Z Betriebswirtsch (2012) 82:745–762
DOI 10.1007/s11573-012-0589-5

FORSCHUNG

ZfB
ZEITSCHRIFT FÜR
BETRIEBSWIRTSCHAFT

The effects of state-dependent human behavior on the design of a serial line



Irina Heimbach · Jörn Grahl · Franz Rothlauf

Abstract: Most research on line design assumes that human operators perform independently from the status of the line. Recent empirical evidence is contradictory. Humans are likely to change their working speed if they might otherwise cause idle time (Schultz et al., Manage Sci 44(12):1595–1607, 1998). This peculiarity of worker behavior is observed in a variety of settings but little is still known about optimal line design that accounts for this more realistic modeling of worker behavior. Therefore, we analyze work allocation in a serial line with limited buffer capacity and adaptive human behavior. An extensive simulation study reveals that optimal work allocation in state-dependent models is different from classical state-independent models. A bowl-shaped work allocation might be suboptimal and design guidelines are more complicated. Depending on the extent of human reactions, a bowl-shaped, balanced, or reversed-bowl work allocation can be preferable.

Keywords: Behavioral operations management · Work allocation · Production line design · State-dependent behavior

Received: 20.06.2011 / **Published online:** 26.05.2012

Author contributions. All authors designed the research. Irina Heimbach and Jörn Grahl conducted the research and wrote the manuscript.

Dipl.-Kffr. I. Heimbach (✉)
Information Systems III/Electronic Markets, Department of Law and Economics,
Technical University Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany
e-mail: heimbach@emarkets.tu-darmstadt.de

Dr. J. Grahl
Johannes Gutenberg-University Mainz,
Jakob-Welder-Weg 9, 55128 Mainz, Germany
e-mail: grahl@uni-mainz.de

Prof. Dr. F. Rothlauf
Gutenberg School of Management and Economics, Johannes Gutenberg-University Mainz,
Jakob Welder-Weg 9, 55099 Mainz, Germany
e-mail: rothlauf@uni-mainz.de

JEL Classification: M11

1 Introduction

Since the 60's, scientists have studied the design of production lines. The optimal allocation of work and interstage buffers has ever been an important means to increase their efficiency. In a seminal paper, Hillier and Boling (1966) showed that lines with variable processing times at stations, if organized in a certain manner, achieve more throughput than lines with equal processing times. They named this allocation a "bowl phenomenon". It implicates the assignment of more work at the end and the beginning of a serial production line. Such a design promises a throughput improvement over a balanced allocation, which can provide considerable cost savings over the lifetime of a line (Hillier and Boling 1966, 1979). A vast amount of simulation- and analytical results confirmed the throughput increase caused by work times that are disposed according to a bowl shape (Hillier and Boling 1979; Pike and Martin 1994). The bowl phenomenon is robust even when the optimal work allocation is estimated inaccurately (Hillier and So 1996).

Almost all studies on line design assume that processing times are independent of past events and of the current state of the production system. Is this assumption realistic? Are workers insensitive to whatever happens in their immediate environment? Former and recent findings show that workers are likely to react on workload (Edie 1954; Kc and Terwiesch 2009), on the feed rate of an assembly line (Franks and Sury 1966) or on their coworkers (Falk and Ichino 2006; Mas and Moretti 2009) and even adjust their working speed to avoid idle time (Schultz et al. 1998). Schultz et al. (1998) denote this occurrence as state-dependent behavior. It is also closely related to Gutenberg's production theory, specifically intensity adjustments (Fandel 2005). What happens in a production line, if workers are adaptive and work is distributed according to the bowl phenomenon? Does adaptive behavior significantly change the throughput? Are guidelines for line design as suggested by previous research still relevant under consideration of the state-dependent behavior? Which modifications should be made to these guidelines, if any?

We study work allocation in serial lines with state-dependent behavior by building a formal model and conducting a series of simulation experiments. We examine the effects of increasing line length and buffer capacities on work allocation guidelines and productivity. Our results show that state-dependent behavior counteracts production losses in production lines caused by human processing time variability. The pattern of work allocation depends on the extent of worker's adaptive behavior. Bowl phenomena hold for moderate processing time adjustments. A balanced pattern is best for realistic human adaption rates. Reversed bowl patterns are best, if workers adjust their operation speed greatly. Further, the effects of increasing line length and buffer capacities differ from what models with independence assumptions predict. Human adaptive behavior compensates for negative effects of lower buffer capacities and increasing number of stations.

This paper is structured as follows. Section 2 provides the theoretical background required to follow our study by introducing serial production lines, previous findings on optimal work allocation and integrating behavioral factors. Section 3 describes the theoretical model. Section 4 presents experimental results and analyzes the findings. It compares the results to those obtained in models with state-independent behavior. The last section concludes the paper and gives suggestions for practice and future research.

2 Literature review

2.1 Work allocation problem

A *serial production line* embraces material, work and storage areas. The material flows in fixed sequence, visiting each work and storage area exactly once (Dallery and Gershwin 1992, pp. 3–4). In the work areas, called *work stations*, human workers accomplish certain operations on the material. After completing an operation, a station can deposit a worked good into a buffer, a storage area, which has limited capacity. The time spent by a station to perform an operation on a production unit is called *processing time*, *operation time*, or *service time* (Dallery and Gershwin 1992, p. 13). This study considers *unpaced lines* (*asynchronous*) where each station operates with an operating time free of external pacing effects.

Human operators have stochastic processing times modeled by an underlying probability distribution. This processing time variability affects the maximum flow rate of material called *production rate*, *efficiency*, or *throughput* as follows. When a worker processes material, she reduces the level of her upstream buffer and increases the level of the downstream buffer when she has finished. When a station fails or takes an especially long time to process a product unit, while its neighbors work at their normal pace, the level of its upstream buffer tends to increase and the level of the downstream buffer tends to decrease. The upstream buffer might become full or the downstream buffer might become empty. One of the adjacent workers is not able to process: either the upstream worker is *blocked*, waiting to be able to put a completed part into the buffer, or the downstream worker is *starved*, waiting for available work in front of an empty buffer. A worker can be blocked and starved simultaneously. Both states describe *idleness* or wasted time in a production line, which, as a consequence, reduce production output (Dallery and Gershwin 1992, p. 5). The disturbance of a smooth material flow is called *coupling effect* or *stochastic interference*. Several suggestions have been made to dampen it:

1. An increase in interstage buffer capacities uncouples the stations. However, this usually raises costs in terms of increased storage space requirements and of increased work-in-process inventory (Pike and Martin 1994, p. 483).
2. Allocating a fixed amount of storage capacity in a certain way improves the efficiency of the line. This problem is known as the *optimal buffer allocation problem* in the literature, which we do not consider in the current study.
3. Idleness can be reduced by shortening a line. Shorter lines exhibit less interference effects. On the other hand, tasks in short lines are less specialized, which can lead to an efficiency decrease (Pike and Martin 1994, p. 483).
4. The efficiency of the line can be increased by distributing workload over stations in a certain way. This problem is known as *optimal work allocation* and is addressed in this study.

