Dynamic Skill Based Routing: a System Dynamics approach to a Policy Definition in Call Center Management.

Stefano Armenia
PhD, MBA, Eng.
DISP - Dept. of Business Engineering - "Tor Vergata" University,
Rome (ITALY)
armenia@disp.uniroma2.it

Alessandro Pietro Saullo
B. Eng.
DISP - Dept. of Business Engineering - "Tor Vergata" University,
Rome (ITALY)
alesaul666@hotmail.com

Habib Sedehi
Communication Science Department - University of Rome "La Sapienza"
Rome (ITALY)
habib.sedehi@uniroma1.it

SYstem Dynamics Italian Chapter (SYDIC)
h.sedehi@systemdynamics.it
www.systemdynamics.it

Abstract

In modern society, Service Systems are considered one of the most profitable assets of both Business Companies and Public Organizations, since they can offer a highly added value for many different kind of products, both tangible and intangible. Among these Service Systems, Call Centers are considered to represent one of the most difficult cases to well understand and hence to manage. This difficulty is mainly due both to their intrinsic dynamics and to the unpredictable behaviour of input data. Furthermore, modern Call Centers tend to focus on the management of various and different types of calls, which most of the times require the government of different situations together with different agent (operator) skill management. So, in this type of scenarios, the main objective is focused on finding an efficient way to organize and manage a multi task Call Center.

In this paper we will try to show how these kind of Call Centers can most efficiently be managed through a Dynamic Skill Based Routing policy, instead of implementing a Static Skill Based Routing or even a call-processing policy which only uses a separate trunks/lines policy. We will try to state once more, how a simulation model can be useful in the management and "sizing" of a Call Center.

In order to achieve this objective , first we will describe both the structure and the typology of Call Centers, hence, by means of the System Dynamics approach ,we will define a possible model structure of a typical Call Center. The paper will end with a case study, where, through the simulation of different policies, we will show how the Dynamic Skill Based Routing is the most efficient way to manage multi-call-types Call Centers, among those presented.

Introduction

Service Systems, and inbound Call Centers in particular, have nowadays become very complex systems, gaining in importance as long as the significance of offering an added value to the customer is discovered. Due to the peculiar nature of their input and the tight connection between measurable (schedules, staffing, lines) and non-measurable components (customer satisfaction, staff burnout and other typical social systems' issues), they are very different from other typical industrial systems. Thus, some companies find themselves in need to have at hand proper tools to analyze the behaviours of their call centers, as well as to help the management both in sizing and in managing them.

Furthermore, since Call Centers usually have to deal today with several types of calls, it is often difficult, or unprofitable, to have agents capable of processing every type of call. This leads to specialization, with environments where agents are divided in groups based upon their skills, each group focusing their effort on one or more types of calls. Automatic Call Distributors (ACD) can assign calls to the agents with the proper set of skills, allowing this type of layout to be used. This approach is called Skill Based Routing (SBR). SBR is further divided into Static Skill Based Routing, where each type of call is routed to only one group of agents, and Dynamic SBR, where calls can be routed to different agents, depending on the Call Center situation; more specifically, each time a call needs to be processed, the ACD chooses, among all the free Agents, the one with the best skill-set related to the actual call.

Many studies have been carried out about different approaches to the problems of sizing and staffing, even in recent times. In this paper, we will approach the problem from a System Dynamics point of view, beginning from the work done in a previous work (Caramia, Armenia, Onori and Giannunzio [1]). In particular, as also anticipated in the cited paper, we will realize a new SD simulation model which is able to represent both Static and Dynamic SBR (with the proper input), with which to help the management in taking sizing and managing decisions. Using this model, we will prove how Dynamic SBR allows for more efficiency and elasticity.

Mathematical and analytical approaches to the problem of Call Center sizing and performance analysis, related to the Skill Based Routing, have recently been proposed by R.A Shumsky [4], that confronted the performance of specialized and flexible servers with a mathematical model, by R. Stolletz and S. Helber [5], that analyzed the performance sensibility of a SBR Call Center through numerical experiments, and by R. B. Wallace and W. Whitt [6], that through a stochastic model verified the effectiveness of even a limited level of resource pooling. Anyway, the strictly analytical models find it hard to take into proper account the “soft” components of a system, i.e. the human components and behaviours. M. Adria and S. D. Chowdhury [7] have recently proposed an analysis about centralization and specialization in Call Centers, analyzing the pros and cons of both of these latter issues.

Our point of view is that a Call Center can be considered a social system, since the human component has a great influence on the behavior of the system as a whole; as described in Caramia, Armenia, Onori and Giannunzio [1], a System Dynamics approach is thus well indicated to express such aspects, surely better than the analytical and mathematical approaches, since SD copes well with the lack of mathematical definitions that is a distinctive characteristic of the human behaviors. Furthermore, a dynamic model allows us to compare and verify the effectiveness of different policies over time. In particular, the use of a System Dynamics approach will allow us to analyze the different policies related to the Skill Based Routing; while verifying these, we must not forget that profitability plays an important role in the policy choice, and thus we will have to take it into account in the model structure.

In the paper, we will first pass through an overview of the structures and types of a Call Center, also explaining the main problems we have chosen to deal with. We will then deepen the motivations that lead us to the choice of the simulation and, specifically, delve into the use of the System Dynamics approach to analyze the system. Afterward we will describe the process we followed in building the model, detailing the main parts of the model itself. Finally, we will use the model to compare the different policies of SBR, and we will conclude by addressing further research and developments of our work.

Call Center Structure and Typologies

All Call Centers share some common traits: basically, they can be broken down into individual sub-systems consisting of entities queuing for some activity/resource, as explained in the queuing theory; more specifically, the main entities encountered during the Call Center (CC) analysis are constituted by calls and people, the latter meant both as service agents, but also as customers (Caramia, Armenia, Onori and Giannunzio [1]).

