

BURNOUT AND FLOATING GOALS IN HIGH-CONTACT SERVICE OPERATIONS

Luis López

INCAE Business School
Alajuela, Costa Rica
Tel: +506 2437-2061
Luis.Lopez@incae.edu

Roy Zúñiga

INCAE Business School
Alajuela, Costa Rica
Tel: +506 2437-2256
Roy.Zuniga@incae.edu

Abstract

This paper explores behavioral issues associated to the management of high-contact service operations. In this type of operations there is a tension between managerial target-setting and the well-being of service agents. Target setting and monitoring to maintain overall efficiency, resource utilization, and output rate often leads to burnout and high attrition rates. This paper looks at how workloads and target performance metrics are adjusted in a service operation and explores the interaction of the mechanisms associated to the management of these goals with burnout and attrition. The paper finds that a simple linear relationship between resource utilization, burnout, and attrition is insufficient to explain observed data. The paper proposes that a feedback non-linear structure is better suited to explore those issues. The proposed feedback structure takes into account agent learning, resource utilization, human agent expectations, and target workload and performance goals. The article explores these issues in the context of a case study of a large high contact service operation.

INTRODUCTION

Facing very competitive environments, many high-contact firms must keep costs under control in order to remain competitive. This is particularly true for firms that sell such high-contact services for outsourcing. Service firms' ability to compete largely depends on how efficiently their resources are managed. Call centers are a good example. According to Armenia, Saullo, & Sedehi (2006), call centers represent, due to their intrinsic dynamics and to the unpredictable behavior of input data, one of the most difficult services to understand and manage. In services such as call centers, customer contracts translate time inefficiencies into non-reimbursable out of pocket costs for the service provider. Some contracts stipulate clear performance metrics and built-in financial penalties, associated to deviations from those stipulations, for the service provider.

Many high-contact contracted service operations need to compete on costs, and managerial goals are very much centered on cost minimization. Because most costs are associated to personnel, which can easily account for 60 to 70% of total costs (Zeynep, Armony & Mehrotra; 2007), and given that high resource utilization implies lower costs per job performed, cost minimization in these environments means managing the utilization of human resources, or, more directly, increasing resource utilization as much as possible.

Increasing resource utilization boundlessly, however, is scarcely possible for several reasons. First, as utilization increases in the presence of inter-arrival and processing time variability, customer waits increase non-linearly. Second, as pressure mounts, for a given level of capacity, agents cut corners, and corner-cutting degrades quality (Oliva, 2001; Oliva and Sterman, 2001). A third, less documented aspect of increasing resource utilization is the repercussions associated to the most important firm resource: the people. Sustained levels of high pressure work over a long enough run may lead to fatigue and an increased predisposition to leave the organization (Maslach, Schaufeli, & Leiter, 2001).

High levels of attrition impair a firm's ability to grow, because people have to be hired to replace those who leave and more need to be hired to increase the stocks of personnel and attain desired levels of growth. Such hiring is costly; it implies sustaining a large organizational structure devoted to recruiting, selecting, and hiring personnel. Hiring also implies huge expenses in training personnel as newcomers must be put through preparation courses in order to learn the basics of the trade. High levels of attrition also mean that the system, as a whole, is less productive because it takes time for workers, however well-trained, to become proficient and up to speed. In the time it takes to attain full proficiency, interactions between inexperienced personnel and customers are more prone to error.

Because of these reasons, managers in many high-contact service operations strive to attain levels of resource utilization that are "right. Bivona (2007) points out that call center managers have to deal simultaneously with conflicting goals, such as maintaining high customer service quality, high employee motivation, and keeping costs low. As Zeynep et al. (2007) show, operational decisions to increase efficiency also affect employees. Yet, the interactions between efficiency oriented measures are seldom viewed in the context of behavioral issues within the operation. Extant related papers typically focus either on operational aspects or on organizational aspects, but not both. As Aksin et al. (2007) indicate,

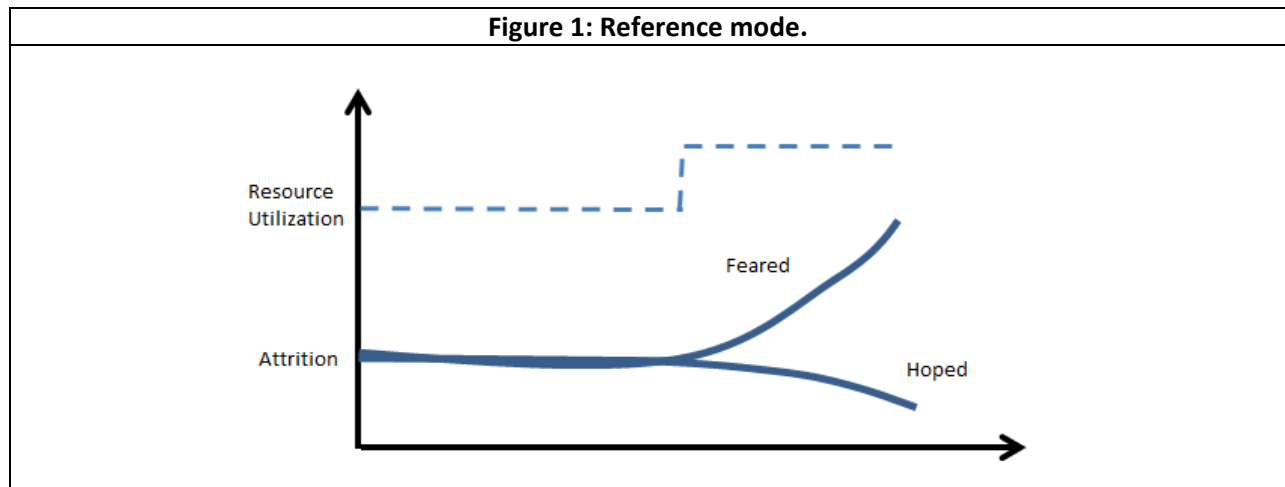
the operations management literature focuses on identifying and understanding operational planning problems such as staffing, scheduling, roster control, and the like. All these aspects have an influence, together with incentives, on employees, but such influence is “typically ignored in operations management models” (Aksin et al.; 2007:678).

