpha \leq \tilde{d}_\alpha, x^0_\gamma - x^1_\alpha \geq \tilde{d}_\gamma) \leq \frac{c^1_\alpha}{\delta^1_\alpha h_\alpha}$ and the optimal cross-training level for α is $x^{1*}_\alpha = x^0_\gamma$ or $x^{1*}_\alpha = 0$, respectively.

Proposition 2.1 also helps us understand when it is not profitable to cross-train. This result is in accordance with the case with perfect cross-training schemes.

Corollary 2.2. *It is not profitable to cross-train offline for task α if the assumption of Proposition 2.1 holds and at least one of $\mathbb{E}[\tilde{\delta}^1_\alpha]$, $\mathbb{P}(x^0_\alpha \leq \tilde{d}_\alpha)$ or $\mathbb{P}(x^0_\gamma \leq \tilde{d}_\gamma)$ is less than $\frac{c^1_\alpha}{h_\alpha}$.*

Intuitively, one should not resort to cross-training if the effectiveness of training programs is not good enough. The first part of Corollary 2.2 quantifies how "good enough" should be interpreted. The second part says that before cross-training one should make sure that there is a solid chance that extra capacity will be needed. Even when extra capacity is needed, demand for the other task may make it impossible to utilize the cross-trained workers.

2.1.1 Independent Demands and Deterministic Productivity Factors

We now investigate the sensitivity of the offline cross-training levels to the changes in productivity factors when the productivity factors are deterministic and demand for different tasks are independent. If we are able to increase the effectiveness of our training policies, one may argue that it is better to exploit this further by increasing our offline cross-training level. On the other hand, another argument is that this increase in effectiveness may significantly reduce the risk of incurring opportunity cost of lost demand and we may not need to invest as much in offline cross-training. Proposition 2.2 provides a condition to decide which of these arguments is applicable.

Proposition 2.2. *Suppose that the offline productivity factor $\tilde{\delta}^1_\alpha$ is deterministically equal to δ^1_α and the demand for α and γ, $\tilde{d}_\alpha$ and $\tilde{d}_\gamma$ are independent. Then, the optimal offline cross-training level ${x^1_\alpha}^*$ for task α increases as productivity factor δ^1_α increases if and only if*

$$H_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha) = \frac{f_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)}{\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha)} \leq \frac{1}{\delta^1_\alpha x^{1*}_\alpha}, \tag{4}$$

where $f_{\tilde{d}_\alpha}(\cdot)$ and $H_{\tilde{d}_\alpha}(\cdot)$ are the marginal probability density and hazard rate functions of task α demand, $\tilde{d}_\alpha$, respectively.

To decide whether it is better to cross-train more than x^{1*}_α, we need to concentrate on the scenarios where the demand exceeds $x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha$. The definition of the hazard rate function implies that $\mathbb{P}(\tilde{d}_\alpha \leq x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha + \epsilon | \tilde{d}_\alpha > x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha) \approx H_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)\epsilon$. Hence, if $H_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)$ is high, even when the demand exceeds $x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha$, the probability that it exceeds by a large margin is low, in which case a small increase in the productivity factor is enough to satisfy the exceeding demand and additional cross-training is not beneficial. However, if $H_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)$ is low, when the demand exceeds $x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha$ most probably it exceeds by a large margin and an increase in the productivity factor makes additional cross-training more attractive.

Under some mild conditions, we can use Proposition 2.1 to infer how the variability of the demand for different tasks affects the cross-training decisions.

Proposition 2.3. *Suppose $\tilde{d}_1$ and $\tilde{d}_2$ are symmetric random variables around zero, i.e., $\mathbb{P}(\tilde{d}_i < -x) = \mathbb{P}(\tilde{d}_i > x)$ for all values of x and $i = 1, 2$. If $\tilde{\delta}^1_\alpha$ is deterministically equal to δ^1_α, $2c^1_\alpha \geq \delta^\alpha_1 h_\alpha$, the demands for different tasks, $\tilde{d}_\alpha$ and $\tilde{d}_\gamma$, are independent, continuous and can be written as $\tilde{d}_\alpha = m\tilde{d}_1 + \mu_1$ and $\tilde{d}_\gamma = n\tilde{d}_2 + \mu_2$, then the optimal cross-training level x^{1*}_α is non-increasing in both m and n.*

Proposition 2.3 essentially states that if demands for tasks are independent and follow symmetric distributions, e.g. a normal or uniform distributon, and if the costs and productivity factors satisfy the conditions above, the optimal cross-training level for task α is decreasing with the variances of demands. Unfortunately, when $\tilde{\delta}_\alpha^1$ is not deterministic, the monotonicity result may not hold. Some counterexamples are presented in Section 3.

2.2 Case 2: Profitable online cross-training in the second stage

Now, we consider the situation where it is profitable to cross-train online in the second stage, i.e., $c_\alpha^2 < \tilde{\delta}_\alpha^2 h_\alpha$ and $c_\gamma^2 < \tilde{\delta}_\gamma^2 h_\gamma$ w. p. 1. The demand is lost only when the available workforce is not able to satisfy the demand even after all the idle workers are cross-trained online in the second stage. Hence, for scenarios in $\Omega_1, \Omega_3, \Omega_4$ and Ω_6 the cost function is as in Section 2.1 and we only need to consider Ω_2 and Ω_5 further.

1. $\Omega_2^a = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 < \tilde{d}_\alpha \leq x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma), \tilde{d}_\gamma < x_\gamma^0 - x_\alpha^1\}$. If $(\tilde{d}_\alpha, \tilde{d}_\gamma) \in \Omega_2^a$, both the initial workforce and the cross-trained workforce are used in performing task α. If $(\tilde{d}_\alpha - x_\alpha^0 - \tilde{\delta}_\alpha^1 x_\alpha^1)/\tilde{\delta}_\alpha^2$ units of the workforce are cross-trained to work on α, the remaining demand can be satisfied. Hence, the second stage cross-training cost is
$$v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = c_\alpha^2 \frac{\tilde{d}_\alpha - x_\alpha^0 - \tilde{\delta}_\alpha^1 x_\alpha^1}{\tilde{\delta}_\alpha^2}.$$
2. $\Omega_2^b = \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma) < \tilde{d}_\alpha, \tilde{d}_\gamma < x_\gamma^0 - x_\alpha^1\}$. For the scenarios in Ω_2^b, it is not possible to satisfy the demand for α even after all the cross-trained workforce work on task α. The decision maker cross-trains the idle workforce of γ and then excess demand is lost. Hence, over Ω_2^b
$$v(x_\alpha^1, x_\gamma^1, \tilde{\xi}) = c_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma) + h_\alpha (\tilde{d}_\alpha - x_\alpha^0 - \tilde{\delta}_\alpha^1 x_\alpha^1 - \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma)).$$

Ω_5^a and Ω_5^b are defined similarly interchanging the subscripts.

Now we formally define achievable capacity of a task which plays a major role in determining the structure of the cross-training problem. For simplicity, we state the definition specifically for task α.

Definition 2.1. *The achievable capacity of task α under first-stage cross-training level x_α^1, $C_\alpha(x_\alpha^1)$, is the expected value of the maximum demand for task α that can be satisfied after* *the randomness is realized and all idle workers for task γ are cross-trained, i.e., defining notation* $x^+ = \max\{x, 0\}$, $C_\alpha(x^1_\alpha) = x^0_\alpha + \mathbb{E}[\tilde{\delta}^1_\alpha \min\{x^1_\alpha, (x^0_\gamma - \tilde{d}_\gamma)^+\} + \tilde{\delta}^2_\alpha(x^0_\gamma - x^1_\alpha - d_\gamma)^+]$.

When $\tilde{\delta}^1_\alpha = \tilde{\delta}^2_\alpha = \tilde{\delta}$ w. p. 1, we get $C_\alpha(x^1_\alpha) = x^0_\alpha + \mathbb{E}[\tilde{\delta}(x^0_\gamma - d_\gamma)^+]$ and the achievable capacity is independent of the first-stage cross-training level. However, when $\mathbb{P}(\tilde{\delta}^1_\alpha > \tilde{\delta}^2_\alpha) > 0$, the demand that can be satisfied without incuring opportunity cost increases as the first-stage cross-training increases.