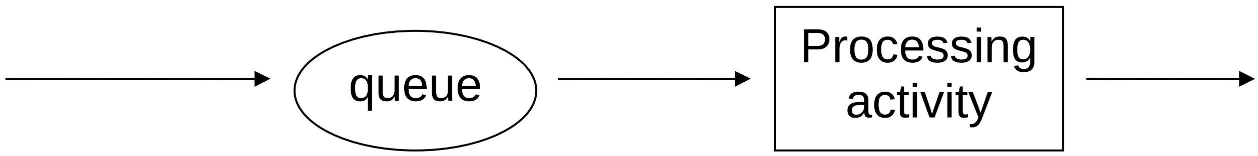

Figure 1 – typical queue

These "entities" can be interact differently, depending on the Call Center type, and on the number and type of the calls to be processed.

The Call Centers can be classified mainly into two categories: 1. Inbound and 2. Outbound.
In **Inbound Call Centers**, customers contact the Call Center asking for some kind of service, and wait until there is an free agent able to provide the requested service. In **Outbound Call Centers**, on the other hand, the agents begin the call, contacting the client offering some kind of service. This second type of Call Center is easier to be managed, because the call origination process is much more under the manager's control.

Inbound Call Centers must often process different call types on the same time; this can be achieved in different ways, with different levels of efficiency. The easier way is to split the Call Center in different and separate blocks depending on the call types, hence assigning a dedicated group of agents to each block. This organization is called *Static Skill Based Routing* (SBR). The Static SBR in fact behaves just like many small call centers, and has a very low flexibility.

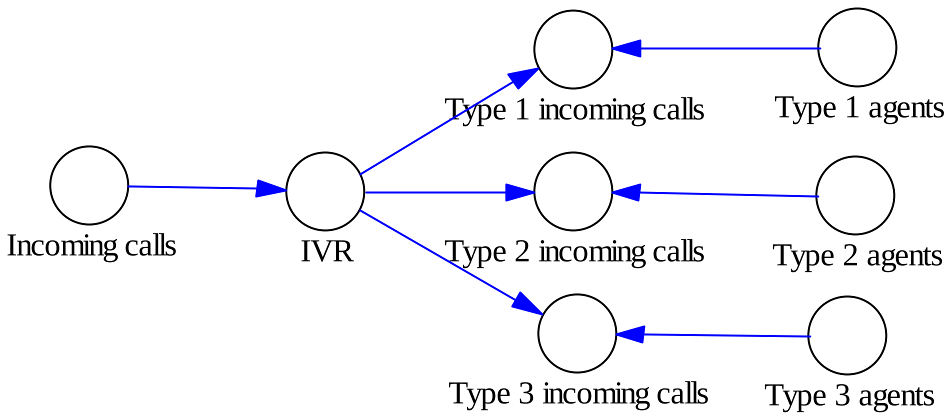

Figure 2 - Static SBR example

Another approach, surely more flexible, is the *Dynamic Skill Based Routing*. This consists of identifying the different calls as they arrive, and assigning them to the agent with the most appropriate set of skills, chosen from a common pool of free agents. This approach is much more efficient (from a Call Center point of view) than the first one, because it combines the quality of the service with the flexibility given by a single group of agents and by a communication infrastructure.

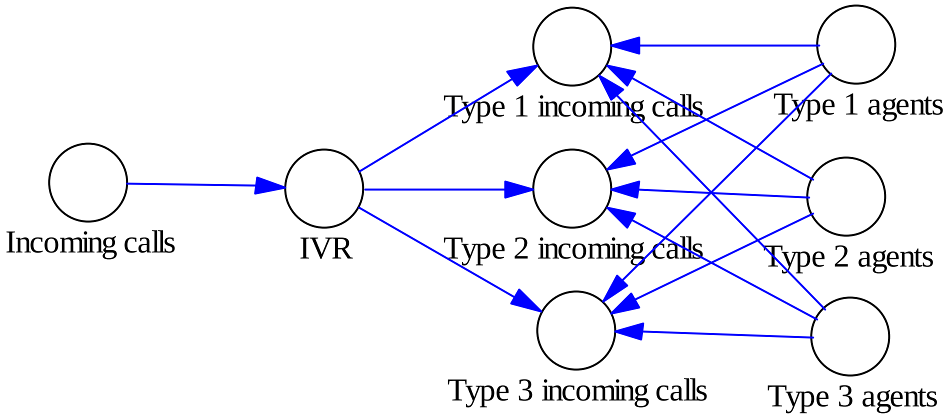

Figure 3 - Dynamic SBR example

Call centers need an automatic system to control the call flow: this is the Automatic Call Distribution system (ACD). In addition the system deals with the first contact with the customer, often offering information about the state of the queue and the estimated time to wait (or in the less advanced systems, offering some nice music). In those Call Centers which are supported by a computer system, the ACD is usually designed to manage the call by routing it through other different and specific communication tools. This is in fact usually achieved through an Interactive Voice Response (IVR) system, that uses a number of recorded messages to identify the incoming requirement service. The ACD then, simply has to direct the call to one of different queues, depending on the type of requirement and eventually on the system status. Each queue consists of a group of lines, called “Trunks”; in the Dynamic SBR, the ACD also chooses the best agent available to process each call, based on parameters set by the management (the most common is the specific preparation of each agent relating to the category of the call requests).

Once the call is processed and the customer leaves the system, the busy line returns immediately free and the agent involved completes the necessary, and so called, After Call Work (ACW), before successively becoming available again.