Efficiency-oriented measures include increasing utilization as much as possible in order to attain efficiency and low costs without incurring in the costs associated to burnout and personnel attrition. In other words, managers strive to find the “right” level of resource utilization (normally as high as possible) that will lead to the “right” level of burnout and which, in turn, will result in the “right” (low) levels of personnel attrition. Such thinking is, schematically, as follows:

Resource utilization → Burnout → Attrition

According to this linear thinking, resource utilization can be increased up to a certain level that is considered to be optimal. Though intuitively appealing, this type of thinking fails to account for the adaptive non-linear nature of organizational systems. People within such systems exhibit adaptive behavior, and such adaptive behavior may defeat managerial intentions. Moreover, this adaptive non-linear behavior of organizations may blur the real causal structure that leads to attrition, thus lowering the effectiveness of managerial interventions to control or reduce attrition.

With this mental model, managers would, for a given level of occupancy, seek to keep attrition down while increasing resource utilization. Figure 1 shows the related reference mode.



This paper explores the dynamics of asset utilization and attrition in high contact service operations. Caramia, Armenia, Onori, & Giannunzio (2003) conclude that the System Dynamics approach helps improve the understanding of call center behaviors. We follow these authors’ suggestion and use the System Dynamics approach along with a preliminary case study from a large service operation to shed

light upon these issues. Our guiding research question is whether a right level of resource utilization can be achieved that does not lead to burnout and attrition. Although the results presented here are preliminary, this article finds that the conventional linear way of thinking about this problem may be misleading, and the results of managerial interventions may be counterintuitive. The paper unveils the causal structure associated to resource utilization and learning in the case study under scrutiny; a structure that may have several aspects in common to other high-contact service operations.

This paper is organized as follows: the next section explains the methodology; the following section –the data section– describes the case by looking at some descriptive data and statistics. With these data a dynamic hypothesis is offered in the following section. The next section explores a preliminary model formulation to study this issue. The last section concludes.

METHODS

This paper originates in the interest of a high contact service operator that has chosen to remain anonymous to establish the right level of resource utilization that will lead to optimality in terms of attrition and other system variables. The underlying research question is: ¿what is the relationship between burnout and resource utilization? The interest in this research question is based on the assumption that burnout is at the heart of attrition. Attrition in this company was becoming problematic because of a tightening labor market which made efforts to replace those who left the organization increasingly difficult.

The research question was operationalized by measuring resource utilization as the time an agent is working as compared to total time available. Burnout was measured periodically using a burnout scale. In addition, several control variables were also tracked. The study lasted 3 months. Table 1 lists the different control variables that were followed, among others. A set of interviews were also made with managers. These were unstructured interviews in which questions regarding the general characteristics of the operation were clarified. The instrument used to collect data was first pilot-tested on a limited basis for uncovering problems in administering and collecting the data. The instrument was thereupon administered to three accounts. In this preliminary study we report only one set of data for one of the accounts.

Table 1: Variables monitored in the case study.	
Variable	Description:
Scheduled Workload	The scheduled workload per agent.
Total available time	The total available time per agent.
Adherence	Time, as a percentage of the total available time, in which the agent was present.
Occupancy	Percentage of agent actually working over the total time.
Jobs	Total number of Jobs received by the person.
Average Work Time	Average time per job in which agent is doing something.
Average Interruption Time	Average time per job in which agent is not in contact with the customer and the customer is waiting for a response from the agent.
Average after job time	Average time after breaking off with customer that the agent spends doing related paperwork.
Average job time	Average total time per job
Wait time	Average time that the next customer is kept waiting after the agent has finished with the previous one.
Idle time	Idle time. The time the agent is not doing anything on the job.

Aside from tracking data about agents, the authors also had detailed discussions and interviews with the company executives. From these discussions we unveiled the managers' mental models, associated to the reference mode depicted in Figure 1. From the qualitative data it was possible to assemble a causal structure for the problem at hand which was corroborated, insofar as possible, with our quantitative information.

The following section presents a description of our preliminary results in the form of a case study, by highlighting the aspects that appear to form the causal structure associated to attrition in the organization.

CASE STUDY RESULTS

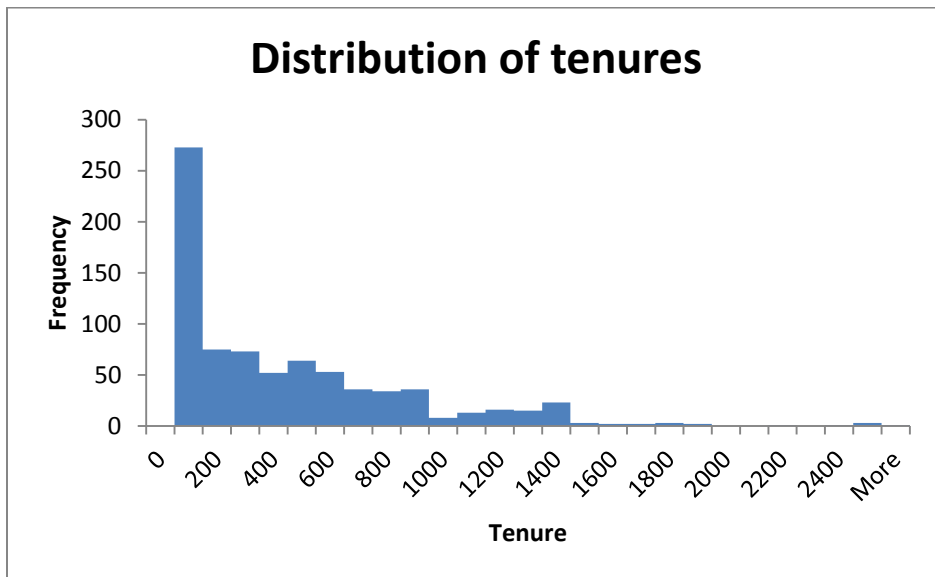
Results presented here are preliminary. At this point we are in the process of performing a more detailed statistical analysis of the collected data. Here we just present preliminary descriptive results.