Similar to Proposition 2.1, we now state a necessary and sufficient condition that should be satisfied by the optimal offline cross-training level. This condition suggests a three-way trade off between the offline and online cross-training costs and the opportunity cost of lost demand due to achievable capacity lost by delaying the training.

Proposition 2.4. *Any solution* ${x^1_\alpha}^* \in [0, x^0_\gamma]$ *that satisfies the equation*

$$\mathbb{E}[\frac{c^2_\alpha \tilde{\delta}^1_\alpha}{\tilde{\delta}^2_\alpha}; \Omega^a_2(x^{1*}_\alpha)] + \mathbb{E}[c^2_\alpha + h_\alpha(\tilde{\delta}^1_\alpha - \tilde{\delta}^2_\alpha); \Omega^b_2(x^{1*}_\alpha)] = c^1_\alpha \tag{5}$$

is an optimal cross-training level for α. If there does not exist any ${x^1_\alpha}^* \in [0, x^0_\gamma]$ *that satisfies (5), then, either* $\mathbb{E}[\frac{c^2_\alpha \tilde{\delta}^1_\alpha}{\tilde{\delta}^2_\alpha}; \Omega^a_2(x^0_\gamma)] + \mathbb{E}[c^2_\alpha + h_\alpha(\tilde{\delta}^1_\alpha - \tilde{\delta}^2_\alpha); \Omega^b_2(x^0_\gamma)] \geq c^1_\alpha$ *or* $\mathbb{E}[\frac{c^2_\alpha \tilde{\delta}^1_\alpha}{\tilde{\delta}^2_\alpha}; \Omega^a_2(0)] + \mathbb{E}[c^2_\alpha + h_\alpha(\tilde{\delta}^1_\alpha - \tilde{\delta}^2_\alpha); \Omega^b_2(0)] \leq c^1_\alpha$ *and the optimal cross-training level for α is* $x^{1*}_\alpha = x^0_\gamma$ *or* $x^{1*}_\alpha = 0$, *respectively. The same result holds for task γ with α and γ interchanged above.*

The opportunity cost plays a role only for the scenarios in Ω^b_2, i.e., for the scenarios where the online cross-training is needed. To understand this further, suppose we defer training one unit of workforce to the second stage. For the scenarios where this additional workforce is not needed for task α or is needed for task γ, we cannot use this additional workforce on task α, then the opportunity cost does not depend on whether this workforce is cross-trained or not offline. If this workforce is not needed for task γ and is needed for task α, we lose $\tilde{\delta}^1_\alpha - \tilde{\delta}^2_\alpha$ units of achievable capacity by postponing cross-training from offline to online. This indicates that the marginal decrease in achievable capacity by cross-training one unit of workforce online instead of offline is $\mathbb{E}[(\tilde{\delta}^1_\alpha - \tilde{\delta}^2_\alpha); \Omega^b_2(x^{1*}_\alpha)]$. Proposition 2.4 suggests that only the opportunity cost for this lost capacity affects our decisions. If the training effectiveness in both stages is the same, then essentially we do not lose any achievable capacity and hence the opportunity cost does not play any role in our decisions. Corollary 2.3 below summarizes these observations.

Corollary 2.3. *If the training effectiveness for programs before and after demand is realized are the same, i.e.,* $\tilde{\delta}^1_\alpha = \tilde{\delta}^2_\alpha$ *w. p. 1, then optimal cross-training decisions do not depend on the opportunity cost* h_α *and any solution* ${x^1_\alpha}^* \in [0, x^0_\gamma]$ *that satisfies the newsvendor-type equation*

$$\mathbb{P}(\tilde{d}_\alpha \geq x^0_\alpha + \tilde{\delta}^1_\alpha {x^1_\alpha}^*, \tilde{d}_\gamma \leq x^0_\gamma - {x^1_\alpha}^*) = \frac{c^1_\alpha}{c^2_\alpha} \tag{6}$$

is an optimal cross-training level for α*. If there does not exist any* ${x^1_\alpha}^* \in [0, x^0_\gamma]$ *that satisfies (6), then, either* $\mathbb{P}(\tilde{d}_\alpha \geq x^0_\alpha + \tilde{\delta}^1_\alpha x^0_\gamma, \tilde{d}_\gamma = 0) \geq \frac{c^1_\alpha}{c^2_\alpha}$ *or* $\mathbb{P}(\tilde{d}_\alpha \geq x^0_\alpha, \tilde{d}_\gamma \leq x^0_\gamma) \leq \frac{c^1_\alpha}{c^2_\alpha}$ *and the optimal cross-training level for* α *is* $x^{1*}_\alpha = x^0_\gamma$ *or* $x^{1*}_\alpha = 0$*, respectively. The same result holds for task* γ *with* α *and* γ *interchanged above.*

2.2.1 Independent Demands and Deterministic Productivity Factors

Another important question is how the offline cross-training depends on the variability in demand. One might expect that as the variance of demand increases, it should become more profitable to delay the cross-training. Proposition 2.5 proves that when $\tilde{\delta}^1_\alpha = \tilde{\delta}^2_\alpha$ w. p. 1, and under mild conditions on training costs and demand distributions, this is indeed true. The proof of Proposition 2.5 follows the same lines as in Proposition 2.3 and is omitted here.

Proposition 2.5. *Suppose* $\tilde{d}_1$ *and* $\tilde{d}_2$ *are symmetric random variables around zero, i.e.,* $\mathbb{P}(\tilde{d}_i < -x) = \mathbb{P}(\tilde{d}_i > x)$ *for all values of* x *and* $i = 1, 2$*. If* $\tilde{\delta}^1_\alpha = \tilde{\delta}^2_\alpha = \delta_\alpha$ *is deterministically,* $2c^1_\alpha \geq c^2_\alpha$*, the demand for different tasks,* $\tilde{d}_\alpha$ *and* $\tilde{d}_\gamma$*, are independent, continuous and can be written as* $\tilde{d}_\alpha = m\tilde{d}_1 + \mu_1$ *and* $\tilde{d}_\gamma = n\tilde{d}_2 + \mu_2$*, then the optimal cross-training level* x^{1*}_α *is non-increasing in both* m *and* n*.*

Obviously, Proposition 2.5 also covers the case with perfect training schemes where $\tilde{\delta}^1_\alpha = \tilde{\delta}^2_\alpha = 1$ w. p. 1. However, the situation is quite different when $\tilde{\delta}^1_\alpha \neq \tilde{\delta}^2_\alpha$. To demonstrate what may go wrong consider the following counter-example. Suppose that $\tilde{d}_\gamma = 1$ deterministically, and $\mathbb{P}(\tilde{d}_\alpha = 9) = \mathbb{P}(\tilde{d}_\alpha = 11) = 0.1$ and $\mathbb{P}(\tilde{d}_\alpha = 10) = 0.8$. Also assume that $x^0_\alpha = 10, x^0_\gamma = 3, c^1_\alpha = 3, c^2_\alpha = 4$ and the unit opportunity cost of lost demand is extremely high as $h_\alpha = 10000$. The given parameters satisfy the conditions of Proposition 2.5 except suppose we have $\tilde{\delta}^1_\alpha = 1 \neq \tilde{\delta}^2_\alpha = 0.5$ w. p. 1. Since there is a small probability that the cross-trained workers will be needed and it is possible to cross-train necessary workers in the second stage, it is optimal not to cross-train any workers offline in the first stage, i.e., $x^{1*}_\alpha = 0$, and if necessary cross-train two task γ workers online to work on α. Now consider the case where $\tilde{d}_\alpha$ follows the distribution with probability mass function $\mathbb{P}(\tilde{d}_\alpha = 8) = \mathbb{P}(\tilde{d}_\alpha = 12) = 0.1$ and $\mathbb{P}(\tilde{d}_\alpha = 10) = 0.8$. If the scenario $\tilde{d}_\alpha = 12$ reveals, one γ worker will be working to satisfy γ demand and even if the remaining two γ workers are cross-trained online their productivity will be equivalent to one original α worker. Hence, it is not possible to satisfy the demand by only online cross-training in the second stage . In order not to risk a high opportunity cost of lost demand, we need to cross-train two workers ($x^{1*}_\alpha = 2$). In this example, the fact that the productivity factors for offline and online cross-training are not equal plays the major role and forces us to invest more in offline cross-training as the variance increases in order not to lose achievable capacity.