A fundamental result on optimal work allocation has been established in Hillier and Boling (1966). Assuming exponentially distributed processing times, they investigate the effects of different workload distributions across lines with 2, 3, or 4 stations and limited inter-stage buffer capacities. They test the issue of deliberately unbalancing the production line in order to increase line productivity. Using analytical methods developed by Hunt (1956), Hillier and Boling show that the output can be increased by assigning unequal processing times to the stations. Further, they conclude that operation times at the first and the last stations are interchangeable. The output rate is maximized by assigning equal service times to the first and the last station. Hillier and Boling describe the best allocation type as "bowl phenomenon", which implies that "a lower average operation time[should be assigned] to the station in the middle of the production line than to the stations on the two ends" (Hillier and Boling 1966, p. 655). The design of a line in a "bowl" manner promises an efficiency improvement (Hillier and Boling 1966, p. 656). In production dimensions of manufacturing, a very small improvement in efficiency can lead to a substantial amount of savings.

The results of Hillier and Boling sparked researchers' interest in work allocation. A review of the follow-up studies suggests the following summary:

1. Assigning mean processing times according to the "bowl phenomenon" leads to an improvement of the production rate over a balanced line, i.e., with equal processing times at stations (Hillier and So 1993).
2. The output rate of a line decreases with the number of sequential workstations. The "bowl phenomenon" holds for longer lines, however, the required average deviation from a balanced assignment, called the *degree of unbalance*, vanishes with the line length (Hillier and Boling 1979; Pike and Martin 1994).
3. The output rate increases as more buffer capacity is available due to a decoupling effect. The "bowl phenomenon" holds for increasing buffer capacities, but the degree of unbalance decreases (Hillier and Boling 1979).
4. The output rate decreases as the variability of service time increases. Operation times with more variability should be allocated to the ends of the station (*variability imbalance*). The degree of unbalance increases with the coefficients of variation (Payne et al. 1972; Kala and Hitchings 1973; Rao 1976; Carnall and Wild 1976).
5. The "bowl phenomenon" is robust, i.e., it holds even when a work process cannot be split exactly into micro-operations according to the analytically determined guidelines (Hillier and So 1996).
6. The "bowl phenomenon" and "variability imbalance" are not sensitive upon different processing time distributions (El-Rayah 1979; Hillier and Boling 1979; Pike and Martin 1994).

These results stem from numerous studies using analytical or simulation methods under the assumption of various processing time distributions, variances, permitting different sizes of buffer capacities and line lengths. These studies suggest possible advantages of unbalancing a serial production line in particular ways. However, all models share the assumption that the means of processing times are independent of each other and remain constant over time. Humans interact with the environment (contrary to machines), they respond to their surroundings. The next section reviews empirical evidence of related human state-dependent behavior.

2.2 Human state-dependent behavior

Edie (1954) and Kc and Terwiesh (2009) observe that people react to their workload. Edie (1954), describing delay times at toll booths, observes that the toll collectors expedite cars faster when traffic per booth is higher. He alludes to the existence of "a common phenomenon in waiting problems involving people who are aware of the amount of congestion" (see Edie 1954, p. 120). Kc and Terwiesch (2009) show in the context of health-care delivery services, that the service speed of workers increases with their workload.

Franks and Sury (1966) study the nature of operator response to pacing in conveyor work. Their main finding is that workers react to the arrival rate and decrease their mean operation times if the line feed rate increases. The study of Doerr et al. (1996) examines primarily effects of different material flow policies and goal-settings on the output in a manual production line. They note that workers increase their working speed in a setting with lower buffer capacities, so that total productivity did not decrease as predicted by an increase in stochastic interference.

Humans react to their social environment, including coworkers or social norms. Working in a group context changes individual task motivation and performance (Steiner 1972). Motivation is defined as "the contemporary (immediate) influence on direction, vigor, and persistence of action" (Atkinson 1964, p. 2). As a team is more than the sum of its members, the output of a team is not the sum of individual outputs. Regarding motivation in a team, there are two possible effects: motivation gains and motivation losses. Motivation gains occur when a team member is more productive than she would be without the team; motivation loss is the counterpart. The latter phenomenon can be explained by the "social loafing" theory, describing the tendency of an individual to reduce her effort at the expense of team members. Motivation losses are stronger when individual effort in a team is difficult to identify and to measure (Latané et al. 1979). Nevertheless, even in situations more prone to "free-riding effects" individuals do not necessarily behave opportunistically. Mas and Moretti (2009) study peer effects at the workplace. They investigate the dependence of the productivity of a cashier in a supermarket on the productivity of her coworkers. The worker's effort is positively correlated with the productivity of workers seeing her and with whom she frequently interacts.

Falk and Ichino (2006) conduct a laboratory experiment. The subjects are asked to fill letters into envelopes. Their payment is independent of the output. There are two treatments: "pair", where workers fill letters at the same time and in the same room, and "single", where each subject works alone. Falk and Ichino (2006) find that the productivity of two workers within pairs is higher than between pairs. They conclude that peer effects increase productivity.

Powell, Schultz and colleagues investigate the mechanism of adaptive worker behavior more deeply. Schultz et al. (1998) hypothesized that lower buffer capacities motivate workers who are slower or faster than their coworkers to adjust their work pace. To test their hypotheses, they design a laboratory experiment with high school students who should enter data while working in groups of three. There are two treatments, one with low buffer capacity, and one with high buffer capacity. Additional buffer capacities should lead to increased productivity. Therefore, the output in a line with lower buffer capacities is expected to be lower. The findings of Schultz et al. (1998) are contradictory. They find no significant difference in production output between groups. The analysis of work times shows that workers adjust their working pace depending on the state of the production line. The processing times are correlated with the capacity of the buffers, the pace of their coworkers and the contents of adjacent buffers. Idle time is lower, and output is higher than predicted by models with independence assumptions. In their experiment, the speeding up of workers completely compensates for productivity losses due to blocking and starving.