The service operator will set up an entire operation for a particular customer. The contract with the customer may be structured in a number of different ways. Typically, the customer will prearrange a minimum level of staff utilization and a few measures of service quality, such as average job time. The service operator, in turn, has an incentive to improve upon the minimum levels specified in the contract in order to increase the profitability of the contract.

The variables in Table 1 were tracked for a period of 3 months for a large account in a service operation. During this time more than 700 service agents were employed in the tracked account for varying periods, though agents would not all be there at the same time because of different work shifts. Some agents were also hired during this period and some left the company within the period. Figure 2 shows the distribution of tenures at the beginning of the study, including repeated measures of tenure (in other words, Figure 2 shows the distribution of tenures for all the observations in the data set which is longitudinal).

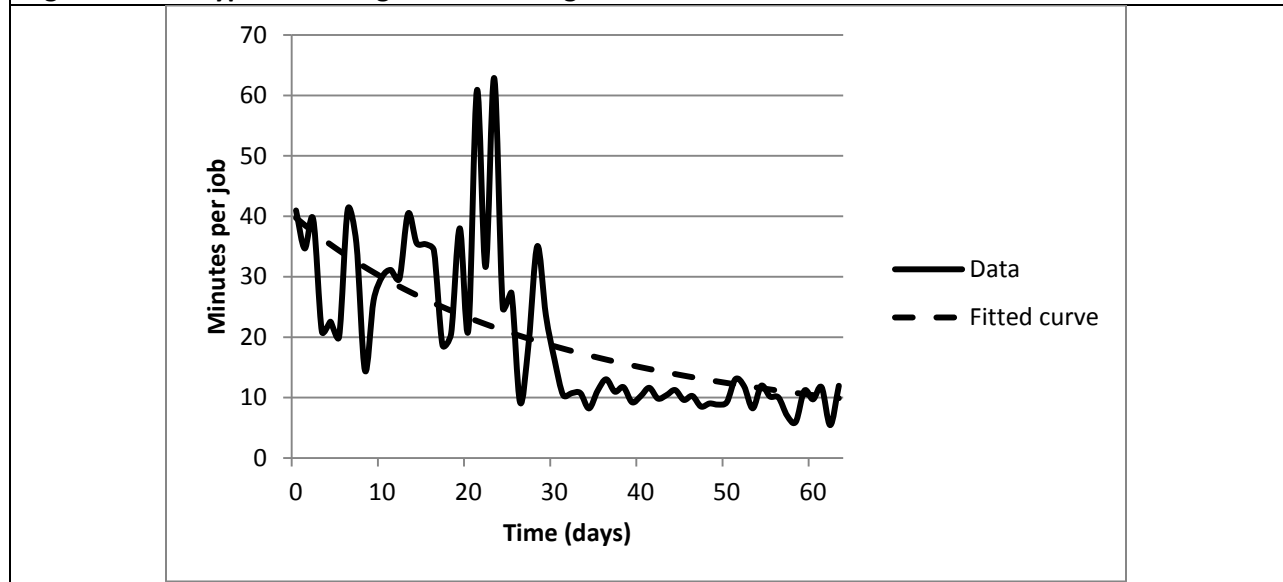
The majority of agents were relatively new. 35% of agents had been working for the company for less than 100 days and 53% for less than a year. 80% of the agents had worked for the company for less than 700 days or approximately two years. The distribution of tenures suggests high rates of attrition. The company had to sustain continuous hiring in order to staff the operation. With hiring, of course, training is needed. All agents in the account had to undergo several weeks of training prior to working in the account. The average tenure for all agents that worked in this account during the period studied was 488 days.

Figure 2: Tenure distribution



Agents go through a fairly brisk learning curve. Figure 3 shows a typical learning curve for an agent. The plot in Figure 3 shows an agent fresh out of training with 77 days of tenure in the firm. The plot shows how the average time used to handle a job diminishes sharply as time passes. The nature of the job remains the same throughout the period. A fitted curve of the form $y = Ax^n$ is also shown in Figure 3. The data suggest that as time doubles, the time required on average to do a job falls to 23% of its original value in this case, and the same is approximately true throughout the sample.

Figure 3: Prototypical learning curve for an agent



Managers, as is typical in high-contact service operations, take into account these differences in productivity for their staffing decisions. Agents with less tenure are scheduled less time, and as agents become more seasoned, the scheduled workload tends to increase. Thus, if a given workload for a day has to be apportioned among available agents, those agents with less experience are assigned a workload that is smaller than the average load. Figure 4 shows this difference in assigned workloads.

Judging by the median line in Figure 4 it appears that less experienced agents are assigned lighter workloads, and as agents acquire more experience the workload becomes larger. The largest workloads appear to be given to those between 100 and 200 days of tenure. The only statistically significant difference is between agents with small tenures of less than 100 days and all the rest. It appears that once people are up to speed, which occurs fairly rapidly due to the brisk learning curve, scheduled work is adjusted to take into account productivity.