When online cross-training is not profitable, Proposition 2.2 states a necessary and sufficient condition for offline cross-training levels to be increasing as we improve the training policies for offline cross-training. When online cross-training is profitable, any improvement we make on our training policies increases both offline and online productivity factors. Proposition 2.6 shows that if the training schemes are improved in such a way that the difference between offline and online productivity factors stays the same or decreases, then offline cross-training will be less attractive.

Proposition 2.6. *Consider two systems with deterministic productivity factors,* $(\bar{\delta}^1_\alpha, \bar{\delta}^2_\alpha)$ *and* $(\hat{\delta}^1_\alpha, \hat{\delta}^2_\alpha)$ *such that* $\bar{\delta}^i_\alpha \leq \hat{\delta}^i_\alpha$ *for* $i = 1, 2$ *and* $\bar{\delta}^1_\alpha - \bar{\delta}^2_\alpha \geq \hat{\delta}^1_\alpha - \hat{\delta}^2_\alpha$*, and suppose that* $\bar{x}^{1*}_\alpha$ *and* $\hat{x}^{1*}_\alpha$ *are the optimal offline cross-training levels respectively. Then,* $\bar{x}^{1*}_\alpha \geq \hat{x}^{1*}_\alpha$*.*

The case when $\bar{\delta}^1_\alpha - \bar{\delta}^2_\alpha > \hat{\delta}^1_\alpha - \hat{\delta}^2_\alpha$ implies that if the second stage productivity factor increases more than the first stage productivity factor, then it is better to postpone the cross-training to the second stage by reducing the offline cross-training. Proposition 2.6 concludes that even when both offline and online productivity factors are improved at the same level, it is better to reduce the offline cross-training. Because in this case, it is possible to cover more demand by cross-training less and the marginal savings in the opportunity cost by increasing the offline cross-training is always better when the productivity factors are low for any given scenario. However, when $\bar{\delta}_{\alpha}^{1} - \bar{\delta}_{\alpha}^{2} < \hat{\delta}_{\alpha}^{1} - \hat{\delta}_{\alpha}^{2}$, one can exploit the increase in the differences to reduce opportunity cost substantially by increasing offline cross-training.

2.2.2 The Effect of Variability in the Productivity Factors on Optimal Cross-Training Levels

Now, we investigate how the variability of the productivity factors affects the total cost. Propositions 2.7 and 2.8 state that the total cost of cross-training policies which is the sum of costs of offline cross-training, online cross-training and the opportunity cost of lost demand increases as the variability of the first and second stage productivity factor increase, respectively. Even though we are explicitly assuming that the second stage cross-training is profitable in this section, an equivalent of Propositon 2.7 can be proved under the assumptions of Section 2.1. However, as there is no second stage cross-training under the assumptions of Section 2.1, there is no equivalent version of Proposition 2.8.

Proposition 2.7. *Suppose $h_{\alpha} \geq \tilde{\delta}_{\gamma}^{1} h_{\gamma}$ w. p. 1, the second stage productivity factor is deterministically equal to δ_{α}^{2}, the first stage productivity factor $\tilde{\delta}_{\alpha}^{1}$ is independent of other parameters and can be expressed as $\tilde{\delta}_{\alpha}^{1} = \delta_{\alpha}^{1} + n\tilde{\Delta}$, where $\tilde{\Delta}$ is a random variable with mean 0 and n is chosen such that $\mathbb{P}(\tilde{\delta}_{\alpha}^{1} \leq 1) = 1$. Then, the total cost is an increasing function of n.*

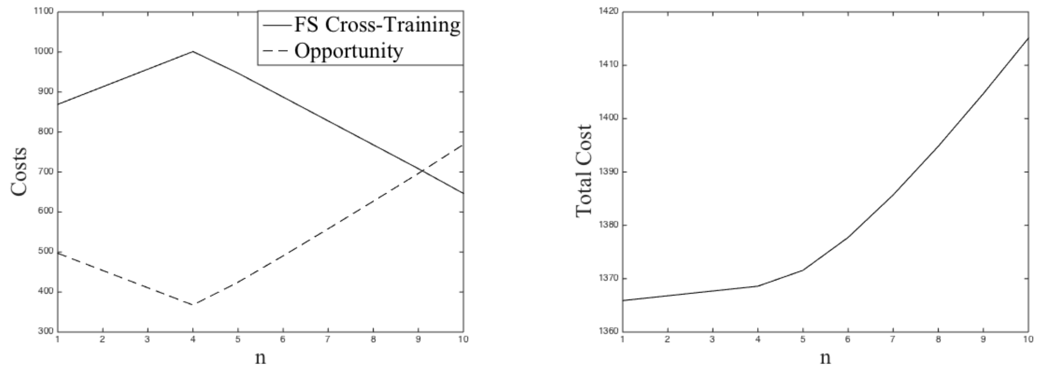
Now, we analyze how the variability of the second stage productivity factor affects the total cost of cross-training policies. Proposition 2.8 proves that the total cost is increasing with respect to the variance of $\tilde{\delta}_{\alpha}^{2}$ if we make sure that online cross-training is profitable for all realizations of the productivity factors. The proof follows the same lines as that of Propositon 2.7.

Proposition 2.8. *Suppose the first stage productivity factor is deterministically equal to δ_α^1 and $h_\alpha \geq \delta_\gamma^1 h_\gamma$. The second stage productivity factor $\tilde{\delta}_\alpha^2$ is independent of other parameters and can be expressed as $\tilde{\delta}_\alpha^2 = \delta_\alpha^2 + n\tilde{\Delta}$, where $\tilde{\Delta}$ is a random variable with mean 0 and n is chosen such that $\mathbb{P}(\tilde{\delta}_\alpha^2 \in [c_\alpha^2/h_\alpha, 1]) = 1$. Then, the total cost is an increasing function of n.*

As the second stage productivity factor only appears in the equations when the online cross-training is profitable, one may be inclined to think that the requirement $\mathbb{P}(\tilde{\delta}_\alpha^2 \in [c_\alpha^2/h_\alpha, 1]) = 1$ does not play much of a role. However, if we cannot ensure that the online cross-training will be profitable after realizing the parameters, then the total cost might actually decrease as the variability of the second stage productivity factor increases. To understand why this is possible consider the following example. To concentrate on the effects of online cross-training parameters assume that offline cross-training is too expensive and $c_\alpha^1 = 10, c_\alpha^2 = 2, h_\alpha = 5, x_\alpha^0 = 50$ and $x_\gamma^0 = 80$. Also suppose that $\tilde{d}_\alpha = 106, \tilde{d}_\gamma = 0$ and $\tilde{\delta}_\alpha^1 = 1$ w. p. 1. The second stage productivity factor is random and $\mathbb{P}(\tilde{\delta}_\alpha^2 = 0.3) = \mathbb{P}(\tilde{\delta}_\alpha^2 = 0.7) = 0.5$. If the second stage productivity factor is high, $\tilde{\delta}_\alpha^2 = 0.7$, then the entire γ workforce is trained online to work on α. If $\tilde{\delta}_\alpha^2 = 0.3$, then online cross-training is not profitable and all the excess demand is lost. The expected total cost is 220. However, if $\mathbb{P}(\tilde{\delta}_\alpha^2 = 0.2) = \mathbb{P}(\tilde{\delta}_\alpha^2 = 0.8) = 0.5$, then the expected total cost is 210. We see that increasing the variance can be perceived as increasing the productivity factors above the mean and decreasing the values below the mean. Pushing the lower values of productivity factors down does not change the achievable capacity as incuring opportunity cost is preferable in any case. However, pushing the higher values of productivity factors up increases the achievable capacity and significantly reduces the second stage cost. Hence, the total cost decreases. Section 3 presents a counterexample where the offline cross-training is suffficiently cheap and the total cost decreases as the variability of the second stage productivity factor increases.