In two follow-up studies, the authors attempt to explain their observations with behavioral theories. Schultz et al. (1999) propose a model based on goal- and norm-directed behavior to explore determinants for individual worker performance. They describe a buffered serial line as a feedback-system. First, a worker can evaluate herself with some standard; second, she can compare her performance to a particular worker; third, she can be aware of causing idleness of a coworker (Schultz et al. 1999, pp. 1666–1667). The cohesiveness, interdependence and feedback in a production line encourage the development of task norms. These norms stimulate the peer pressure to adjust the working pace (Schultz et al. 1999, pp. 1674–1675). A line with lower buffer capacities provides more feedback than a line with higher ones. According to goal theory, increased feedback leads to more goal achievement, i.e., to a productivity increase (Locke and Latham 1990; Latham and Locke 1991).

Schultz et al. (2010) study the *regression-to-the-mean effect* which postulates that slower worker speed up and faster workers slow down (Schultz et al. 2010, p. 178). The motivational theory explaining regression towards the mean is inequity theory (Adams 1963). It states that workers desire equity between themselves and other employees. An individual compares her own output-input ratio to the ratio of other team members. If she observes inequity, several reaction alternatives are possible: to change inputs, to change outputs, to “leave the field”, to “psychologically distort” the outcomes or incomes of herself or of the referent person, and finally to change the referent person (Adams 1963, pp. 427–430). For a serial line, the implications are that workers slow down (decreasing the input) or work faster (increasing the input) (see Schultz et al. 2010, p. 179). Schultz et al. (1998) confirmed that slower workers speed up, but failed to demonstrate that workers slow down.

Powell and Schultz (2004) integrate these findings into a simulation model with the goal to investigate the relation between throughput and line length in production lines with state-dependent behavior. They show that line length has less impact on throughput than previously thought. They explain this phenomenon by built-in balancing mechanism that compensates for the effects of stochastic interference. Powell and Schultz (2004) show that in some cases efficiency actually increases with line length. Their findings are contradictory to results obtained for state-independent models. No existing study integrates state-dependent behavior into normative models of work allocation.

3 Model

Our model of state-dependent behavior is based on the one by Powell and Schultz (2004). The production line consists of N stations in series. Every production unit must be processed through each of the N single-worker stations in the same fixed sequence. There are $N-1$ buffers with fixed capacity C between the stations. The notation $c_{i,i+1}$ denotes the amount of units in the buffer $b_{i,i+1}$ between stations i and $i + 1$. Following Hillier and Boling (1966) we assume that the first worker is never starved and the last worker is never blocked, i.e., there is always a product available to be processed at the first station, and the last buffer space is unlimited. Following Powell and Schultz (2004 p. 1098), the production discipline is *blocking-after-servic*e. After completing her work on a job, a worker stops to work if there is no available space in the downstream buffer.

As discussed in Sect. 2.2, workers react to the processes around them and the behavior of their coworkers. They are motivated to avoid causing idleness in the line. Consequently, operators will work faster when their downstream buffer is likely to be empty and their downstream neighbor is likely to be starved. Similarly, a worker will work faster when her upstream buffer is likely to be full and her upstream coworker is likely to be blocked.

Each worker starts with a nominal processing time exponentially distributed with mean w_i, i.e., the processing time if the worker worked alone. Starting a new piece of work, the worker observes the contents of adjacent buffers and, if required, adjusts the mean of her processing time distribution. When she is still working on the product, she adjusts her processing speed if the status of the adjacent buffers changes. Work flows from worker 1 towards worker N. Thus, for worker i, worker $i - 1$ is her upstream neighbor and worker $i + 1$ is her downstream neighbor. Likewise, $\mathrm{b}_{i-1,i}$ is her upstream buffer and the $b_{i,i+1}$ her downstream buffer. Each buffer can hold up to C units of work-in-progress. We reduce the mean nominal processing time when the upstream buffer is likely to be full and the downstream buffer is likely to be empty. Powell and Schultz (2004) assume a nominal mean processing time of 2 time units per item for each worker. Since we will vary processing times at each station as well as the buffer capacities, we modify their speed-up mechanism in this model. The parameter f is the maximum processing time adjustment factor measured in percent. We assume that each worker will adjust with the same rate, e.g., 50 % of her nominal mean processing time. f ranges from 0 to 0.9. $f = 0$ is identical to the state-independent model.

The adjusted working time is

$$w_i^{adj} = w_i - f * \frac{w_i}{2} * \frac{c_{i-1,i}}{C} - f * \frac{w_i}{2} \frac{(C - c_{i,i+1})}{C},$$

where w_i^{adj} is the adjusted processing time at station i.

The mean nominal processing time w_i is reduced by f * 100 %. Because workers react evenly on both adjacent buffers we split the maximum possible adjustment by half. Further, the adjustment depends on the adjacent buffer sizes. For the downstream buffer the processing time is adjusted by $(C - c_{i,i+1})/C$, where $c_{i,i+1}$ is the current number of items in the buffer. The fewer items are in the buffer, the higher is the processing time adjustment. Finally, when the downstream buffer is full, $(C - c_{i,i+1})$ is zero and there is no adjustment to the processing time, so there is no adjustment for the downstream buffer.

For the upstream buffer the processing time is adjusted by $c_{i-1,i}/c$. When the upstream buffer is empty, we have $c_{i-1,i}/c = 0$, i.e., the processing time does not react on the upstream buffer. When the upstream buffer is full and the downstream buffer is empty, the worker achieves the maximal adjustment. In this case, a worker works at her highest speed, because both neighbors are likely to be idle or become idle soon.

Following Powell and Schultz (2004) we assume a quasi-continuous adjustment of processing times, i.e., a worker adjusts her speed while still working on an item, if the adjacent buffer states change. The state of a buffer changes if either the upstream worker passes on her finished item or the downstream worker takes an item off. These events trigger the following actions. If a new item enters buffer $b_{i,i+1,}$ it impacts the downstream worker, so that her adjusted remaining processing time is

$$w_{i+1}^{rem,adj} = w_{i+1}^{rem} - f * \left(\frac{1}{C}\right) * w_{i+1}^{rem},$$

where w_{i+1}^{rem} is the remaining time and $w_{i+1}^{rem,adj}$ the adjusted remaining time of worker $i + 1$. If an item leaves the buffer, the upstream worker adjusts her mean processing time according to

$$w_{i}^{rem,adj} = w_{i}^{rem} - f * \left(\frac{1}{C}\right) * w_{i}^{rem}.$$

Further, our model allows processing time reductions, but no increases. There is no empirical evidence of worker's slowing down (Schultz et al. 1998). Alternatively, we can reinterpret initial mean processing time so that the reinterpreted formula would capture worker's slowing down too.