Figure 4: Box Plot of scheduled workload in seconds

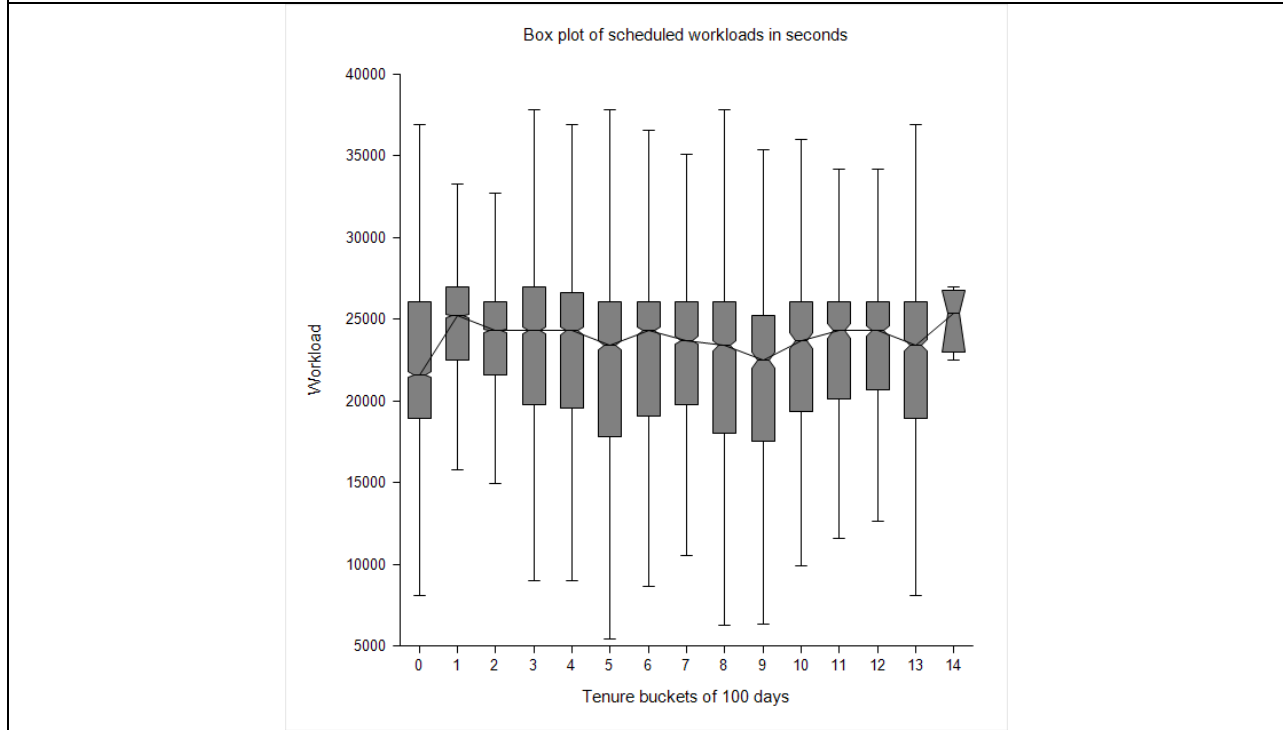
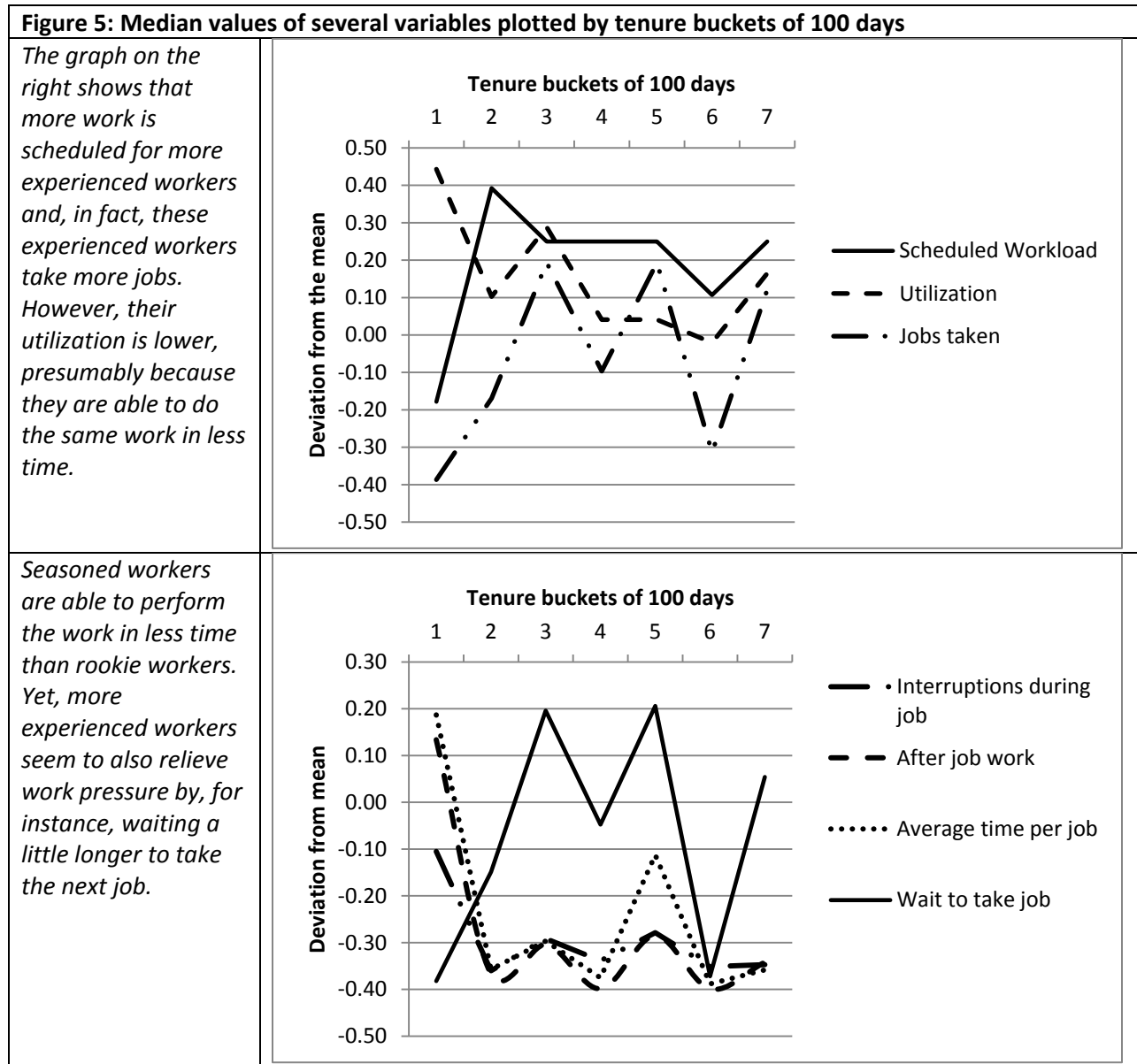


Table 2 summarizes median values for different standardized variables for 7 tenure buckets of 100 days, which means the data refers to all people in the sample who had worked for 2 years or less –approximately 80% of all agents. Table 2 reports the median to avoid distortions introduced by extreme outliers.