3 Numerical Experiments

In this section, we numerically analyze how varying different parameters affects cross-training policies. The problems for each experiment are designed to highlight the caveats for the



(a) Cross-training and opportunity costs (b) Total cost

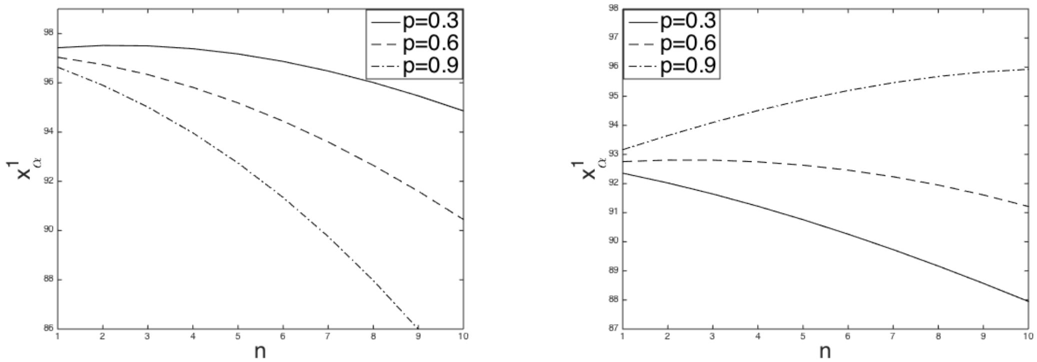
Figure 2: Costs with respect to the variability of demand for α when productivity factor is random

theory in the previous sections. In the examples, we only concentrate on cross-training costs for task α and due to separability we do not need to specify cost parameters related to task γ.

3.1 Effects of Demand Variability on Cross-Training Levels When Productivity Factors are Random

Proposition 2.3 states that the offline cross-training decreases as the variability in demand increases when the second stage cross-training is not profitable, the first stage productivity factors are deterministic and costs satisfy some mild conditions. This is in agreement with the postponement literature. In this section, we provide an example to show the crucial role played by the deterministic nature of productivity factors. In this example, we have $(x_\alpha^0, x_\gamma^0) = (10, 100)$, $c_\alpha^1 = 55, h_\alpha = 100$, $\mathbb{P}(\tilde{\delta}_\alpha^1 = 1) = 0.1$ and $\mathbb{P}(\tilde{\delta}_\alpha^1 = 0.6) = 0.9$ and $\mathbb{P}(\tilde{\delta}_\alpha^2 = 0) = 1$. Using the notation in Proposition 2.3, we have the expected value of $\tilde{d}_\alpha$ $\mu_1 = 25, \tilde{d}_1 \sim$Uniform$(-1, 1)$, i.e., for any given $n > 0$, $\tilde{d}_\alpha \sim$Uniform$(25 - n, 25 + n)$. To concentrate on the effects of variability of demand α, we assume $\tilde{d}_\gamma = 0$ w. p. 1.

Figure 2(a) shows that when the first stage productivity factor is random, the optimal offline cross-training first increases as the variability of demand increases up to a certain threshold and then decreases. The trend in the optimal opportunity cost is the opposite



(a) $\delta = 0.6$ (b) $\delta = 0.8$

Figure 3: The effect of productivity factor variability on the optimal cross-training levels

and it increases (decreases) as the optimal offline cross-training level decreases (increases). However, the variability of demand always has an adverse affect on the total cost, and regardless of the trend in the other costs, the total cost increases as the variability increases.

For $0 < n < 4$, $\mathbb{P}(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^{1*} < \tilde{d}_\alpha | \tilde{\delta}_\alpha^1 = 0.6) = 1$ and $\mathbb{P}(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^{1*} < \tilde{d}_\alpha | \tilde{\delta}_\alpha^1 = 1) < 0.5$. Hence, as n increases in this interval the first probability is not affected, but the second probability increases and one should increase the optimal cross-training level to recover Equation (2). This behavior is due to not having $\mathbb{P}(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^{1*} < \tilde{d}_\alpha | \tilde{\delta}_\alpha^1 = \delta) > 0.5$ for all values of δ.

3.2 The Effect of the Variability of the First Stage Productivity Factors when Second Stage Cross-Training is Profitable

We now investigate how the variability of the first stage productivity factor affects the offline cross-training levels. In this example, we take $x_\alpha^0 = 0, x_\gamma^0 = 100, c_\alpha^1 = 50, c_\alpha^2 = 120$,$h_\alpha = 400$, $\tilde{d}_\alpha$ ∼Uniform(0,100) and $\tilde{d}_\gamma = 0$ w. p. 1. We assume that the first stage productivity factor for task α is random and $\mathbb{P}(\tilde{\delta}_\alpha^1 = \delta - 0.01n) = p$ and $\mathbb{P}(\tilde{\delta}_\alpha^1 = \delta + 0.01n) = 1 - p$. The second stage productivity factor is deterministically equal to 0.6. In our experiments, we vary δ, n and p and obtain the optimal offline cross-training levels.

Figure 3 shows that when the mean of the first stage productivity factor is equal to the second stage productivity factor, the offline cross-training levels tend to decrease as the variability of the first stage productivity factor increases. However, when the mean of the offline productivity factor is strictly greater than the second stage productivity factor, the optimal offline cross-training level may increase as the variability of the first stage productivity factor increases.

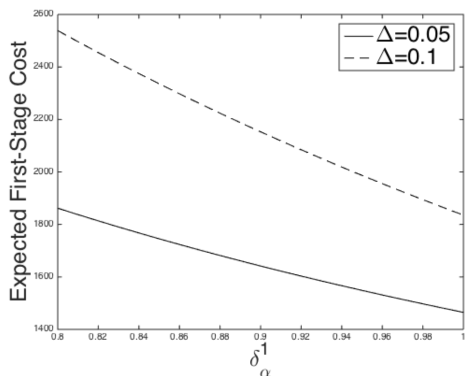
3.3 The Change in Cross-Training Costs as Productivity Improves

Proposition 2.6 shows that if the productivity factors are improved while keeping the difference between the first and the second stage factors constant, one tends to invest less in offline cross-training. Now, we perform experiments to understand how the optimal second stage opportunity and total costs respond to the changes in productivity factors. In this example, we take $x^0_\alpha = 0, x^0_\gamma = 100, c^1_\alpha = 50, c^2_\alpha = 70, h = 200, \tilde{d}_\alpha$ ∼Uniform(0,100) and $\tilde{d}_\gamma = 0$ w. p. 1. We also assume that the productivity factors are deterministic and satisfy $\delta^2_\alpha = \delta^1_\alpha - \Delta$, and perform experiments by changing δ^1_α and Δ.

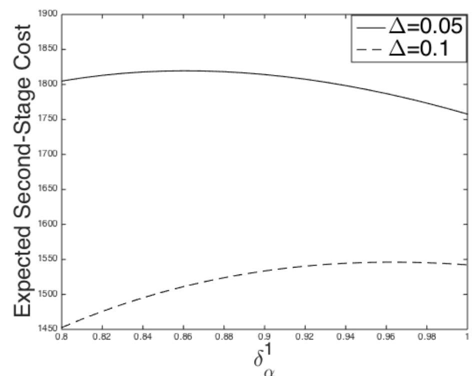
Figure 4(a) confirms the result of Proposition 2.6 and Figures 4(c) and 4(d) are in accordance with our expectations as both opportunity and total costs decrease as the productivity factors improve. Figure 4(b) is interesting and shows that if the difference between productivity factors is relatively small, the second stage cross-training cost first increases and then decreases. However, when the difference between the productivity factors is relatively large, the second stage cost increases as the productivity factors are improved.

3.4 Capacity Constraints on Online Cross-Training

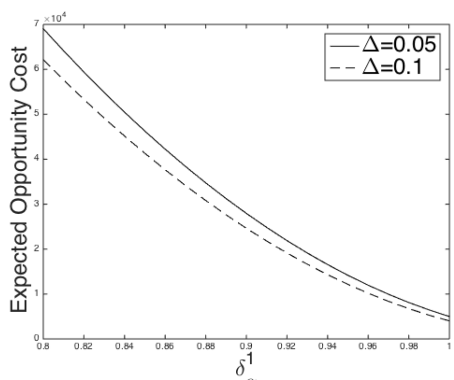
Our model assumes that all the capacity can be cross-trained if needed. In many real world situations, the resources are limited and it is very difficult to cross-train workers after the demand is revealed and when the workers are already trying to satisfy the observed demand. This might impose a capacity on the number of workers that can be cross-trained online. In this section, we investigate how such a capacity constraint affects the optimal cross-training levels and related costs. We take $x^0_\alpha = 0, x^0_\gamma = 100, c^1_\alpha = 50, c^2_\alpha = 70, h = 200, \tilde{d}_\alpha$ ∼Uniform(0,100) and $\tilde{d}_\gamma = 0, \tilde{\delta}^1_\alpha = 0.9, \tilde{\delta}^2_\gamma = 0.7$ w. p. 1. We denote the online cross-training capacity as K_α.