The Problem

As we already said, in this paper we will propose a model in order to help managers of inbound call centers, those ones that see more than one type of incoming calls, to confront alternative managing options and thus forecast the outcome of a certain change in the call center structure. To be coherent with past works, and in order to confront results with previous model efficiency and performances, we have built a model that works on a relatively short time span (30 minutes) thus mainly focusing on tactical aspects of a Call Center management. In future research, we are planning to expand the model from a tactical level to a strategic level, which means expanding the simulation time window at first realizing a one day simulation (thus better evidencing both incoming calls fluctuations over a 24-hours period as well as the dynamics of agents shifts and schedule, burnout, etc…) and then moving to a next stage of one week/month simulations, thus allowing us to take into account also aspects concerning customer relationship and quality of service effects, and much more.

Anyway in call center management, as it happens in most physical systems, a general rule is that one of the most significant problems consists in the difficulty to find the right sizing, and the right solution to the inevitable problems, without previously experimenting the possible solutions. The reason for this is that since the focus of a Call Center's performance is towards customers and towards a correct management of the relationship with them, it is easy to understand how in this field there is little to no space for mistakes: once the customer is unsatisfied, and the service gets reknownably poor, it is really hard to gain back customers' trust. Yet, since the Call Center system presents so many uncontrollable aspects (first of all, the fact that it has to operate under uncertain demand, without any forewarning; but also the fact that the quality of the Agents' work is less controllable, at least directly, than in other productive systems), it is very difficult to formulate "operation standards" or produce "best practices" for those managers who would need to have them at hand when needed. In truth, even problems very different from each other, may lead to very similar results, but on the other hand it's also true that what worked once may not be as effective again. Thus, a model that allows to simulate a Call Center, making it possible to conduct also unusual analysis and experimentation, even those that would not normally (say, mathematically) be easily feasible or sometimes even ethical in the real system (like, for examples, studies taking into account the effects of excessive burnout on staff performance), can be a very useful tool in the hands of the management.

Why use simulation

Senior managers, as well as analysts and advisors, aiming at optimizing the decisions concerning the management of the system they interact with, may find a suitable support by simulation methods instead of using those approaches eventually derived from traditional and analytical methods, and in particular in service systems where the environment is extremely dynamic and volatile. Moreover, business strategy would take advantage of such an approach which may virtually avoid many of the risks associated with mistakes in the assessment of the alternatives. To achieve good results, the manager, as it is the case in many similar problems, has to seek the proper trade off between service capability, a high value of which usually means also high investments, and a better service to customers. A challenging aspect of such an exercise is that the results may be highly sensitive to parameter variability (Jon, Vivek, Bill [3]).

For example, a small adjustment in call routing may have a significant debilitating change on customer satisfaction, or a minor reduction in trunk-line capacity may cause too many “busies” and raise the potential loss of customers (which, according to the well known Customer Based View, may cause on a longer term a negative impact and hence reduce the overall revenues and profits of the company). These relationships follow a circular pattern that must be carefully defined and analyzed if a high quality service is among those main objectives of the management actions.

Differently from what happens with the use of Operations Research tools like, for example, linear programming where finding an optimal solution is a straightforward process, “simulating” a Service System can present some disadvantages and difficulties if one seeks to find an optimal solution. One way to attempt to achieve similar results using simulation is to make changes to the model by setting and tuning appropriately its parameters, run the simulations and see if an improvement has been achieved or not, and repeat. Although the process may consume large amounts of computer time, nowadays there are some simulation packages which also support some optimization tools.

A further advantage of using simulation, as opposed to analytical methods, is due to the fact that it’s possible to deal much more easily with time-dependent behavior; where the mathematics of queuing theory is hard and only valid for certain statistical distributions, the mathematics used in computer simulation need not to be complicate and can cope with any statistical distribution. In some situations where we may want to model some particular features of the system, it is virtually impossible to build the equations that the queuing theory would demand for (e.g. for features like queue switching, queue dependent work rates) and, finally, the simulation approach is usually much easier to be quickly mastered by managers than queuing theory or difficult analytical methods. (Caramia, Armenia, Onori and Giannunzio [1])

The System Dynamics approach

As we already said, the core of the System Dynamics approach is building a computer model which represents the analyzed system.

This model has to be able to deal with the most difficult aspects of a Call Center, that are:

- high dynamism;
- non-linear relationships;
- different types of data processed (both hard and soft);
- multi feedback processes;
- many interdependent components.

With this issues in mind, we have to build a model that is both complex enough to take into account all these aspects, and comprehensible enough to be understood easily by the future users.

To build the model following a System Dynamics approach, we have to consider a system not only as a conglomerate of separate components, but as a whole, focusing on the relations between the components more than on the components themselves, thus being able to take into account all the cause-effect relations that define the real system. In this analysis, we will have to consider how all these relations contribute to generate causal loops, that is those chains of relations where a variable, while changing, gets to influence its own behaviour over time. Loops can be both Positive or Negative. Positive Feedback Loops exist when a certain increment in a variable leads in time to a further increase of the variable itself. These loops are typical of the so called amplifying systems, in which some parameters, unless kept under control, can determine exponential evolutions, preventing the achievement of a stable equilibrium. In Negative Feedback Loops, on the other hand, the increment of a variable generates a decrease of that variable over time. These loops are typical of stable or oscillatory behaviour systems.

Since in simulation we have the support of a computer to perform calculations, we can express the non linear relations between the components without worrying about calculus complexity. Anyway, we must someway pay a price in this since it’s often necessary to pay some more attention to the so called soft data. Considering various sources from which to take the information, since we need to express soft relationships as closely as we can to reality, we need to base our inferences and theories on as much of this sources as we can, like interviews, numerical data, direct observation, or else; even if it is very difficult to consider all the available information, this is, however, a great modelling process improvement if confronted with the Discrete Simulation approach, that cannot consider this kind of data at all.

The multiple feedback loops present in the process, together with a number of qualitative variables (soft data) that influence the whole system, create an optimal environment which thus suggests to tackle the problem of the policy definition through S.D. approach.