Table 2 : Medians for relevant variables by tenure

	Tenure up to (days)						
	100	200	300	400	500	600	700
Scheduled Workload	-0.18	0.39	0.25	0.25	0.25	0.11	0.25
Adherence	0.02	0.25	0.25	0.33	0.33	0.33	0.33
Utilization	0.44	0.10	0.29	0.04	0.04	-0.02	0.16
Jobs taken	-0.39	-0.17	0.19	-0.10	0.19	-0.31	0.12
Working during job	0.01	-0.31	-0.28	-0.34	-0.12	-0.35	-0.32
Interruptions during job	-0.10	-0.36	-0.29	-0.34	-0.28	-0.35	-0.35
After job work	0.13	-0.37	-0.30	-0.40	-0.29	-0.40	-0.34
Average time per job	0.19	-0.36	-0.30	-0.38	-0.11	-0.39	-0.36
Wait to take job	-0.38	-0.15	0.20	-0.05	0.21	-0.37	0.05

Table 2 shows that there are differences in several variables as agents become more experienced. As agents acquire more seniority, a larger workload is scheduled for them, and they indeed take more jobs on a per unit of time basis. However, more tenured personnel are able to perform the jobs in less time than their less seasoned counterparts. To highlight such differences Figure 5 exhibits groups of variables from Table 1 in graphical form. Each graph is described to its left of Figure 5.



Aside from having different mean values for the variables listed, less experienced workers are also less consistent in their responses. Table 3 shows the standard deviations of several relevant standardized variables. This table shows that experienced workers are more consistent in terms of the average times

required to perform a job, and also in terms of other variables such as the interruptions during each job and the work that has to be done after each contact with the customer. The variability in handling times is usually not considered when scheduling service systems, but it can add significantly to customer wait times and other measures of service quality.

Table 3 : Standard deviations for relevant variables by tenure

	Tenure up to (days)						
	100	200	300	400	500	600	700
Workload	0.77	0.77	0.90	1.15	1.16	1.19	1.06
Adherence	1.37	0.44	0.55	0.62	0.61	0.61	0.63
Utilization	0.97	0.98	0.88	0.98	1.03	1.09	1.03
Jobs taken	0.90	0.99	1.09	1.07	0.99	0.99	0.96
Working during job	1.18	0.95	0.97	0.96	0.99	0.93	0.98
Interruptions during job	1.53	0.77	0.77	0.72	0.85	0.74	0.70
After job work	1.56	0.64	0.61	0.86	0.81	0.84	0.62
Average time per job	1.21	0.91	0.94	0.93	0.96	0.91	0.93
Wait to take job	0.90	1.00	1.09	1.07	0.99	0.99	0.95

In Table 2 and in the related graphs we observe that there are important differences between newbies and seasoned workers. It appears that workloads are adjusted in accordance with agent learning. Qualitative data from interviews seems to support this impression. Because certain metrics are defined in the customer contract, metrics are adjusted to adhere to the contracted metric. As one manager said: “... Internally we strategically adjust goals to guarantee that we are able to reach the customer’s goals or, according to the account’s needs, we lower or raise the goal.”

Agents, however, are measured against the current metric, and a good portion of their work take home pay will be dependent on achieving the currently established goals. As one manager stated:

In order to get it [the bonus pay], there is a set of filters. First agents must comply with several requirements in order to apply for the bonus. Then their performance is evaluated and confronted with the metric goals. The percentage [of additional pay they are able to get] depends much on the metric, on the employee performance against goals, and there are certain ranges to generate those incentives.

Salaries are increased also on the basis of achieving established goals and performance. As one manager indicated:

The salary only increases depending on regular salary increases made in the period that the agent has been in the company and if this agent is applying for performance incentives or not. Employees have an entry level salary that is usually similar to or less than the salary of people already in the company.

The data suggested that different workloads are given to agents as a function of their experience, or, in other words, goals adjust to a given level of experience. Similarly, as the information provided by managers seems to suggest, the work pay is not directly tied to intrinsic improvements in productivity, but rather to achieving certain goals. In order to obtain the bonus pay, the agent has to “apply” for it. The success of the application will depend on whether the agent has matched or surpassed certain goals. However, such goals can be “strategically” adjusted depending on company’s needs.

The information also suggests, particularly that of Table 2 and Figure 5, that as agents become more experienced they are able to release work pressure by shaving off minutes here and there without noticeably degrading their performance. However, managers, aware of this, by design increase the workload for seasoned workers. Hence, both performance goals and workload are floating variables. They are adjusted to match increases in productivity or contract pre-set requirements, such as resource utilization.

As time passes the probability of a worker remaining in the firm falls precipitously. Such propensity to leave the company is likely to be associated to several plausible reasons. Managers, however, seem to be fixated on the issue of resource utilization and burnout, which leads to attrition. Forced to compete on price, they seek to minimize cost per job by increasing the utilization of workers (by changing workloads) and minimizing salary expenses (by changing the goals that would permit an agent to attain bonuses and salary increases). However, the structure is far from linear. In the following section we describe a dynamic hypothesis.