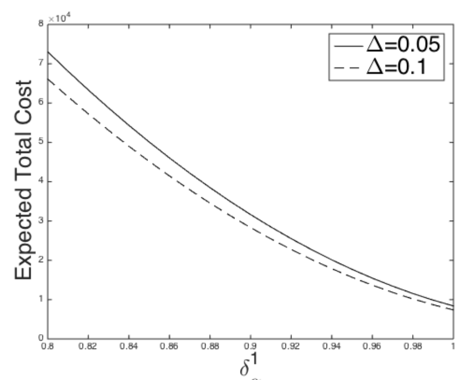
(a) First stage cost



(b) Second stage cost

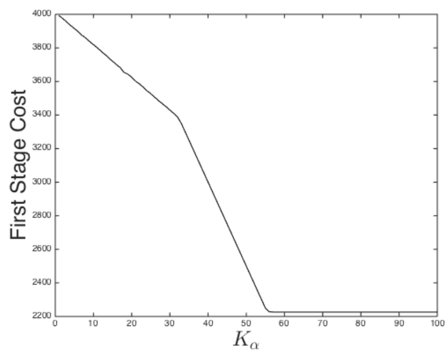


(c) Opportunity cost

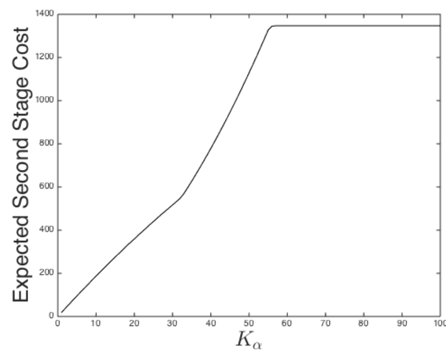


(d) Total cost

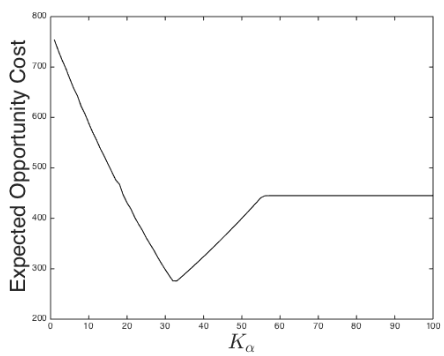
Figure 4: The sensitivity of various costs to changes in productivity factors



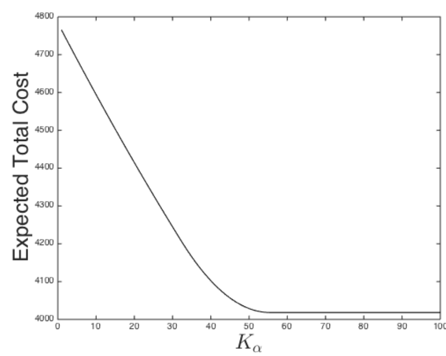
(a) First stage cost



(b) Second stage cost



(c) Opportunity cost



(d) Total cost

Figure 5: The sensitivity of various costs to online cross-training capacity

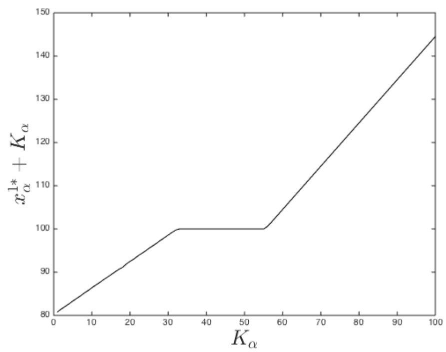


Figure 6: The sum of optimal cross-training and online cross-training capacity vs. online cross-training capacity.

Figure 5 shows that the optimal offline cross-training cost first decreases slowly up to a certain point and then decreases rapidly and after a certain threshold it stays constant. The behavior of the expected online cross-training cost is exactly the opposite. More interestingly, we see that the opportunity cost decreases as the offline cross-training cost decreases slowly and then increases as the offline cross-training cost decreases rapidly. We also see that the total cost is a convex function of the online cross-training capacity.

To understand this behavior, we plot the sum of optimal offline cross-training and online cross-training capacity versus the online cross-training capacity in Figure 6. We see that the offline cross-training cost decreases slowly when the sum is less than, x_γ^0, the total number of workers that can be cross-trained. When this is the case, increasing the capacity yields a reduction in the opportunity cost. As the online cross-training capacity is increased we reach a situation where the sum under consideration is equal to x_γ^0 and the optimal offline cross-training level is still greater than the optimal level without the capacity constraint. When this is the case, an increase in the online cross-training capacity constraint will be exactly equal to the decrease in the optimal offline cross-training level, which implies a decrease in the achievable capacity and hence, the opportunity cost increases. When the optimal offline cross-training level hits the optimal level without capacity constraints, the capacity constraint does not have any effect on costs as all the remaining workers can be cross-trained online as needed.

4 Concluding Remarks

In this work, we have studied the effects of imperfect training schemes on the cross-training policies. We have considered a two-stage model, where the workers can be cross-trained offline in the first stage, before the demand is realized, and online in the second stage as the demand is revealed. The cross-trained workers are assumed to be less productive than the workers who are originally trained to do a specific task and the productivity of the cross-trained workers may depend on when they are cross-trained (offline or online). We have defined the achievable capacity as the maximum demand that can be satisfied from a task after all random parameters are realized and all the idle workers from the other task are cross-trained. We have shown that when the first stage and second stage training schemes are equally effective, the achievable capacity does not depend on when the training is done and the cross-training decisions are independent of the opportunity cost of lost demand. However when the offline and online training policies differ in their effectiveness, deferring cross-training implies a significant decrease in the achievable capacity and hence, the decision maker needs to consider a three-way trade off between cross-training costs of offline and online schemes and opportunity cost of lost demand.

We have also analyzed how the variability of demand and productivity factors affect our cross-training decisions. We have shown under some mild conditions that when the productivity levels of workers trained at different times are consistent, we tend to postopone cross-training as the demand or productivity factors become more variable. However, when the workers cross-trained online are less productive, we show via counter-examples that we may wish to increase increase offline cross-training as variability increases to avoid losing precious achievable capacity.

The insights provided in this paper can be used to devise effective solution methods to address the case with more than three tasks. When there are three or more tasks, the cross-training problem is no longer separable and the analysis and solution methodology need to be modified. We have also developed a two-stage stochastic integer program to aid decision makers to design cross-training policies in the presence of multiple tasks. We do not present this model in this paper to keep the focus on managerial insights. The integer programming model is available from the authors upon request.

Acknowledgements

The authors are grateful to the department editor, Michael Pinedo, the senior editor and the referees for their constructive comments which helped improve the paper significantly.

A Proofs of Propositions in Section 2

A.1 Proof of Proposion 2.1

Proof. We derive the first order optimality conditions by setting the derivatives equal to 0. Since the bounds on productivity factors $\tilde{\delta}^1$ and $\tilde{\delta}^2$ do not depend on the decision variables, using Leibniz rule

$$\frac{\partial \mathbb{E}[\mathbb{E}[g(x^1_\alpha, x^1_\gamma, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]]}{\partial x^1_\alpha} = \mathbb{E}\left[\frac{\partial \mathbb{E}[g(x^1_\alpha, x^1_\gamma, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]}{\partial x^1_\alpha}\right] = c^1_\alpha + \mathbb{E}\left[\sum_{i=1}^{6} \frac{\partial \mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi}); \Omega_i|\tilde{\delta}^1, \tilde{\delta}^2]}{\partial x^1_\alpha}\right].$$

The second stage cost function $v(x^1_\alpha, x^1_\gamma, \tilde{\xi})$ is constant with respect to x^1_α on Ω_4, Ω_5 and Ω_6 and the bounds of these regions do not depend on x^1_α. Also on Ω_1, the second stage cost is uniformly equal to 0. Hence, the derivatives of expectation over these regions are all equal to 0. To simplify the notation, we use $f(\tilde{d}_\alpha, \tilde{d}_\gamma)$ to denote the density function of demand vector when productivity factors are given. The derivative of expectation over Ω_2 is

$$\begin{aligned}\frac{\partial \mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi}); \Omega_2|\tilde{\delta}^1, \tilde{\delta}^2]}{\partial x^1_\alpha} &= -h_\alpha \tilde{\delta}^1_\alpha \int_{x^0_\alpha + \tilde{\delta}^1_\alpha x^1_\alpha}^{\infty} \int_0^{x^0_\gamma - x^1_\alpha} f(d_\alpha, d_\gamma) dd_\gamma dd_\alpha \\ &\quad - \int_{x^0_\alpha + \tilde{\delta}^1_\alpha x^1_\alpha}^{\infty} h_\alpha (d_\alpha - x^0_\alpha - \tilde{\delta}^1_\alpha x^1_\alpha) f(d_\alpha, x^0_\gamma - x^1_\alpha) dd_\alpha\end{aligned}$$