The measure of performance is the mean output rate $R(\boldsymbol{w})$ where vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)$ denotes the initial work allocation.

4 Results

4.1 Simulation experiments

We build the simulation with AnyLogic available at http://www.xjtek.com and code human behavior in Java. This study addresses the optimal work allocation problem under assumption of state-dependent behavior. Thus, we look for the distribution of workload by which the productivity is optimized. We seek to

$$\text{maximize } R(\boldsymbol{w})$$

subject to $\sum_{i=1}^{N} w_i = N$, $w_i > 0$ for $i = 1, 2, \ldots N$, where $R(\boldsymbol{w})$ is the production rate, N the number of stations, and w_i the mean processing time at station i.

Our experiments are based on the study of Hillier and Boling (1966). They derive production rates for 2,3-station lines with buffer capacities from 0–4, for a 4-station line without buffers, while varying mean processing times at stations. We study line lengths of 3 and 4 and buffer capacities from 1–4. We do not consider 2-station lines and cannot obtain results for line settings without buffers, because we assume, that workers react on the buffer levels and then adjust their processing times.

The mean processing times are set as in Hillier and Boling (1966), but we additionally assume different speed up values f. The results of Hillier and Boling (1966) correspond to our state-dependent model with $f = 0$. Following Hillier and Boling (1966) we assume exponentially distributed processing times at stations. Each work allocation can be characterized by its degree of unbalance δ, calculated as

$$\delta = \frac{\sum_{i=1}^{N} |w_i - 1|}{N}.$$

For each line configuration we measure the production rate $R(\boldsymbol{w})$, which is the number of finished goods in a given time span.

We ran 10 repetitions, each of which spanned 600,000 time units for a given configuration. At the beginning of the simulation, the first worker starts on her first work piece and the others are idle. This state is called *initialization* or *transient state*. Initialization states can distort the results. This problem has been solved in different ways. Powell and Schultz (2004) follow Law and Kelton (1991) and start their simulation with half-full buffers. Hillier and So (1996) eliminate the data of the first production sequences. We follow the latter approach and eliminate the first 200,000 production units. Run times as well the transient state production units were determined empirically.

4.2 The general impact of state-dependent behavior on the work allocation

In this section, we investigate the general effect of adaptive behavior on the production rate and the corresponding work allocation. We concentrate on the effects of a "bowl phenomenon"-like work allocation. The number of stations $N = 3$ and buffer capacity $C = 1$. The work allocation is symmetric, i.e., the first and the last station have identical nominal mean processing times. As a normalization, the mean processing time of the second station is calculated as $w_2 = 3 - 2 * w_1$, so that $\sum w_i = 3 = N$. We vary the mean processing time at the first station in[0.02; 1.16] in steps of 0.01. Hillier and Boling (1966) provide results for variation of mean processing times in[1; 1.15] in steps of 0.01. The speed up parameter is increased from zero up to 0.9 in steps of 0.1. Each combination of work allocations and speed-up parameter corresponds to a production rate value.

Figure 1 shows production rate curves for $f \leq 0.5$. The horizontal axis depicts different initial work allocations represented by the processing time assigned to the first station. The point 1.06 denotes $w_1 = w_3 = 1.06$, and $w_2 = 0.88$. The lowest curve is the production rate function for the model without adaptive behavior. Each production rate function has a unique maximum. The production rate increases with increasing speed-up parameter independently of the work allocation. That is plausible, because of, first, the entire cycle time reduces and second, the idleness is reduced. Independent of the underlying speed-up parameter, the output for a given work allocation is always higher than predicted by the state-independent model. Lines with lower mean processing times on the ends benefit from workers' adaptive behavior. Boundary stations induce pressure on the middle stations so that they begin to work faster. The production rate increases. A line configuration due to bowl phenomenon would not seize a chance to use this potential output gain.

Assuming that workers adjust their processing times by some maximal value f, the best found work allocation does not remain the same. In the state-independent model, the best work allocation follows a bowl-shape with $w_1 = 1.08$, $w_2 = 0.84$, $w_3 = 1.08$. In models with mean processing time adjustment, this allocation would not deliver the maximum

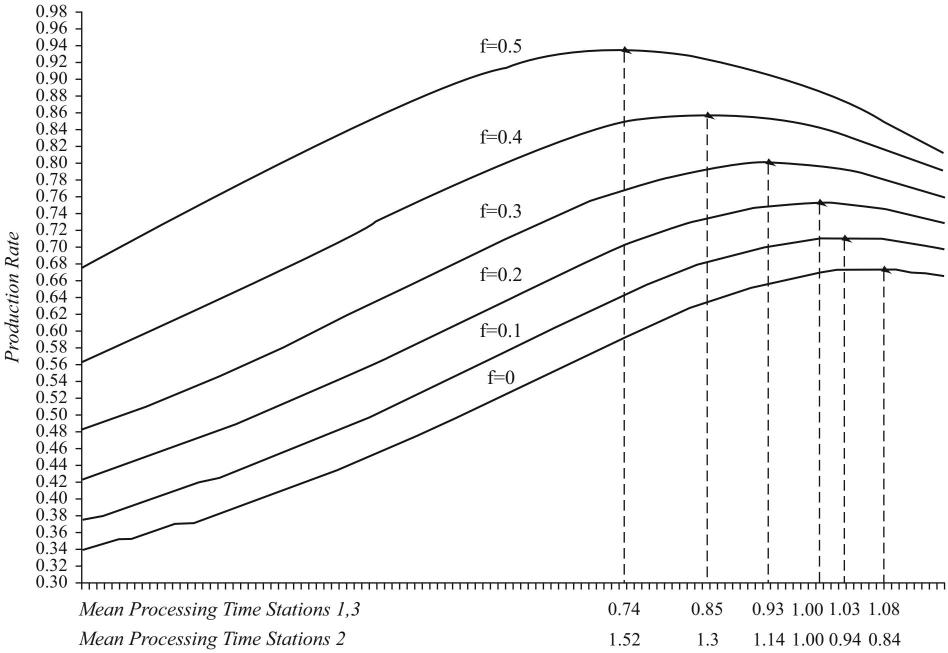

Fig. 1: Optimal work allocation for $\boldsymbol{N=3}$, $\boldsymbol{C=1}$

possible output. For example, the maximum production rate for $f=0.1$ is achieved by $w_1=1.03$, $w_2=0.94$, $w_3=1.03$. By increasing f, the optimal work assignment moves towards a balanced line (i.e., $w_1=w_2=w_3=1$) and for $f>0.2$ represents a reversed bowl. Figure 2 shows the optimal work allocations for different speed-up parameters in a different way. At this point, we conclude that state-dependent behavior has an impact on the work allocation.