DYNAMIC HYPOTHESIS

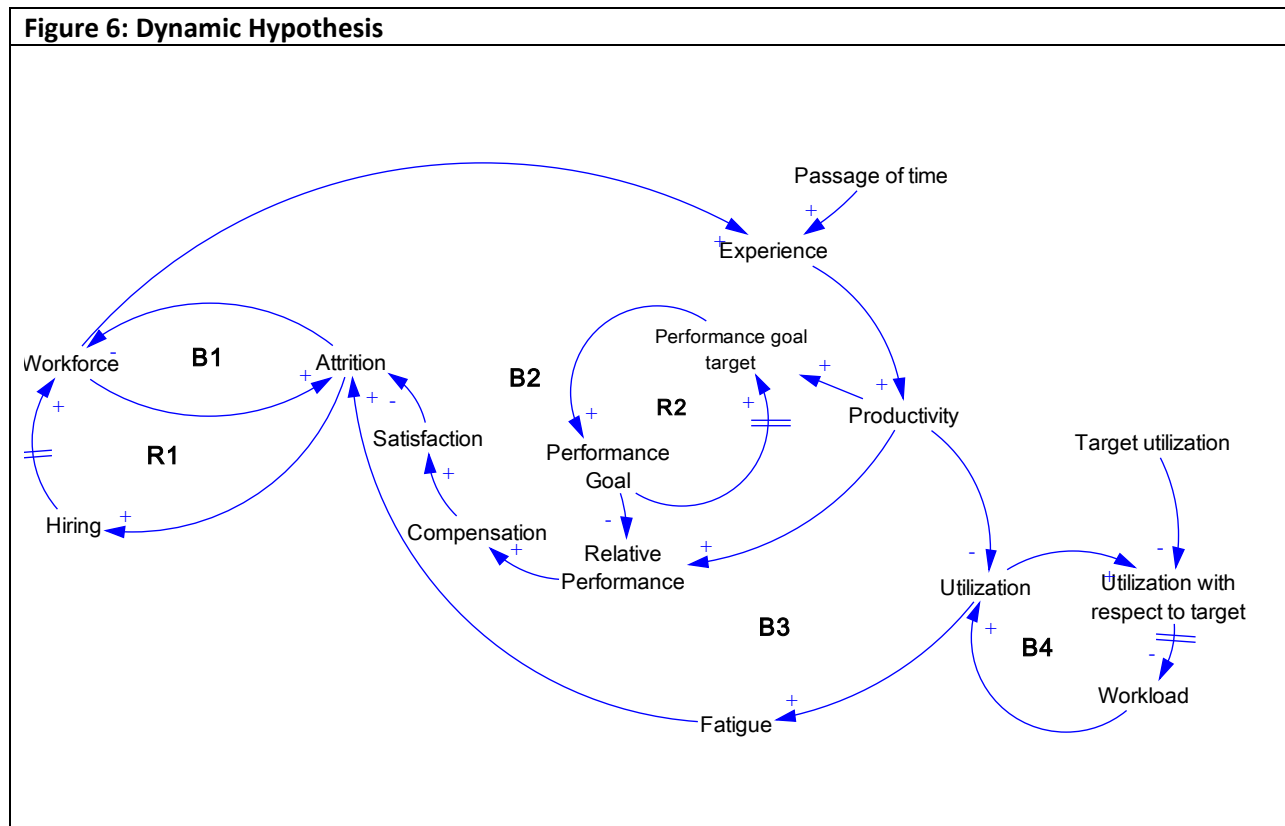
The data explored in the previous section, its exploratory nature notwithstanding, suggests a more intricate relationship between resource utilization and attrition than the relationship that managers of the firm under study suggested:

Utilization → Burnout → Attrition

There are several features in the previous data that point to the system adaptive characteristics and a structure full of feedback. These features are:

- 1- There is learning: as agents become more experienced they are able to handle the work in less time.
- 2- Targets are adjusted by management: managers have the ability to adjust occupancy either by readjusting the workload or by changing the number of people scheduled. For a given projected workload people are given a scheduled workload, but this workload seems to take into account worker experience.
- 3- As people become more experienced they tend to find ways to relieve the load while still accomplishing targets.
- 4- As individual agent load is reduced, workload is adjusted to achieve a desired occupancy.
- 5- The probability of an agent remaining in the firm goes down with the passage of time and the acquisition of experience.

These relationships are depicted in the dynamic hypothesis of Figure 6.



The dynamic hypothesis has one reinforcing loop and 4 negative feedback loops. These loops are described in turn:

R1: Hiring loop. As people leave there is more hiring. This, in turn, moves workforce in the same direction. With a larger workforce Attrition becomes larger than it otherwise would have been.

B1: Attrition loop. This loop balances the workforce through attrition.

B2: Productivity loop: This balancing loop relates to experience gains by workers who are, thus, able to achieve performance goals and be compensated accordingly. With compensation satisfaction is higher than it otherwise would have been and attrition goes down.

B3: Fatigue loop: As utilization increases so does fatigue. Fatigue brings attrition to a level that is higher than it otherwise would have been which reduces the workforce. Workforce and experience move in the same direction. As experience decreases overall, productivity decreases, thus increasing utilization.

B4: Floating workload: As occupancy decreases the workload is increased in order to achieve a given target utilization.

R2: Floating performance goal: As productivity increases the performance target increases.

There are other loops, such as experience gain through hiring, which are not shown for clarity.

SIMULATION MODEL

Appendix 1 shows the model for the causal loop diagram of Figure 6. The model uses an experience curve formulation as suggested in Sterman (2000).

The data showed that with the doubling of tenure the time required to handle a job becomes about 23% of what it initially was. Thus, a fractional change in productivity per doubling of experience was determined with the following (Sterman, 2000).

$$Productivity = Reference\ Productivity * \left(\frac{Average\ Experience}{Reference\ experience} \right)^c$$

Where $c = \log_2(1+f_p)$.

We introduced an effect of compensation on attrition by assuming that compensation moves with the performance relative to the goal, but a dissonance is created between relative performance (to the goal), and personal improvement (with respect to initial productivity). We assume that as the gap between the actual gains in productivity to the relative (to the floating goal) gains in productivity widens, and given that compensation is tied to the latter, satisfaction decreases and, therefore attrition increases. We assume that as resource utilization becomes closer and closer to 100% agent fatigue increases abruptly (because of work pressure). The increase in fatigue increases the percentage of people who leave the company.