Similarly, we can calculate the derivative of expectation over Ω_3

$$\frac{\partial \mathbb{E}[v(x^1_\alpha, x^1_\gamma, \tilde{\xi}); \Omega_3|\tilde{\delta}^1, \tilde{\delta}^2]}{\partial x^1_\alpha} = \int_{x^0_\alpha + \tilde{\delta}^1_\alpha x^1_\alpha}^{\infty} h_\alpha (d_\alpha - x^0_\alpha - \tilde{\delta}^1_\alpha x^1_\alpha) f(d_\alpha, x^0_\gamma - x^1_\alpha) dd_\alpha$$

Canceling the boundary terms, we get

$$\frac{\partial \mathbb{E}[\mathbb{E}[g(x^1_\alpha, x^1_\gamma, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]]}{\partial x^1_\alpha} = c^1_\alpha - h_\alpha \mathbb{E}[\tilde{\delta}^1_\alpha; x^0_\alpha + \tilde{\delta}^1_\alpha x^1_\alpha \leq d_\alpha, x^0_\gamma - x^1_\alpha > d_\gamma]. \tag{7}$$

The partial derivative with respect to x_γ^1 can be found similarly. Setting these terms to equal 0, we get equation (2). Now, we need to show that the (x_α^1, x_γ^1) pairs that solve these equations, actually minimizes the expected cost by showing that expected cost function is convex in decision variables. Equation (7) suggests that the cross-partials are 0 and the Hessian matrix is positive semidefinite if the second partial derivatives with respect to x_α^1 and x_γ^1 are both nonnegative.

$$\frac{\partial^2 \mathbb{E}[\mathbb{E}[g(x_\alpha^1, x_\gamma^1, \tilde{\xi})|\tilde{\delta}^1, \tilde{\delta}^2]]}{(\partial x_\alpha^1)^2} = -h_\alpha \frac{\partial \mathbb{E}[\tilde{\delta}_\alpha^1 \int_{x_\alpha^0+\tilde{\delta}_\alpha^1 x_\alpha^1}^{\infty} \int_0^{x_\gamma^0 - x_\alpha^1} f(d_\alpha, d_\gamma) dd_\gamma dd_\alpha]}{\partial x_\alpha^1}.$$

Observe that if $x_\alpha^1 < \bar{x}_\alpha^1$, then

$$\{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 + \tilde{\delta}_\alpha^1 \bar{x}_\alpha^1 < \tilde{d}_\alpha, 0 \leq \tilde{d}_\gamma \leq x_\gamma^0 - \bar{x}_\alpha^1\} \subseteq \{(\tilde{d}_\alpha, \tilde{d}_\gamma) : x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 < \tilde{d}_\alpha, 0 \leq \tilde{d}_\gamma \leq x_\gamma^0 - x_\alpha^1\}.$$

Using this relation and the fact that the density function and the productivity factor $\tilde{\delta}_\alpha^1$ are always positive, we get

$$\lim_{\Delta \to 0} \frac{\mathbb{E}[\tilde{\delta}_\alpha^0 (\int_{x_\alpha^0+\tilde{\delta}_\alpha^1 (x_\alpha^1+\Delta)}^{\infty} \int_0^{x_\gamma^0-(x_\alpha^1+\Delta)} f(d_\alpha, d_\gamma) dd_\gamma dd_\alpha - \int_{x_\alpha^0+\tilde{\delta}_\alpha^1 x_\alpha^1}^{\infty} \int_0^{x_\gamma^0 - x_\alpha^1} f(d_\alpha, d_\gamma) dd_\gamma dd_\alpha)]}{\Delta} \leq 0.$$

Plugging this back into the second derivative and repeating the same procedure for x_γ^1, we conclude that the Hessian is positive semidefinite and the expected cost function is convex. This also implies that if there is no solution satisfying 2 in $[0, x_\gamma^0]$ the optimal solution is either 0 or x_γ^0 as suggested in cases 2 and 3.

Since the distribution is assumed to be continuous, the derivative in (7) is continuous in x_α^1. If the expectation takes both negative and positive values over $x_\alpha^1 \in [0, x_\gamma^0]$, then intermediate value theorem ensures us that (2) will be satisfied for some x_α^{1*}. Hence, the three cases stated in the proposition cover all possible situations. □

A.2 Proof of Corollary 2.2

Proof. Using the fact that $\tilde{\delta}_\alpha^1 \geq 0$ w. p. 1, we get

$$\mathbb{E}[\tilde{\delta}_\alpha^1; x_\alpha^0 + \tilde{\delta}_\alpha^1 {x_\alpha^1}^* \leq \tilde{d}_\alpha, x_\gamma^0 - {x_\alpha^1}^* > \tilde{d}_\gamma] \leq \mathbb{E}[\tilde{\delta}_\alpha^1] \leq \frac{c_\alpha^1}{h_\alpha}.$$

Using the third part of Proposition 2.1 the result follows. To prove the second and third parts of the corollary, we use the same methodology realizing $\tilde{\delta}^1_\alpha \leq 1$ w. p. 1. □

A.3 Proof of Proposition 2.2

Proof. The optimal offline cross-training level x^{1*}_α is a function of δ^1_α. Taking the implicit derivative of (2) with respect to δ^1_α, we get

$$\begin{aligned}
0 &= \mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha)\mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma) + \delta^1_\alpha \frac{\partial \mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma)}{\partial \delta^1_\alpha}\mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma) \\
&\quad + \delta^1_\alpha \mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha)\frac{\partial \mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma)}{\partial \delta^1_\alpha} \\
&= \mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma)\left(\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha) - \delta^1_\alpha f_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)x^{1*}_\alpha\right) \\
&\quad - \delta^1_\alpha \frac{\partial x^{1*}_\alpha}{\partial \delta^1_\alpha}\left(\mathbb{P}(x^0_\gamma - x^{1*}_\alpha > \tilde{d}_\gamma)\delta^1_\alpha f_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha) + \mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha) f_{\tilde{d}_\gamma}(x^0_\gamma - x^{1*}_\alpha)\right),
\end{aligned}$$

which implies that $\frac{\partial x^{1*}_\alpha}{\partial \delta^1_\alpha} \geq 0$ if and only if $\mathbb{P}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha \leq \tilde{d}_\alpha) \geq \delta^1_\alpha f_{\tilde{d}_\alpha}(x^0_\alpha + \delta^1_\alpha x^{1*}_\alpha)x^{1*}_\alpha$. □

A.4 Proof of Proposition 2.3

Proof. Using equation (3) and independence, we get

$$\mathbb{P}\left(\frac{x^0_\alpha + \delta^1_\alpha {x^1_\alpha}^* - \mu_1}{m} \leq \tilde{d}_1\right)\mathbb{P}\left(\frac{x^0_\gamma - {x^1_\alpha}^* - \mu_2}{n} > \tilde{d}_2\right) = \frac{c^1_\alpha}{\delta^\alpha_1 h_\alpha} \geq \frac{1}{2}.$$

Now, we can infer that the probabilities on the left-hand side of the inequality should be greater than 0.5. Then, using the fact that $\tilde{d}_1$ and $\tilde{d}_2$ are symmetric random variables

$$\frac{x^0_\alpha + \delta^1_\alpha {x^1_\alpha}^* - \mu_1}{m} \leq 0 \text{ and } \frac{x^0_\gamma - {x^1_\alpha}^* - \mu_2}{n} \geq 0. \tag{8}$$

The left-hand side of equation (3) decreases as m increases. Hence, if statement 1 of Corollary 2.1 is true, we need to decrease x^{1*}_α to recover the equality. If statement 2 is true, we may either wish to stay at x^0_γ or we may wish to decrease x^1_α. For the third statement, we do not need to take any action. Hence, this proves that the optimal cross-training level is non-increasing in m. Similar arguments show that x^{1*}_α is non-increasing in n. □

A.5 Proof of Proposition 2.4

Proof. Similar to the proof of Proposition 2.1 we need to derive the first and second order optimality conditions. On regions $\Omega_1, \Omega_3, \Omega_4$ and Ω_6, the structure is the same as in Proposition 2.1 and the problem is separable in cross-training levels x_α^1 and x_γ^1.