The modelling process

Analyze the system morphology

The first step towards the building of our model is the analysis of the main parts of the relative real system. To do such analysis, block diagrams have been designed, representing the flows that take place in the Call Center. These diagrams represent the frame of the process to be analyzed, and strictly correspond to what happens in the system and to the parts we intend to include in the model.

In the examined case, we start representing the incoming calls flow. The incoming calls must obligatorily pass through a common route, since they haven't been identified yet. They enter the Interactive Voice Response System, that proceeds to identify (for example) the two call types (**Figure 4 – second column**). In the case in which the IVR lines are under-dimensioned, some calls receive a busy signal and are blocked. The IVR then splits the calls among (for example) three separate queues, two of which are dedicated only to one type of call (queue 1 to calls marked as dedicated to A-type calls and queue 2 marked as dedicated to B-type calls), while the third can receive both call types (A or B) indiscriminately (**Figure 4 – third and fourth columns**). Priority is given to the dedicated queues, since the shared one has been introduced as a backup in order to be able to deal with possible unexpected call peaks, both A and B. Should this queue be under-dimensioned, some calls would be blocked after the IVR.
After that, as the customers enter the queue, they are notified about the average waiting time before reaching an operator; there is thus a chance that some of the customers may find that the queue is too long, thus abandoning the call and hanging up upon entering the queue (Balking effect). Also, if the waiting time is too long, some of the customers in queue may consider that they have waited too long and thus decide to abandon the call as well (Reneging effect).

When there are free agents (either specialists or generalists) the first customer in the queue is put in contact with the first available agent, and the processing of the call begins. **The priority is given to specialists, so as to keep the generalists ready to process unexpected peaks from either type of call, also because the specialists are more prepared than the generalists**.
After the talk is over, the customer abandons the line, the agent goes on with some after call work and the call enters in the state "handled", becoming either satisfied or unsatisfied, depending on the quality of the service (**Figure 4 – last two columns**).

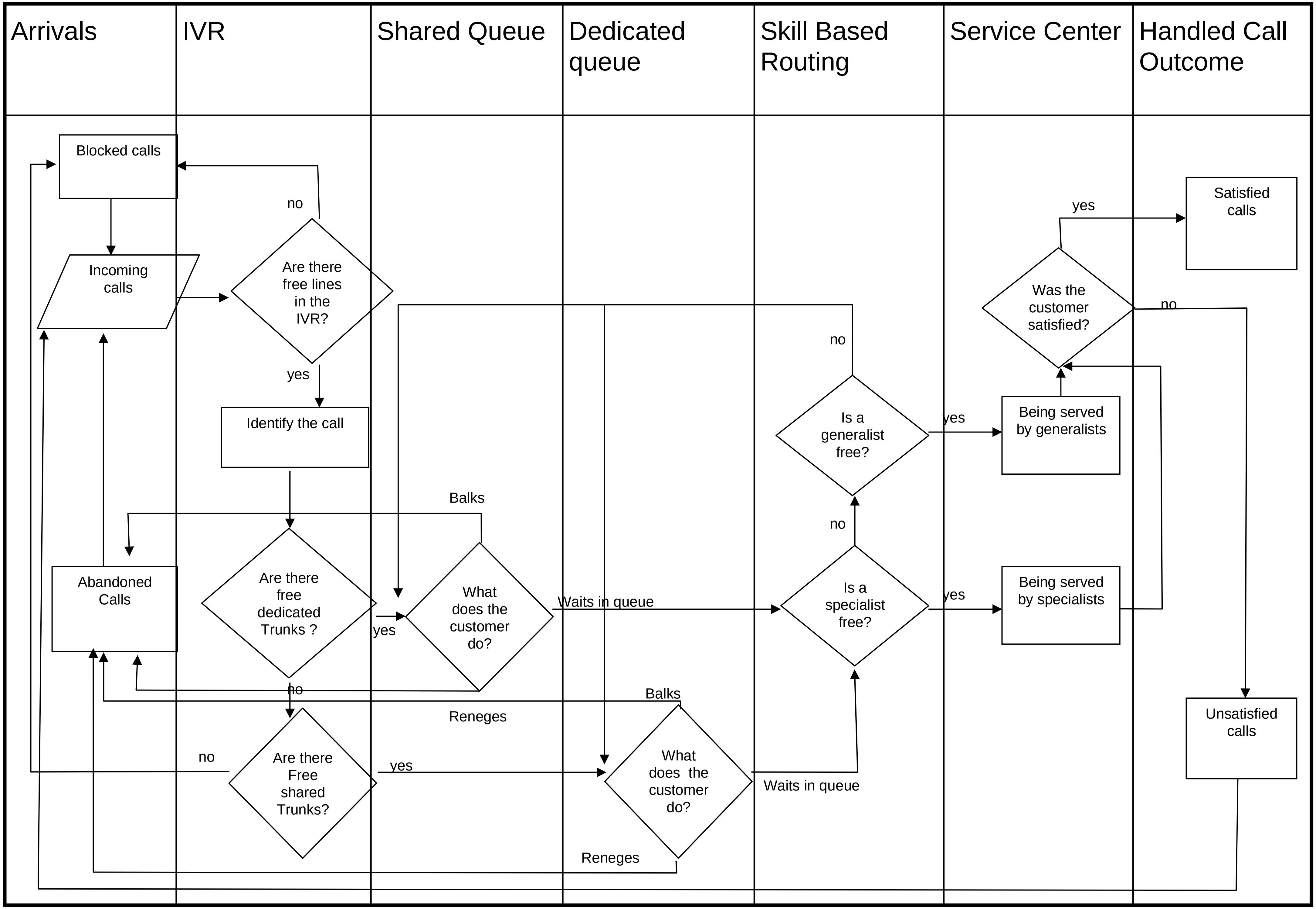

Figure 4 - Call flow diagram

We now need to examine the flows related to agents' behaviour, which are similar for the three types of agents.