Table 1 depicts the optimal allocations for different speed-up parameters and corresponding degrees of unbalance δ. If the workers can adjust their processing speed by 20 %, the optimal $\delta=0$, indicating that *in some configurations a balanced line can definitely be the best work allocation*. Powell and Schultz (2004) consider a speed up of ≈ 20 % realistic.

The individual magnitude of workers' processing time adjustments are not observable in practice. In what direction should a line designer unbalance a line? Table 1 indicates that the optimal degree of unbalance is a convex function of the underlying model speed-up parameter. For $f \leq 0.2$ the direction of unbalance is to assign the work in a "bowl shape", for $f>0.2$ the maximum production rate is achieved by a work assignment in a "reversed bowl pattern". *Without knowledge of the real speed-up parameters and individual behaviors, production line models that use the independence assumption can significantly misestimate the optimal work allocation*.

Figure 1 indicates that for small degrees of unbalance the output is higher than predicted by the model with independence assumption. A real production line is a kind of

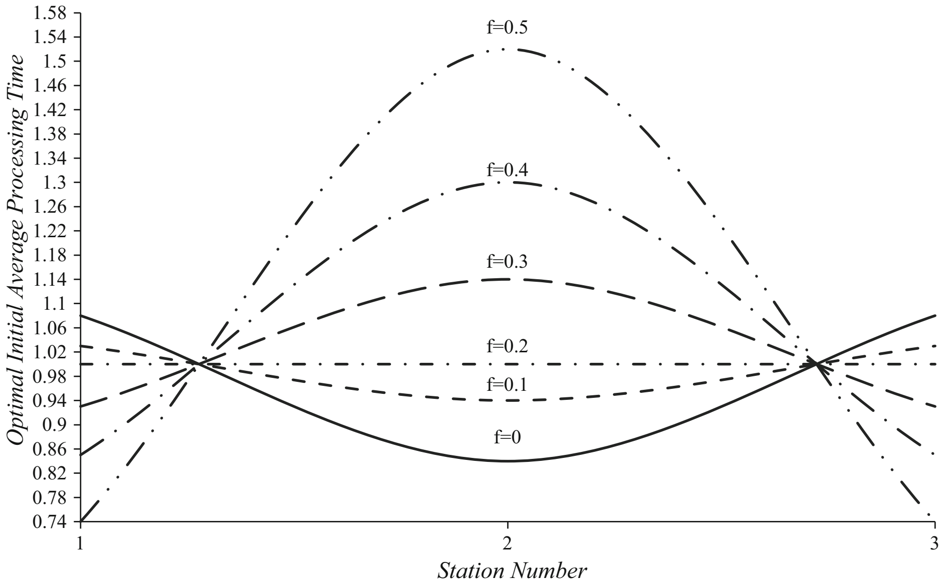

Fig. 2: Optimal work allocations for different speed-up parameters

Table 1: Degrees of unbalance $N = 3$, $C = 1$

w_i	f 0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
w_1	1.08	1.03	1	0.93	0.85	0.74	0.58	0.36	0.18	0.03
w_2	0.84	0.94	1	1.14	1.3	1.52	1.84	2.28	2.64	2.94
w_3	1.08	1.03	1	0.93	0.85	0.74	0.58	0.36	0.18	0.03
δ	0.11	0.04	0	0.09	0.20	0.35	0.56	0.85	1.09	1.29

"black-box". A line designer can only decide how to allocate the work and observe the output of the production system. Under certain conditions, however, he does not observe any difference in output, regardless of whether the line is unbalanced in "bowl" or in "reversed bowl" shape. For example, assuming that the real underlying speed-up parameter is 0.2, assignments $\boldsymbol{w}' = (0.93, 1.14, 0.93)$ and $\boldsymbol{w}'' = (1.06, 0.88, 1.06)$ have approximately the same degree of unbalance $\delta' = 0.08$, $\delta'' = 0.09$ and deliver similar output rates $R(w') = 0.7493$, $R(w'') = 0.7492$. The optimum is, however, a balanced line. This example illustrates, that adaptive behavior complicates the determination of the optimal work allocation.

4.3 The impact of increasing buffer capacity

Buffer capacities were constant in Sect. 4.2. We now vary them from 1 to 4. For fixed capacity and speed up parameters, Table 2 shows optimal work allocations $w' = (w_1, w_2, w_3)$, degrees of unbalance δ and production rates $R(w')$. The column "improvement" depicts the improvement over the output of a line with $C = 1$. The last column describes the type of allocation (bowl, balanced, or reversed bowl).

Table 2: Optimal work allocations for different values of f and C, $N = 3$

f	C	$R(w')$	Improvement (%)	w_1	w_2	w_3	δ	Allocation type
0	1	0.6737	0.0	1.08	0.84	1.08	0.11	Bowl
	2	0.7370	9.4	1.06	0.88	1.06	0.08	Bowl
	3	0.7796	15.7	1.04	0.92	1.04	0.05	Bowl
	4	0.8099	20.2	1.02	0.96	1.02	0.03	Bowl
0.1	1	0.7116	0.0	1.03	0.94	1.03	0.04	Bowl
	2	0.7895	10.9	1.00	1.00	1.00	0.00	Balanced
	3	0.8268	16.2	1.01	0.98	1.01	0.01	Bowl
	4	0.8577	20.5	1.01	0.98	1.01	0.01	Bowl
0.2	1	0.7529	0.0	1.00	1.00	1.00	0.00	Balanced
	2	0.8481	12.7	1.00	1.00	1.00	0.00	Balanced
	3	0.8791	16.8	0.99	1.02	0.99	0.01	Rev. bowl
	4	0.9105	20.9	1.00	1.00	1.00	0.00	Balanced
0.3	1	0.8006	0.0	0.93	1.14	0.93	0.09	Rev. bowl
	2	0.9148	14.3	0.95	1.10	0.95	0.07	Rev. bowl
	3	0.9379	17.1	0.97	1.06	0.97	0.04	Rev. bowl
	4	0.9679	20.9	0.98	1.04	0.98	0.03	Rev. bowl
0.4	1	0.8580	0.0	0.85	1.30	0.85	0.20	Rev. bowl
	2	0.9908	15.5	0.94	1.12	0.94	0.08	Rev. bowl
	3	1.0040	17.0	0.94	1.12	0.94	0.08	Rev. bowl
	4	1.0324	20.3	0.95	1.10	0.95	0.07	Rev. bowl
0.5	1	0.9350	0.0	0.74	1.52	0.74	0.35	Rev. bowl
	2	1.0796	15.5	0.88	1.24	0.88	0.16	Rev. bowl
	3	1.0782	15.3	0.92	1.16	0.92	0.11	Rev. bowl
	4	1.1053	18.2	0.92	1.16	0.92	0.11	Rev. bowl
0.6	1	1.0551	0.0	0.58	1.84	0.58	0.56	Rev. bowl
	2	1.1837	12.2	0.83	1.34	0.83	0.23	Rev. bowl
	3	1.1644	10.4	0.88	1.24	0.88	0.16	Rev. bowl
	4	1.1881	12.6	0.89	1.22	0.89	0.15	Rev. bowl
0.7	1	1.2734	0.0	0.36	2.28	0.36	0.85	Rev. bowl
	2	1.3107	2.9	0.81	1.38	0.81	0.25	Rev. bowl
	3	1.2643	− 0.7	0.84	1.32	0.84	0.21	Rev. bowl
	4	1.2840	0.8	0.85	1.30	0.85	0.20	Rev. bowl
0.8	1	1.7633	0.0	0.18	2.64	0.18	1.09	Rev. bowl
	2	1.4597	− 17.2	0.76	1.48	0.76	0.32	Rev. bowl
	3	1.3812	− 21.7	0.81	1.38	0.81	0.25	Rev. bowl
	4	1.3833	− 21.6	0.83	1.34	0.83	0.23	Rev. bowl
0.9	1	3.3740	0.0	0.03	2.94	0.03	1.29	Rev. bowl
	2	1.6487	− 51.1	0.72	1.56	0.72	0.37	Rev. bowl
	3	1.5167	− 55.0	0.76	1.48	0.76	0.32	Rev. bowl
	4	1.5119	− 55.2	0.80	1.40	0.80	0.27	Rev. bowl