The derivative of the second-stage cost function over Ω_2^a and Ω_2^bcan be calculated as:

$$\begin{aligned}
\frac{\partial \mathbb{E}[v(x_\alpha^1, x_\gamma^1, \tilde{\xi}); \Omega_2^a | \tilde{\delta}^1, \tilde{\delta}^2]}{\partial x_\alpha^1} &= -\int_0^{x_\gamma^0 - x_\alpha^1} \int_{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1}^{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma)} \frac{c_\alpha^2 \tilde{\delta}_\alpha^1}{\tilde{\delta}_\alpha^2} f(\tilde{d}_\alpha, \tilde{d}_\gamma) d\tilde{d}_\alpha d\tilde{d}_\gamma \\
&+ \int_0^{x_\gamma^0 - x_\alpha^1} (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2) c_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma) f(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma), \tilde{d}_\gamma) d\tilde{d}_\gamma, \\
\frac{\partial \mathbb{E}[v(x_\alpha^1, x_\gamma^1, \tilde{\xi}); \Omega_2^b | \tilde{\delta}^1, \tilde{\delta}^2]}{\partial x_\alpha^1} &= -\int_0^{x_\gamma^0 - x_\alpha^1} \int_{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma)}^{\infty} (c_\alpha^2 + h_\alpha (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2)) f(\tilde{d}_\alpha, \tilde{d}_\gamma) d\tilde{d}_\alpha d\tilde{d}_\gamma \\
&- \int_0^{x_\gamma^0 - x_\alpha^1} (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2) c_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma) f(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma), \tilde{d}_\gamma) d\tilde{d}_\gamma \\
&- \int_{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1}^{\infty} h_\alpha (\tilde{d}_\alpha - (x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1)) f(\tilde{d}_\alpha, x_\gamma^0 - x_\alpha^1) d\tilde{d}_\alpha.
\end{aligned}$$

Aggregating all the results and cancelling the boundary terms as appropriate, we obtain

$$\begin{aligned}
\frac{\partial \mathbb{E}[g(x_\alpha^1, x_\gamma^1, \tilde{\xi}) | \tilde{\delta}^1, \tilde{\delta}^2]}{\partial x_\alpha^1} &= c_\alpha^1 - \int_0^{x_\gamma^0 - x_\alpha^1} \int_{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1}^{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma)} \frac{c_\alpha^2 \tilde{\delta}_\alpha^1}{\tilde{\delta}_\alpha^2} f(\tilde{d}_\alpha, \tilde{d}_\gamma) d\tilde{d}_\alpha d\tilde{d}_\gamma \\
&- \int_0^{x_\gamma^0 - x_\alpha^1} \int_{x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma)}^{\infty} (c_\alpha^2 + h_\alpha (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2)) f(\tilde{d}_\alpha, \tilde{d}_\gamma) d\tilde{d}_\alpha d\tilde{d}_\gamma.
\end{aligned} \tag{9}$$

Setting these terms to equal 0, we get equation (5). Now, we need to show that the (x_α^1, x_γ^1) pairs that solve these equations are optimal by checking the Hessian matrix. We do not need to consider the cross-partials and the Hessian matrix is positive semidefinite if the second partial derivatives with respect to x_α^1 and x_γ^1 are both nonnegative. We take the second derivative with respect to x_α^1 and get

$$\begin{aligned}
\frac{\partial^2 \mathbb{E}[g(x_\alpha^1, x_\gamma^1, \tilde{\xi}) | \tilde{\delta}^1, \tilde{\delta}^2]}{(\partial x_\alpha^1)^2} &= \frac{c_\alpha^2 (\tilde{\delta}_\alpha^1)^2}{\tilde{\delta}_\alpha^2} \int_0^{x_\gamma^0 - x_\alpha^1} f(x_\alpha^0 + \tilde{\delta}_\alpha^1 x_\alpha^1, \tilde{d}_\gamma) d\tilde{d}_\gamma \\
-(\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2) \left(\frac{c_\alpha^2 \tilde{\delta}_\alpha^1}{\tilde{\delta}_\alpha^2} - c_\alpha^2 - h_\alpha (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2) \right) & \int_0^{x_\gamma^0 - x_\alpha^1} f(x_\alpha^0 - \tilde{\delta}_\alpha^1 x_\alpha^1 + \tilde{\delta}_\alpha^2 (x_\gamma^0 - x_\alpha^1 - \tilde{d}_\gamma), \tilde{d}_\gamma) d\tilde{d}_\gamma \\
+ \left(c_\alpha^2 + h_\alpha (\tilde{\delta}_\alpha^1 - \tilde{\delta}_\alpha^2) \right) \int_{x_\alpha^0 + \tilde{\delta}^1 x_\alpha^1}^{\infty} f(\tilde{d}_\alpha, x_\gamma^0 - x_\alpha^1) d\tilde{d}_\alpha.
\end{aligned}$$

On the right-hand side, the first term is positive and the assumption $\tilde{\delta}_\alpha^1 > \tilde{\delta}_\alpha^2$ ensures that the third term is positive w. p. 1. The positivity of the second term follows from both $\tilde{\delta}_\alpha^1 > \tilde{\delta}_\alpha^2$

and $c_\alpha^2 < \tilde{\delta}_\alpha^2 h_\alpha$. Hence the objective function is convex and the solution which satisfies (9) minimizes the total cost.

When the expectation satisfies the inequality for the second case, the derivative of the cost function is negative for any value in $[0, x_\gamma^0]$. Hence, to minimize the cost, we need to set x_α^{1*} to the maximum possible value. The third case can be proven similarly. $\square$

A.6 Proof of Propositon 2.6

Proof. First to simplify the notation we define

$$
\begin{aligned}
\bar{P}_1(y) &= \mathbb{P}(x_\alpha^0 + \bar{\delta}_\alpha^1 y \leq \tilde{d}_\alpha \leq x_\alpha^0 + \bar{\delta}_\alpha^1 y + \bar{\delta}_\alpha^2 (x_\gamma^0 - y - \tilde{d}_\gamma), \tilde{d}_\gamma \leq x_\gamma^0 - y), \\
\bar{P}_2(y) &= \mathbb{P}(\tilde{d}_\alpha > x_\alpha^0 + \bar{\delta}_\alpha^1 y + \bar{\delta}_\alpha^2 (x_\gamma^0 - y - \tilde{d}_\gamma), \tilde{d}_\gamma \leq x_\gamma^0 - y) \\
\bar{P}(y) &= \bar{P}_1(y) + \bar{P}_2(y) = \mathbb{P}(x_\alpha^0 + \bar{\delta}_\alpha^1 y \leq \tilde{d}_\alpha, \tilde{d}_\gamma \leq x_\gamma^0 - y), \\
\bar{C}(y) &= \frac{c_\alpha^2 \bar{\delta}_\alpha^1}{\bar{\delta}_\alpha^2} \bar{P}_1(y) + (c_\alpha^2 + h_\alpha(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2)) \bar{P}_2(y).
\end{aligned}
$$

Similarly, define $\hat{P}_1(y), \hat{P}_2(y), \hat{P}(y)$ and $\hat{C}(y)$ by interchanging $\bar{\delta}$ with $\hat{\delta}$. First we prove that $\bar{C}(y)$ and $\hat{C}(y)$ are non-increasing functions of y. Let $y_1 < y_2$ and define $\Delta_1 = \mathbb{P}(\Omega_2^a(y_1) \setminus \Omega_2^a(y_2))$ and $\Delta_2 = \mathbb{P}(\Omega_2^a(y_2) \setminus \Omega_2^a(y_1))$ where "\" is the set difference operator. Then,

$$
\bar{C}(y_2) - \bar{C}(y_1) = \frac{c_\alpha^2 \bar{\delta}_\alpha^1}{\bar{\delta}_\alpha^2}(\Delta_2 - \Delta_1) - (c_\alpha^2 + h_\alpha(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2))\Delta_2 \leq \left(\frac{(c_\alpha^2 - h_\alpha \bar{\delta}_\alpha^2)(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2)}{\bar{\delta}_\alpha^2} \right) \Delta_2 \leq 0,
$$

where the last inequality follows from the assumption that online cross-training is preferable over incurring opportunity cost.