The agents begin the simulation in the "free agents" state, i.e. agents available to serve. Free agents are subject to three different flows (**Figure 5**):

1. the first one represents the answering process; when there are free agents and there are calls in queue, the first customer in queue is put in contact with a free agent, and the processing begins; when the call is over, the agent has to complete some After Call Work, and then returns to the state of free agent.
2. the second flow represents the agents getting on pause; this flow is ruled by factors that on the whole are not under the manager direct control.
3. the third flow represents the agents that, when the service request is low, are assigned to other tasks in the call center, to be called back when the request grows.

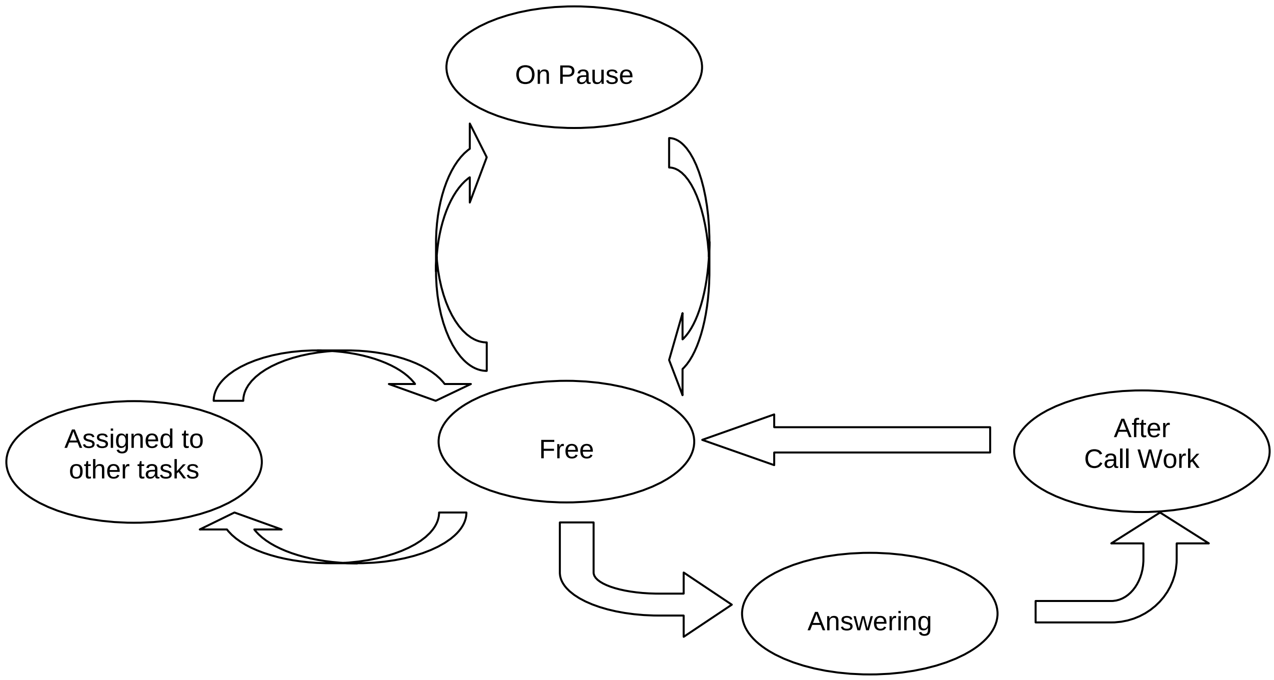

Figure 5 - Agents' behavior

Identify Cause-Effect relationships

Once we identify the components of the main flows of a Calls Center, as the one we are studying, we have to analyze the cause-effect relationships defined by the interactions among these parts. This is done by means of causal loop diagrams, that, qualitatively depict the relations between the variables and components of the system, thus allowing us to identify the positive and negative loops. To find out which kind of relationships we can expect to observe, we can rely on the classical theory and on some past works (Caramia, Armenia, Onori and Giannunzio [1]).

Let us start from the so called "five fundamental principles that govern a call center" and characterize its dynamics (see Cleveland and Mayben [2]):

1. For a given call load, while service level increases, agent occupancy decreases (mostly because the call load is evenly distributed among them)
2. By keeping improving the Service Level, a point of diminishing returns will be reached (limit to the growth)
3. Given a Service Level, larger agent groups are more efficient than smaller ones (agents show a higher occupancy percentage)
4. All other things being equal, pooled groups are more efficient than specialized ones (handle more calls with same number of agents, same call load with fewer agents, same call load with same number of agents at a higher Service Level)
5. Given a call load, if staff is increased, then the Average Speed of Answer will decrease and the trunk load will go down

Moreover, we have found evidence, also in the literature (Busacca-Valdani [8], Sterman et al. [9]), that the quality of the service offered depends both on factors connected with the effort and commitment of the management towards a qualitative service, as well as also with customer expectations and satisfaction (e.g. calls handled qualitatively by agents), and on variables which are typical of a call center structure, i.e. Service Level (SL), Average Handling Time (AHT), Average Speed of Answer (ASA), free communication line, and so on.

In particular, we can say that an effort of the management towards making agents handle calls in a qualitative way may be viewed as a tendency to skill agents so that they can meet customer expectations, thus improving the customer satisfaction index. In fact, reducing errors and rework has been seen to have a positive impact on service level, morale, customer satisfaction and, last but not least, costs. Typically, customers expect a CC to be accessible, to be treated courteously, to promptly do what they ask, not to deal with poorly trained agents, to be responsive to what they need and want, to do it right the first time, to be socially responsive and ethical, and so on.

Under this premises, a call is thus defined as “qualitative” when:

1. the customer is satisfied with the received service and of his/her experience in the CC,
2. the Agent captures all needed/useful information,
3. all data entry in the After Call Work phase is correctly done,
4. the Agent provided correct response and the caller received clear and correct information in a time not perceived as too long.