Resource utilization is defined as a function of workload, workforce and productivity; as follows:

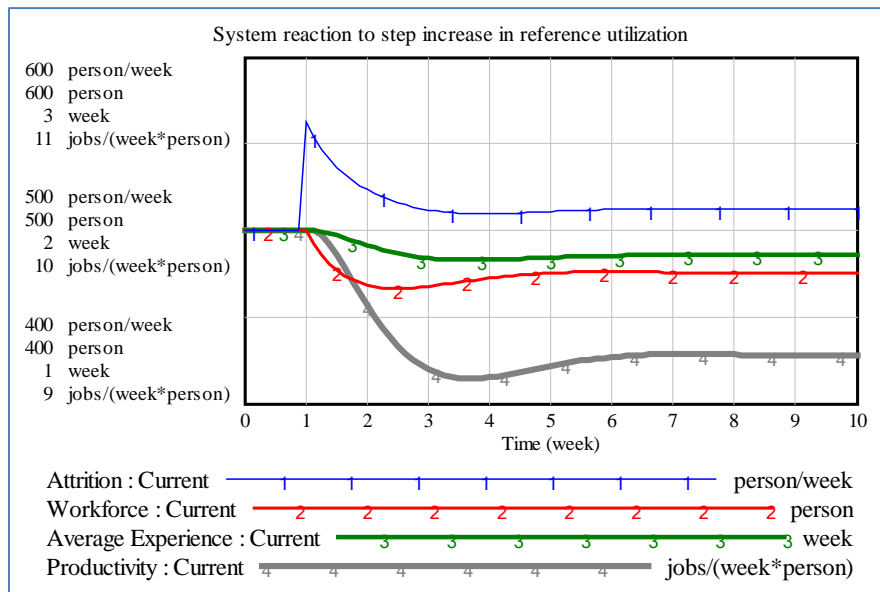
$$Resource\ utilization = (workload)/(workforce*productivity)$$

We assume workload and performance goal adjustments do not occur instantaneously. There is also a time for the new hires to go through training and become part of the workforce.

To test the response of the system to different disturbances we first started the model in equilibrium and then entered different step disturbances to examine the unfolding dynamics.

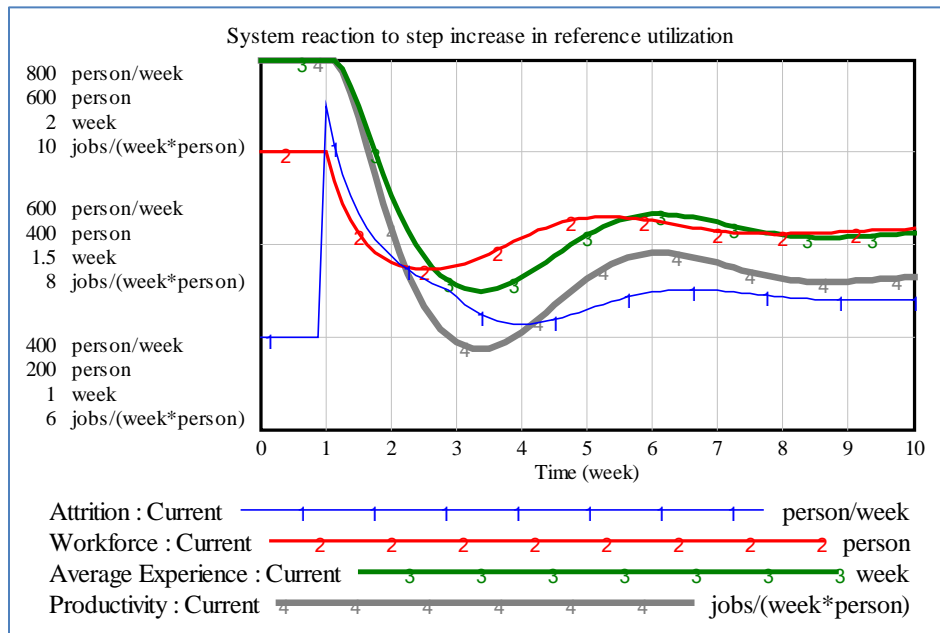
When utilization is increased abruptly, there is a sudden increase in attrition, due to fatigue. As a result, the workforce decreases and the average experience decreases also, with the system reaching, after some slight oscillation, a new equilibrium at a lower level of productivity and with a smaller workforce. This is observed in Figure 7.

Figure 7: System reaction to step increase in resource utilization.



A sudden increase in attrition has a more dramatic effect. Figure 8 shows the effects of an abrupt 50% increase in attrition. We can see that the Workforce goes immediately down, and average experience decreases with it. With the decrease in average experience, productivity goes down. At some point the people that were hired to replace those who left start entering the stock of workforce. With the increase in workforce both, average experience and productivity, increase. However, as productivity increases utilization decreases, goals are adjusted, and after some overshooting the system reaches a new equilibrium which is at a lower level than the equilibrium it started from. This comes as no surprise, as the model is just replacing attrition with equal hiring, but, given that there is a delay in between –due to training delays– the system settles down at a lower level of productivity.

Figure 8: Effect of abrupt increase in attrition

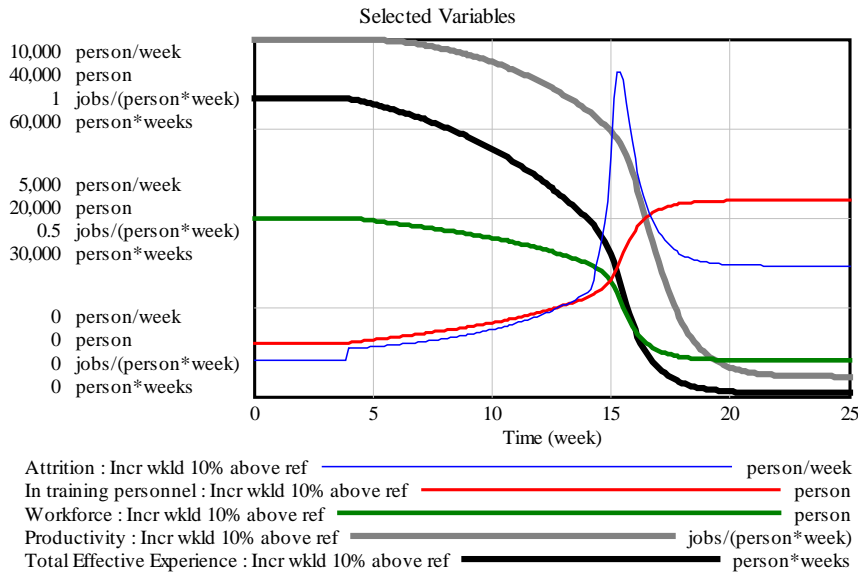


These experiments show that an increase in attrition, or an increase in utilization, will end up with a productivity equilibrium level that is smaller than the initial equilibrium. Attrition is pushed even further as both the fatigue and the dissonance loops kick in.