Hillier and Boling (1966) find that as C increases, the extent of potential improvement over the balanced line decreases and the bowl phenomenon becomes less pronounced: the degree of unbalance diminishes (see Hillier and Boling 1966, p. 655). Table 2 confirms their findings in models with independence assumption. In models with state-dependent behavior the effect of increasing buffer capacities on the productivity and optimal work allocation is not straightforward to understand. Therefore, we distinguish between speed-up intensities. First, let us consider work allocations:

- $f < 0.2$: The work allocation is the "bowl" type.
- $f = 0.2$: A balanced line is the best allocation.
- $f > 0.2$: A production line should be unbalanced in a "reversed bowl" shape.

Table 2 indicates that with increasing buffer capacity the degree of unbalance diminishes for most cases, i.e., the optimal processing time configuration moves toward the balanced line. Next, we consider production rates:

- $f \leq 0.6$: Productivity increases with buffer capacity but with decreasing returns. The largest improvement is achieved by changing C from 1 to 2.
- $f = 0.7$: A buffer capacity of two leads to the largest output gain of 2.9 %. $C = 3$ delivers a deterioration of output, while $C = 4$ leads to smaller output gain.
- $f > 0.7$: The largest improvement is achieved by a buffer capacity of one. Increasing buffer capacity reduces the output rate.

Figure 3 depicts the interrelation between C, f and R. Speed-up and providing buffer capacity have positive effects on productivity. Speed up has two effects: first, the output increases because the whole cycle time reduces; second, the idle time decreases and leads to more output. Additional buffer capacity makes two adjacent stations more independent of each other; therefore, it reduces the idle time in the line. These conclusions suggest that more buffer capacity under the assumption of adaptive behavior would lead to increased productivity. However, we cannot observe this for $f \geq 0.7$. Moreover, we observe that output curves reach their maximum at $C = 2$ ($f = 0.7$), or $C = 1$ ($f > 0.7$) and then decrease. One possible explanation is the negative relation between speed up and buffer capacity. The more buffer capacity is provided, the less potential exists for speeding, the less pressure in the line which leads to the adjustment of the processing times. The effect of speeding up is higher than of providing additional buffer capacity. We hypothesize that if $C = 1$ speeding up achieves its maximum effect, and if $C > 1$ the negative effect of supplying additional buffer capacity outweighs the positive effect.

We conjecture that buffer capacity and speeding up in conjunction provide an additional design factor in serial production lines. A line designer should find an appropriate amount of interstage buffer capacities which triggers a natural mechanism of adaptive human behavior.

4.4 The impact of increasing line length

As discussed in Sect. 2.1, an increase in line length decreases productivity due to an increase in stochastic interference. This section investigates the effects of line length in

Fig. 3: The effects of speeding up and additional buffer capacity on the productivity

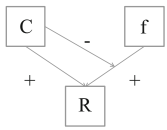

Table 3: Production rate differences between balanced 3 and 4-station lines, $C = 1$

f	Mean R_3	sd R_3	Mean R_4	sd R_4	Production rate decrease (%)	p-Value
0	0.6701	0.0007	0.6313	0.0009	− 6.15	0.00
0.1	0.7106	0.0010	0.6749	0.0008	− 5.30	0.00
0.2	0.7529	0.0007	0.7225	0.0009	− 4.20	0.00
0.3	0.7963	0.0006	0.7722	0.0012	− 3.12	0.00
0.4	0.8406	0.0011	0.8226	0.0009	− 2.19	0.00
0.5	0.8827	0.0014	0.8716	0.0009	− 1.27	0.00
0.6	0.9221	0.0017	0.9173	0.0011	− 0.52	0.00
0.7	0.9544	0.0024	0.9538	0.0014	− 0.07	0.48
0.8	0.9785	0.0016	0.9791	0.0015	0.06	0.42
0.9	0.9947	0.0010	0.9941	0.0015	− 0.06	0.29

models with state-dependent behavior. We extend the production line to 4 stations. The mean processing time at the first and last stations is ranged in the interval $[0.04; 1.14]$ in steps of 0.04. The interval for mean processing times differs from our setting with 3 stations due to computational issues, these implementation differences have no impact on the results. Hillier and Boling (1966) provide results for mean processing time settings in the interval $[1; 1.20]$ in steps of 0.02. Processing times for middle stations are $(4 - 2 * w_1)/2$. $C \in 1,2,3,4$ and f varies from 0 to 0.9 in steps of 0.1.

First, we consider the impact on the productivity. Table 3 shows production rates for 3 and 4-station lines, denoted as R_3 and R_4. With purpose to obtain undistorted effects of increasing line length, we compare two line settings which differ from each other only in the number of stations, so we set $C = 1$ and consider balanced work allocation, i.e., $w_i = 1 \forall i$. The production rate decrease is calculated as $1 - R_4/R_3$. For speed-up parameters $f \leq 0.7$, the output decreases with the number of stations, for $f > 0.7$ there are no statistically significant differences in output rates, as indicated by p-values for paired Student t-test in Table 3. *Therefore, state-dependent behavior offers opportunities to counteract negative effects of stochastic interference which enlarges with increasing number of stations*. These results are similar to findings of Powell and Schultz (2004).