Other important features which set the quality of the answer are that the caller does not get transferred around the CC or does not get rushed, the customer has confidence that his call was effective, the caller does not feel necessary to verify, check-up, repeat or even call back. Even the Agent must be satisfied of how he just handled the call, the caller must not get any busies or be placed on hold for too long; by this way, in the end, the CC mission is accomplished. The equation that puts in strict relationship quality effort and highly skilled or experienced agents seems not to be too far from reality [1].

Considering this conclusions, we can then draw some causal loop diagrams regarding the various aspects of the call center. In the following figures each component will follow this legend:

[<*name of the area of the component*>**]** <*name of the component*>

and parts that are inherently destined to be represented as levels in our model are marked with a surrounding square.

The diagram in **Figure 6** concerns the Interactive Voice Response System; we notice how an increment in the “A-type” Incoming Calls (A-calls) determines an increment of the A-calls blocked, since some of the blocked calls are subject to retrials, in the IVR. Since the IVR lines are shared among the calls, an increment of A-calls in the IVR determines on average a decrement of B-calls in the IVR; thus, the more B-calls will be blocked and the more they will increase through the retrial process dynamics. This leads as said to an increment of B-calls in the IVR, creating a negative loop.

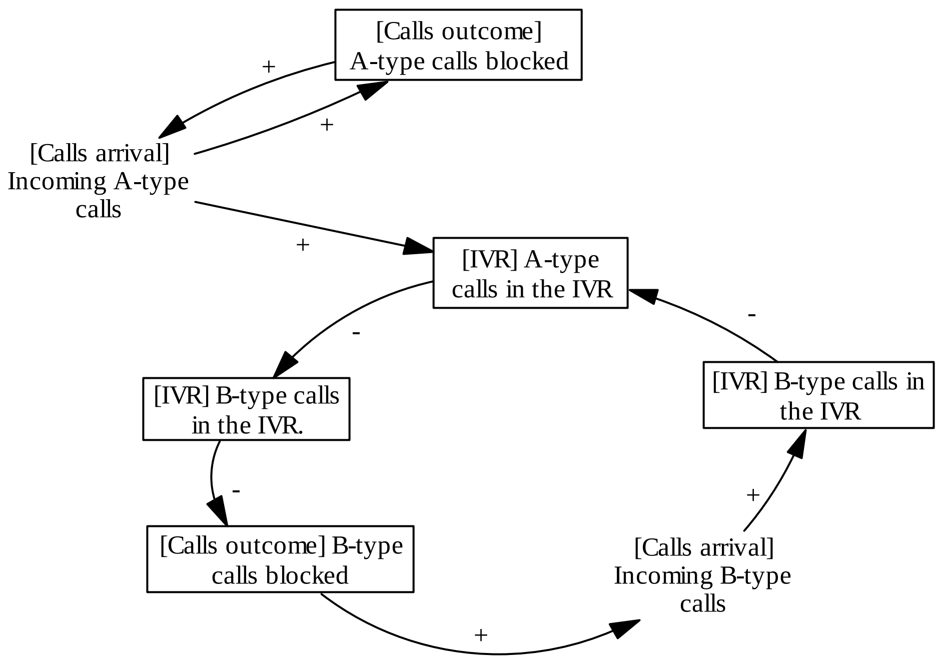

Figure 6 - IVR causal relations

Figure 7 represents the loops concerning the main service center; as the calls in the IVR increase, we have an increment in the calls awaiting to be sent to the queue. Depending on the status of the queues, this increment can lead to an increment in the dedicated queue, in the shared queue or in the blocked calls. Should the blocked calls increase, the incoming calls would increase, too, thus creating a positive feedback loop. If instead we have an increase in the calls in the shared queue, since the lines are shared we would have on average a decrease of the other type of calls in the shared queue. This would lead to an increment of the other calls blocked, in turn increasing, through the retrials, the other calls in the shared queue, thus completing a negative feedback loop. On the other hand, if we have an increase in the dedicated lines, there will be no consequences on the processing of calls of the other type.