Thus, by increasing the workload to levels that induce fatigue, and hence attrition, it is possible that the system is gaining efficiency, because resources are more utilized, but at the expense of experience and productivity. Attrition, however, is not solely due to fatigue and burnout. As managers modulate target goals and workloads, agents are not compensated according to intrinsic productivity gains, but rather by productivity relative to a goal that is floating. As a result, agents feel shortchanged and decide to leave. The result is similar: we end up with a system less productive on the whole, with less average experience, and at a lower workforce level than it started from.

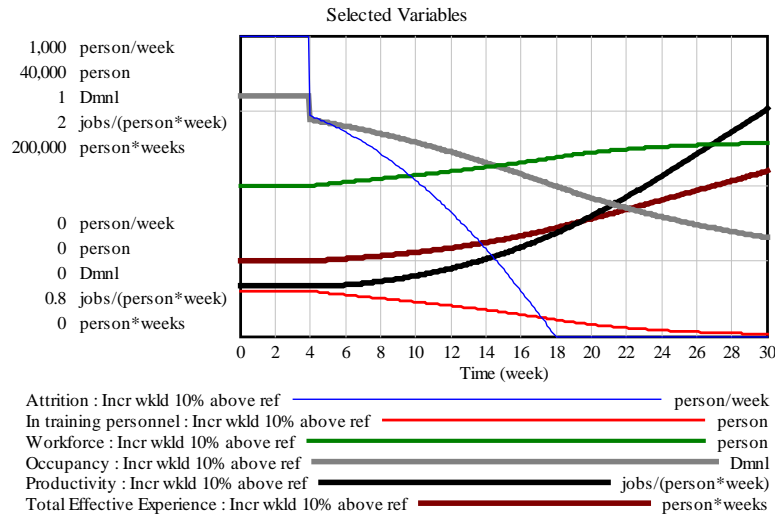
A possible policy change implies changing the structure of the system. If both floating goals are removed from the system, or, in other words, goals are set and fixed, the system exhibits much different behavior. An increase in utilization, by increasing the workload, still shows an undesirable outcome, because the system settles in a lower level of productivity, and a high level of attrition that needs to be compensated by maintaining a large number of people in training to resupply the system. Figure 9 shows such behavior.

Figure 9: Effect of increasing workload by 10%, no floating goals.



A *decrease* in workload (with no floating goals active) renders a somewhat unexpected behavior. The system initially decreases its productivity, because of the smaller resource utilization, but then, as the learning effect starts becoming important, the system actually improves its productivity overall, exhibiting a “worse-before-better” effect. Evidently the notion of actually decreasing the workload to improve system productivity is very alien to managers, and one difficult to implement. Figure 10 shows this effect.

Figure 10: Effect of decreasing workload by 10%, no floating goals.



CONCLUSIONS

In service operations, particularly in highly competitive environments, there is a tension between service level, resource utilization, efficiency and quality. As companies try to increase efficiency, attrition may increase. Many models indicate how the workforce should be managed in order to avoid such high levels of attrition while still achieving high levels of efficiency and resource utilization. One such common recommendation is to establish career paths. However, the need to compete on costs makes it difficult to do so. Although there are categories, the nature of work across categories remains essentially the same and practically no job enrichment is added as people gain experience. Some systems recommend, for instance, skill-based routing: giving the more complex jobs to the more experienced people. Such systems, however, in the end, just increase the implicit workload of agents without any compensatory measures. Such managerial approaches do not take into account that in managing the workforce we are dealing with an adaptive system.

The approach shown in this paper permits to look at attrition as a non-linear complex phenomenon. By understanding the causal structure that leads to attrition, managers can become more aware of the limitations and possible unintended consequences of their actions. Moreover, the paper shows that there is a very counterintuitive policy that could lead to the system's long run performance improvement, one that runs counter to managerially accepted practice.

This paper, as mentioned, is only preliminary. We hope to expand on the data we have collected and to advance this research by refining our simulation model and its assumptions.

REFERENCES

Armenia, S., Saullo, A. and Sedehi, H. (2006). Dynamic Skill based Routing: a System Dynamics approach to a Policy Definition in Call Center Management, Conference Proceedings. The 24th International Conference of the System Dynamics Society.

Bivona, E. (2007). Exploring Intellectual Capital Investment Policies in a Call Center through a "System Dynamics" Resource Based View. Conference Proceedings, The 25th International Conference of the System Dynamics Society.

Caramia, M., Armenia, S., Onori, R. and Giannunzio, V. (2003). Service Quality and Customer Abandonment: a System Dynamics approach to Call Center Management, Conference Proceedings, The 21st International Conference of the System Dynamics Society.

Maslach, C.; Schaufeli, W.B., and Leiter, M.P. (2001). Job Burnout. *Annual Review of Psychology*, 52, 397-422.

Oliva, R. (2001). Tradeoffs in responses to work pressure in the service industry. *California Management Review*, 43, 4, 26-43.

Oliva R. and Sterman, J. D. (2001). Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* 47, 7, 894-914.

Sterman J. D., (2000) *Business Dynamics: Systems Thinking and Modeling for a Complex World (Irwin Professional/McGraw-Hill, New York)*.

Zeynep, A., Armony, M. and Mehotra, V. (2007). The modern call center: a multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16, 6, 665-688.

APPENDIX 1: Stock and Flow structure of burning and floating goals model