Next, we study how optimal work allocations depend on the line length. Table 4 shows optimal work allocations for each combination of speed-up parameter and buffer capacity. Similar to findings in the 3-station line, for $f < 0.2$ the pattern of the optimal work assignment is a bowl, for $f = 0.2$ a balanced line is the better one, for $f > 0.2$ a production line should be unbalanced in reversed bowl shape. Columns δ_3 and δ_4 show the average amounts of unbalance in the optimal allocations in 3 and 4-station lines with the same parameters of C and f. Similar to the 3-station line, the degree of unbalance in the 4-station line decreases with increasing buffer capacity. Since $\delta_3 \in [0, 4/3]$ and $\delta_4 \in [0; 1]$, we cannot identify effects of increasing line length on the degree of unbalance, by comparing both values directly. Therefore, δ_3 is standardized as $\tilde{\delta}_3 = \delta_3 * 3/4$. The last column in Table 4 gives absolute differences in degrees of unbalance in optimal allocations for $f = 3,4$ and is calculated as $\Delta = \tilde{\delta}_3 - \delta_4$.

Hillier and Boling (1966) postulate that the optimal degree of unbalance remains about the same as N increases (Hillier and Boling 1979, pp. 723–724). Our experimental results for the state-independent model ($f = 0$) confirm their finding: the absolute difference is

Table 4: Optimal work allocations for different values of f and C, $N = 4$

C	f	w_1	w_2	w_3	w_4	Allocation	δ_4	δ_3	$\tilde{\delta}_3$	Δ
1	0	1.1	0.9	0.9	1.1	Bowl	0.1	0.11	0.08	0.02
2		1.06	0.94	0.94	1.06	Bowl	0.06	0.08	0.06	0.00
3		1.04	0.96	0.96	1.04	Bowl	0.04	0.05	0.04	0.00
4		1.04	0.96	0.96	1.04	Bowl	0.04	0.03	0.02	0.02
1	0.1	1.04	0.96	0.96	1.04	Bowl	0.04	0.04	0.03	0.01
2		1.04	0.96	0.96	1.04	Bowl	0.04	0.00	0.00	0.04
3		1.02	0.98	0.98	1.02	Bowl	0.02	0.01	0.01	0.01
4		1.02	0.98	0.98	1.02	Bowl	0.02	0.01	0.01	0.01
1	0.2	1	1	1	1	Balanced	0	0.00	0.00	0.00
2		0.98	1.02	1.02	0.98	Rev. bowl	0.02	0.00	0.00	0.02
3		1	1	1	1	Balanced	0	0.01	0.01	− 0.01
4		0.98	1.02	1.02	0.98	Rev. bowl	0.02	0.00	0.00	0.02
1	0.3	0.94	1.06	1.06	0.94	Rev. bowl	0.06	0.09	0.07	− 0.01
2		0.96	1.04	1.04	0.96	Rev. bowl	0.04	0.07	0.05	− 0.01
3		0.96	1.04	1.04	0.96	Rev. bowl	0.04	0.04	0.03	0.01
4		0.96	1.04	1.04	0.96	Rev. bowl	0.04	0.03	0.02	0.02
1	0.4	0.84	1.16	1.16	0.84	Rev. bowl	0.16	0.20	0.15	0.01
2		0.9	1.1	1.1	0.9	Rev. bowl	0.1	0.08	0.06	0.04
3		0.92	1.08	1.08	0.92	Rev. bowl	0.08	0.08	0.06	0.02
4		0.90	1.1	1.1	0.9	Rev. bowl	0.1	0.07	0.05	0.05
1	0.5	0.72	1.28	1.28	0.72	Rev. bowl	0.28	0.35	0.26	0.02
2		0.86	1.14	1.14	0.86	Rev. bowl	0.14	0.16	0.12	0.02
3		0.88	1.12	1.12	0.88	Rev. bowl	0.12	0.11	0.08	0.04
4		0.90	1.1	1.1	0.9	Rev. bowl	0.1	0.11	0.08	0.02
1	0.6	0.56	1.44	1.44	0.56	Rev. bowl	0.44	0.56	0.42	0.02
2		0.8	1.2	1.2	0.8	Rev. bowl	0.2	0.23	0.17	0.03
3		0.84	1.16	1.16	0.84	Rev. bowl	0.16	0.16	0.12	0.04
4		0.84	1.16	1.16	0.84	Rev. bowl	0.16	0.15	0.11	0.05
1	0.7	0.4	1.6	1.6	0.4	Rev. bowl	0.6	0.85	0.64	− 0.04
2		0.76	1.24	1.24	0.76	Rev. bowl	0.24	0.25	0.19	0.05
3		0.80	1.2	1.2	0.8	Rev. bowl	0.2	0.21	0.16	0.04
4		0.80	1.2	1.2	0.8	Rev. bowl	0.2	0.20	0.15	0.05
1	0.8	0.2	1.8	1.8	0.2	Rev. bowl	0.8	1.09	0.82	− 0.02
2		0.68	1.32	1.32	0.68	Rev. bowl	0.32	0.32	0.24	0.08
3		0.76	1.24	1.24	0.76	Rev. bowl	0.24	0.25	0.19	0.05
4		0.76	1.24	1.24	0.76	Rev. bowl	0.24	0.23	0.17	0.07
1	0.9	0.08	1.92	1.92	0.08	Rev. bowl	0.92	1.29	0.97	− 0.05
2		0.64	1.36	1.36	0.64	Rev. bowl	0.36	0.37	0.28	0.08
3		0.68	1.32	1.32	0.68	Rev. bowl	0.32	0.32	0.24	0.08
4		0.72	1.28	1.28	0.72	Rev. bowl	0.28	0.27	0.20	0.08

about zero. In state-dependent cases, all differences do not exceed 0.1 but take on different values, so that we cannot see any interrelation. *With increasing line length the effect of adaptive behavior on the optimal degree of unbalance is unclear*.

5 Conclusions and further research

The primary goal of our study is to gain insights into the effects of complex human behavior on the performance of a production line. Thus we contribute to behavioral operations, a relatively novel scientific stream considering behavioral issues in real operating systems, (see, e.g., Bendoly et al. 2006; Gino and Pisano 2008). Bendoly et al. (2006) allude to a gap between theoretical concepts and practice (Bendoly et al. 2006, p. 737). They point out that unconsidered behavioral issues impact negatively the operational success.