The condition $\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2 \geq \hat{\delta}_\alpha^1 - \hat{\delta}_\alpha^2$ implies that $\bar{\delta}_\alpha^1 / \bar{\delta}_\alpha^2 \geq \hat{\delta}_\alpha^1 / \hat{\delta}_\alpha^2$, and $c_\alpha^2 < \bar{\delta}^2 h_\alpha$ implies that $c_\alpha^2 \bar{\delta}_\alpha^1 / \bar{\delta}_\alpha^2 < c_\alpha^2 + h_\alpha(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2)$. Then, Equation (9) indicates

$$
\begin{aligned}
\hat{C}(\bar{x}_\alpha^{1*}) - c_\alpha^1 &= \hat{C}(\bar{x}_\alpha^{1*}) - \bar{C}(\bar{x}_\alpha^{1*}) \\
&\leq \frac{c_\alpha^2 \bar{\delta}_\alpha^1}{\bar{\delta}_\alpha^2}(\hat{P}_1(\bar{x}_\alpha^{1*}) - \bar{P}_1(\bar{x}_\alpha^{1*})) + (c_\alpha^2 + h_\alpha(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2))(\hat{P}_2(\bar{x}_\alpha^{1*})) - \bar{P}_2(\bar{x}_\alpha^{1*}))) \\
&\leq (c_\alpha^2 + h_\alpha(\bar{\delta}_\alpha^1 - \bar{\delta}_\alpha^2))(\hat{P}(\bar{x}_\alpha^{1*}) - \bar{P}(\bar{x}_\alpha^{1*})) \leq 0
\end{aligned}
$$

Then, $\hat{x}_\alpha^{1*} \leq \bar{x}_\alpha^{1*}$ follows as $\hat{C}(y)$ is non-increasing. $\square$

A.7 Proof of Proposition 2.7

Proof. As the problem is separable, the cost related to the demand for task γ is a constant with respect to $\tilde{\delta}^1_\alpha$ and for ease of notation we denote it as C. If online cross-training is not profitable , $c^2_\alpha > h_\alpha \delta^2_\alpha$, we can write the second stage cost function as follows:

$$v(x^1, \xi) = h_\alpha \max\{0, \tilde{d}_\alpha - x^0_\alpha - \tilde{\delta}^1_\alpha x^1_\alpha, \tilde{d}^1_\alpha - x^0_\alpha - \tilde{\delta}^1_\alpha (x^0_\gamma - \tilde{d}_\gamma)\} + C.$$

Similarly, if online cross-training is profitable, $c^2_\alpha \leq h_\alpha \delta^2_\alpha$,

$$\begin{aligned} v(x^1, \xi) &= \max \Big\{ 0, c^2_\alpha \frac{d_\alpha - x^0_\alpha - \delta^1_\alpha x^1_\alpha}{\delta^2_\alpha}, \\ &\qquad c^2_\alpha (x^0_\gamma - x^1_\alpha - d_\gamma) + h_\alpha (d_\alpha - x_\alpha - \delta^1_\alpha x^1_\alpha - \delta^2_\alpha (x^0_\gamma - x^1_\alpha - d_\gamma)) \Big\} + C. \end{aligned}$$

Both functions are convex with respect to δ^1_α. Then, the function $v(x^1, n) = \mathbb{E}(v(x^1, d, \delta^1_\alpha + n\Delta, \delta^2_\alpha))$ is convex with respect to n. Then, using Jensen's inequality for any $n > 0$ $v(x^1, n) \geq v(x^1, 0)$. If $0 < n_1 < n_2$, then using convexity

$$v(x^1, n_1) \leq \frac{n_2 - n_1}{n_2} v(x^1, 0) + \frac{n_1}{n_2} v(x^1, n_2) \leq \frac{n_2 - n_1}{n_2} v(x^1, n_1) + \frac{n_1}{n_2} v(x^1, n_2).$$

Manipulating, these equations we conclude that $v(x^1, n_2) \geq v(x^1, n_1)$ for any x^1. Let x^{1*} and x^{1**} be the optimal offline cross-training levels for n_1 and n_2 respectively. Then,

$$v(x^{1*}, n_1) \leq v(x^{1**}, n_1) \leq v(x^{1**}, n_2)$$

which concludes the proof. □

References

Bassamboo, A., R.S. Randhawa, J.A. Van Mieghem. 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Science* **56** 1285–1303.

Bassamboo, A., R.S. Randhawa, J.A. Van Mieghem. 2012. A little flexibility is all you need: On the value of flexible resources in queueing systems. *Operations Research* **60** 1423–1435.

Brusco, M.J., T.R. Johns. 1998. Staffing multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences* **29**(2) 499–515.

Chakravarthy, S.R., S.R. Agnihothri. 2005. Optimal workforce mix in service systems with two types of customers. *Production & Operations Management* **14**(2) 218–231.

Chou, M.C., G.A. Chua, C-P. Teo. 2010. On range and response: Dimensions of process flexibility. *European Journal of Operational Research* **207** 711–724.

Davis, D.D., H.V. Kher, B.J. Wagner. 2009. Influence of workload imbalances on the need for worker flexibility. *Computers and Industrial Engineering* **57** 319–329.

Feitzinger, E., H. L. Lee. 1999. Mass customization at Hewlett-Packard: The power of postponement. *Harvard Business Review* **75**(1) 116–121.

Gel, E.S., W.J. Hopp, M.P. Van Oyen. 2007. Hierarchical cross-training in work-in-process-constrained systems. *IIE Transactions* **39** 125–143.

Gnanlet, A., G.G. Wendell. 2009. Sequential and simultaneous decision making for optimizing health care resource flexibilities. *Decision Sciences* **40** 295–326.

Hopp, W.J., E. Tekin, M.P. Van Oyen. 2004. Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science* **50**(1) 83–98.

Hopp, W.J., M.P. Van Oyen. 2004. Agile workforce evaluation: a framework for cross-training and coordination. *IIE Transactions* **36** 919–940.

Iravani, S.M.R., B. Kolfal, M.P. Van Oyen. 2007. Call-center labor cross-training: it's a small world after all. *Management Science* **53**(7) 1102–1112.

Jordan, W.C., S.C. Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Science* **41**(4) 577–594.

Jordan, W.C., R.R. Inman, D.E. Blumenfeld. 2004. Chained cross-training of workers for robust performance. *IIE Transactions* **36** 953–967.

Lyons, R.F. 1992. Cross-training: A richer staff for leaner budgets. *Nursing Management* **23**(1) 43–44.

McCune, J.C. 1994. On the train gang. *Management Review* **83**(10) 57–60.

Netessine, S., G. Dobson, R.A. Shumsky. 2002. Flexibly service capacity: Optimal investment and the impact of demand correlation. *Operations Research* **50**(2) 375–388.

Pinker, E.J., R.A. Shumsky. 2000. The efficiency-quality trade-off of cross-trained workers. *Manufacturing & Service Operations Management* **2**(1) 32–48.

Tanrisever, F., D. Morrice, D. Morton. 2012. Managing capacity flexibility in make-to-order production environments. *European Journal of Operational Research* **216**(2) 334 – 345.

Van Mieghem, J.A. 1998. Investment strategies for flexible resources. *Management Science* **44**(8) 1071–1078.

Van Mieghem, J.A., N. Rudi. 2002. Newsvendor networks: dynamic inventory management and capacity investments with discretionary pooling. *Manufacturing & Service Operations Management* **4**(4) 313–335.

Walsh, M., W.M. Carlyle, J.W. Fowler, C. Wee. 2000. Modeling operator cross-training in wafer fabrication given uncertain demand: A stochastic programming approach. J. K. Cochran, J. W. Fowler, S. Brown, eds., *Proc. of the Int. Conf. on Modeling and Analysis of Semiconductor Manufacturing*. 226–232.

Wang, Y., W. Gilland, B. Tomlin. 2010. Mitigating supply risk: Dual sourcing or process improvement. *Manufacturing & Service Operations Management* **12**(3) 489–510.