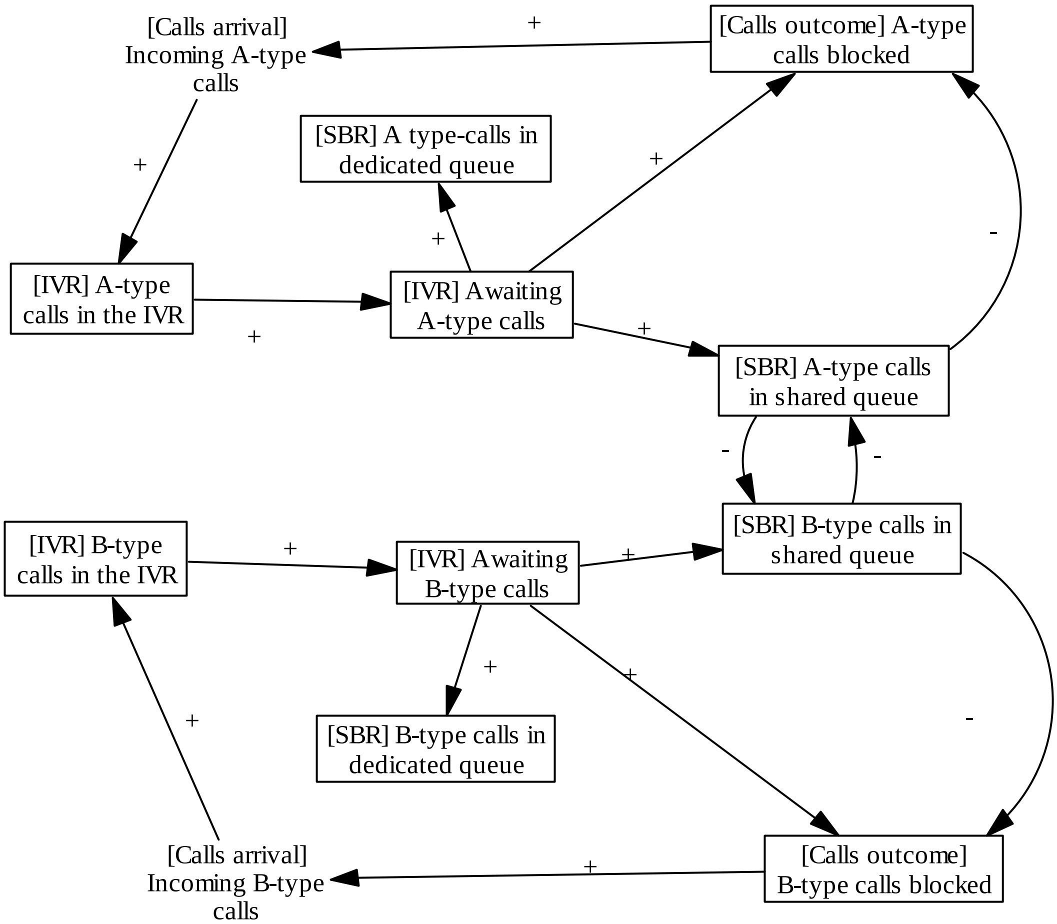

Figure 7 - Queue insertion causal relations

The diagram in **Figure 8** represents the consequences that an increment of the calls in the queues has on the staff behavior. In that diagram, we considered an increment in the A-calls dedicated queue, but similar diagrams are applicable to the other queues as well. When the calls in queue increase, we have, at least at first, an increase in the served calls, thus incrementing the profit and then the Quality factor (an indicator that considers both speed and quality of service, and the quality of the system on economical terms). As we see from the block diagrams, though, an increment of the calls in the queue also means, on average, an increment of the abandoned calls, thus increasing the incoming calls and creating a positive loop. It also means an increase in the waiting time for the customers to be answered, thus causing even more abandoning and creating another positive loop.

An increment of the calls in queue also means an increment of the staff occupancy, thus shortening the agent's free time and increasing the agents' stress; this, in turn, also causes an increment both in the waiting time and in the possible mistakes. Increasing mistakes means increasing unsatisfied calls, thus reducing the quality factor and increasing the incoming calls (creating another positive loop).

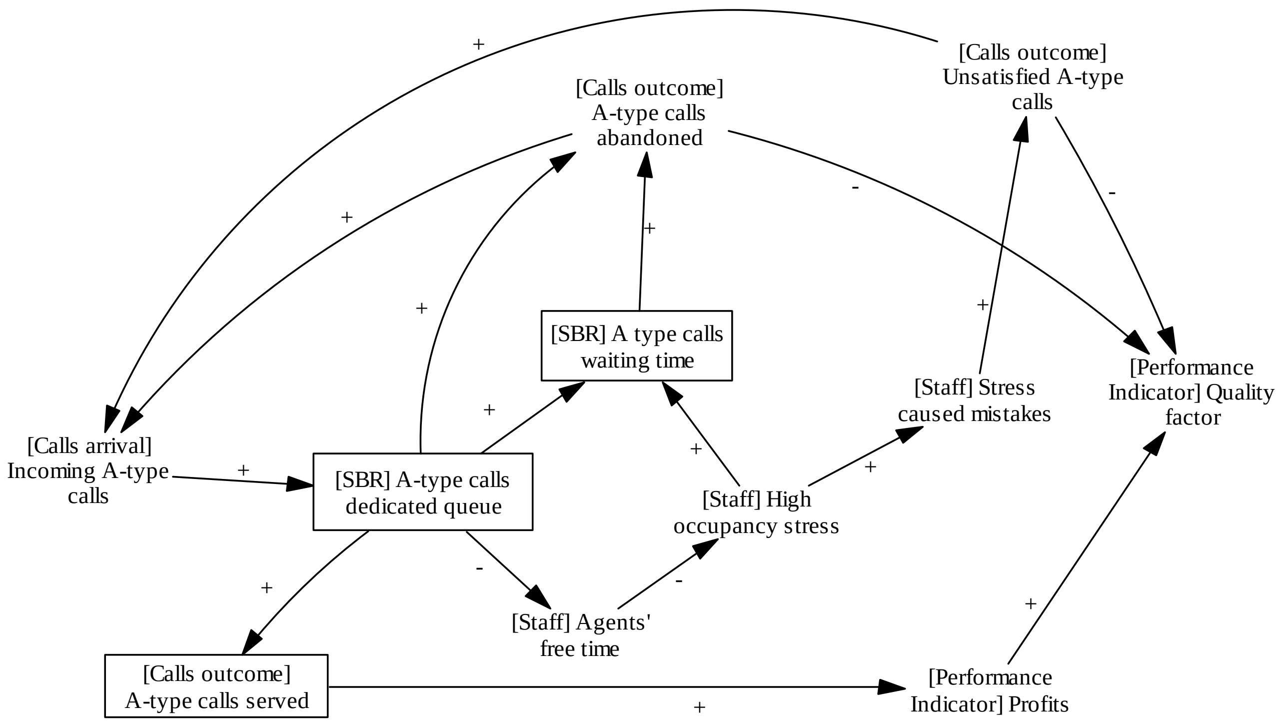

Figure 8 - Service Center causal loop relationships

Figures 9 and 10 represent the feedback loops related to an increment in the level of a certain agents' type (in this case, A-calls specialists). Increasing the staff, leads at first (as we defined the process of agents flow in the system, see figure 5) to an increase in the number of free agents, thus in turn determining a reduction of the queue length and consequently the blocked and abandoned calls. This brings to an increase in the overall call-handling process quality, and causes a reduction of the incoming calls, since less unsatisfied calls means less retrials, thus further shortening the queues and creating positive feedback loops.