Our study confirms this deliberation and shows that adaptive behavior has a significant impact on optimal work allocation. We consider adaptive behavior in production lines with variable human operation times. Variability is unavoidable in human operating times and causes idleness and throughput decrease. Some improvement can be achieved by intelligent work allocation. Hillier and Boling (1966) find that the best way to assign work in a line is characterized by a "bowl", which implicates the allocation of more work to the end and the beginning of a line. Their seminal work and the follow-up studies assume that the means of the operation time distributions are independent and remain constant over time. Some recent results indicate that production line workers adjust their work rates with the goal to prevent idleness of their coworkers. We integrate this finding into a simulation model and examine how work should be allocated in this line.

Experimental evidence suggests that adaptive behavior has a positive impact on the production rate. For all adjustment parameters the output is higher than predicted by the models with independence assumption. Adaptive behavior compensates for production losses which occur due to blocking and starving. We find that high-quality work allocations in state-dependent models need not match the allocations predicted by state-independent models. Moreover, the maximum production rate is achieved by different types of allocations: bowl, balanced, or reversed bowl. The allocation type depends on the underlying speed-up parameter. We find that the "bowl phenomenon" holds only if the underlying speed-up parameter is less than 20 %. If the maximal real adjustment parameter is about 20 % the balanced line promises the best output rate.

The degree of unbalance diminishes with increasing buffer capacity. The relation between line length and degree of unbalance remains unclear. The combination of buffer capacity and of the speeding up parameter is a system-design variable. The extent of processing time adjustment depends on the buffer capacity. In summary, without knowledge of the real speed-up parameters and individual behaviors, production line models that use the independence assumption can significantly misestimate the optimal work allocation.

Adaptive behavior can counteract predicted negative effects of lower buffer capacities, or of in- creasing number of stations. Knowing that workers behave adaptively at the workplace, line engineers and management can adapt their design principles. Lower buffer capacities enhance even the magnitude of the speeding up, which leads to more production rate. Therefore, the line designer can deliberately design the production line with lower buffer capacities anticipating that workers will feel more pressure and will adjust their working pace.

Our work is not without limitations. We assume that all workers are similar in their behavior. However, people are differently sensible to the environment. One person can react more strongly to avoid the cause of idleness, whereas another person does not care about his coworkers and does not adjust his working pace at all. The very recent study of Schultz et al. (2010) found that individuals have indeed different speed up behavior: Some workers adjust their operation time very strongly, some do not. We consider the integration of these results into models of production lines a fruitful area of future research.

References

Adams SJ (1963) Toward an understanding of inequity. J Abnorm Soc Psychol 67:422–436
Atkinson JW (1964) Introduction to motivation. Van Nostrand, Princeton
Bendoly E, Donohue K, Schultz KL (2006) Behavior in operations management: assessing recent findings and revisiting old assumptions. J Operat Manage 24:737–752
Carnall CA, Wild R (1976) The location of variable work stations and the performance of production flow lines. Int J Prod Res 14(6):703–710
Dallery A, Gershwin SB (1992) Manufacturing flow line systems: a review of models and analytical results. Queueing Syst 12:3–94
Doerr KH et al (1996) Impact of material flow policies and goals on job outcomes. J Appl Psychol 81:142–152
Edie LC (1954) Traffic delays at toll booths. J Oper Res Soc Am 2(2):107–138
El-Rayah TE (1979) The efficiency of balanced and unbalanced production lines. Int J Prod Res 17(1):61–75
Falk A, Ichino A (2006) Clean evidence on peer effects. J Labor Econ 24(1):39–57
Fandel G (2005) Produktion I: Produktions- und kostentheorie. Springer Verlag, Berlin
Franks IT, Sury RJ (1966) The performance of operators in conveyor-paced work. Int J Prod Res 5(2):97–112
Gino F, Pisano G (2008) Toward a theory of behavioral operations. Manuf Serv Operat Manage 10(4):676–691
Hillier FS, Boling R (1966) The effect of some design factors on the efficiency of production lines with variable operation times. J Ind Eng 17(12):651–658
Hillier FS, Boling RW (1979) On the optimal allocation of work in symmetrically unbalanced production line systems with variable operation times. Manage Sci 25(8):721–728
Hillier FS, So KC (1993) Some data for applying the bowl phenomenon to large production line systems. Int J Prod Res 31(4):811–822
Hillier FS, So KC (1996) On the robustness of the bowl phenomenon. Eur J Oper Res 89:496–515
Hunt GC (1956) Sequential arrays of waiting lines. Oper Res 4(6):674–683
Kala R, Hitchings GG (1973) The effects of performance time variance on a balanced, four-station manual assembly line. Int J Prod Res 11(4):341–353
Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: an econometric analysis of hospital operations. Manage Sci 55:1486–1498
Latané B, Kipling W, Harkins S (1979) Many hands make light the work: the causes and consequences of social loafing. J Pers Soc Psychol 37(6):822–832
Latham GP, Locke EA (1991) Self-regulation through goal setting. Organ Behav Hum Decis Process 50:212–247
Law AM, Kelton WD (1991) Simulation Modeling and Analysis, 2nd edn. McGraw-Hill, New York
Locke EA, Latham GP (1990) A theory of goal setting and task performance. Prentice Hall, Engelwood Cliffs
Mas A, Moretti E (2009) Peers at work. Am Econ Rev 99(1):112–145
Payne S, Slack N, Wild R (1972) A note on the operating characteristics of 'balanced' and 'unbalanced' production flow lines. Int J Prod Res 10(1):93–98
Pike R, Martin GE (1994) The bowl phenomenon in unpaced lines. Int J Prod Res 32(3):483–499

Powell SG, Schultz KL (2004) Throughput in serial lines with state-dependent behavior. Manage Sci 50(8):483–499

Rao NP (1976) A generalization of the 'bowl-phenomenon' in series production systems. Int J Prod Res 14(4):437–443

Schultz KL, Juran DC, Boudreau JW (1999) The effects of low inventory on the development of productivity norms. Manage Sci 45(12):1664–1678

Schultz KL, Schoenherr KL, Nembhard D (2010) An example and a proposal concerning the correlation of worker processing times in parallel tasks. Manage Sci 56(1):176–191

Schultz KL et al (1998) Modeling and worker motivation in JIT production systems. Manage Sci 44(12):1595–1607

Steiner ID (1972) Group process and productivity. Academic Press, New York

Thompson WW, Burford RL (1988) Some observations on the bowl phenomenon. Int J Prod Res 26(8):1367–1373