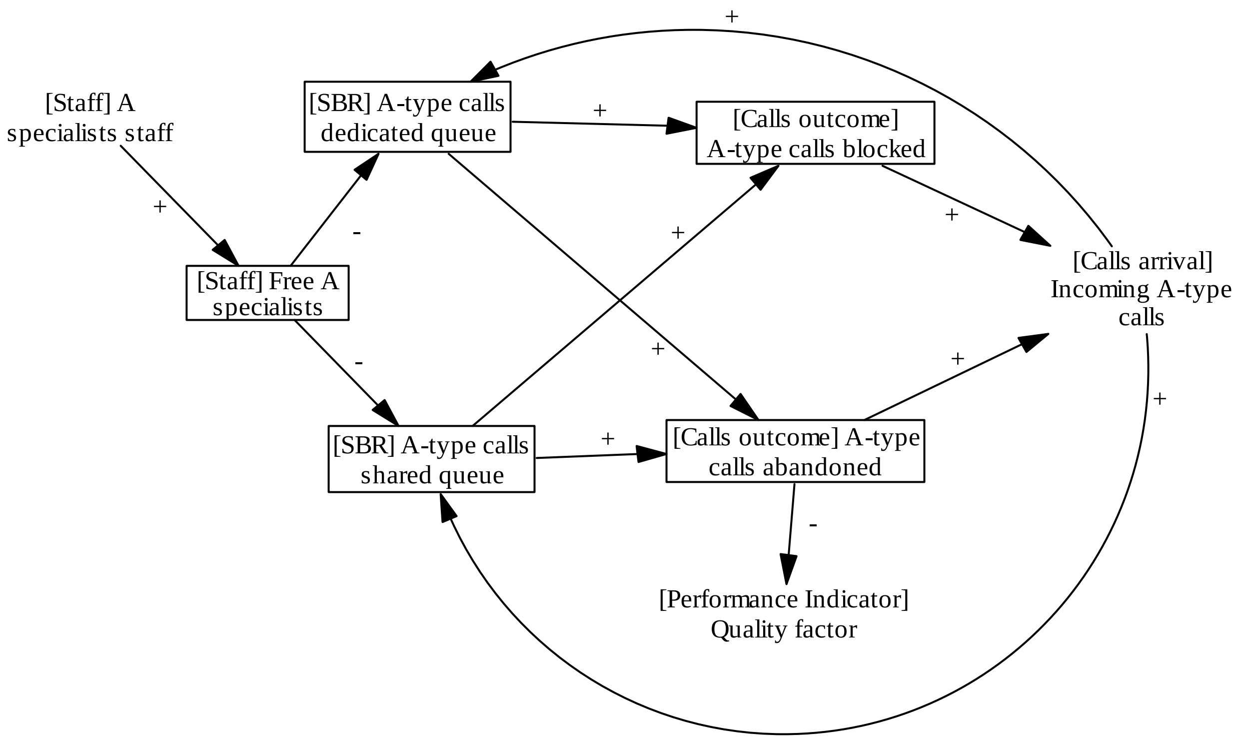

Figure 9 - Staff causal loop relationships – 1

In the second diagram (**Figure 10**), we see how an increase in the staff generates an increase in the staff free-time, thus reducing stress and errors and reducing, over time, the number of unsatisfied calls. This also helps increase the Quality factor, reduces the incoming calls due to unsatisfied retrilas, shortens the queues and eventually increases the agents free time even further, thus creating two parallel positive loops, as it is evidenced in the following figure.

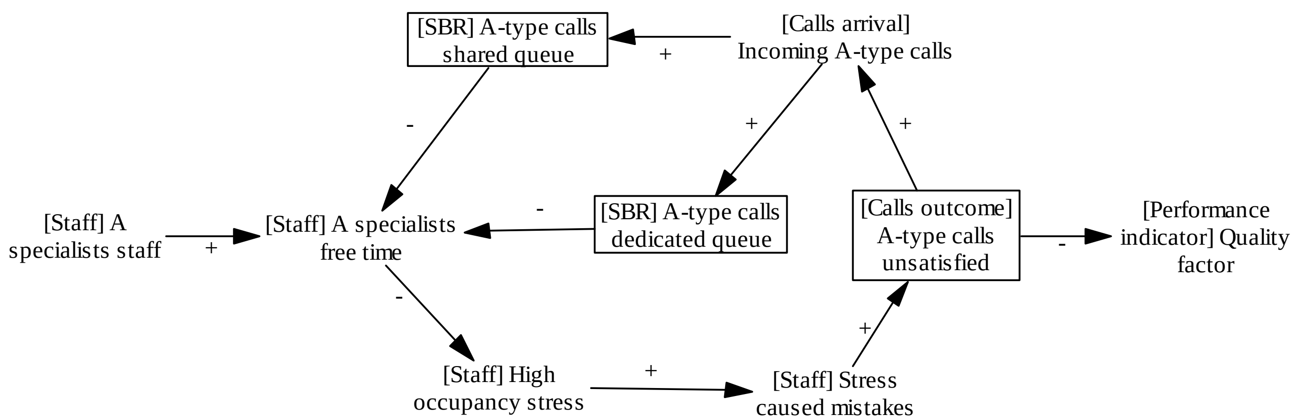

Figure 10 - Staff causal loop relationships – 2

Another meaningful diagram concerns the consequences of the increase in the Agents' training and thus in the level of the average experience (**Figure 11**). Such an increase reduces the incompetence mistakes, thus reducing the unsatisfied calls. As we have already seen, this increases the quality factor and reduces the incoming calls, creating positive loops similar to those in **Figure 8**.

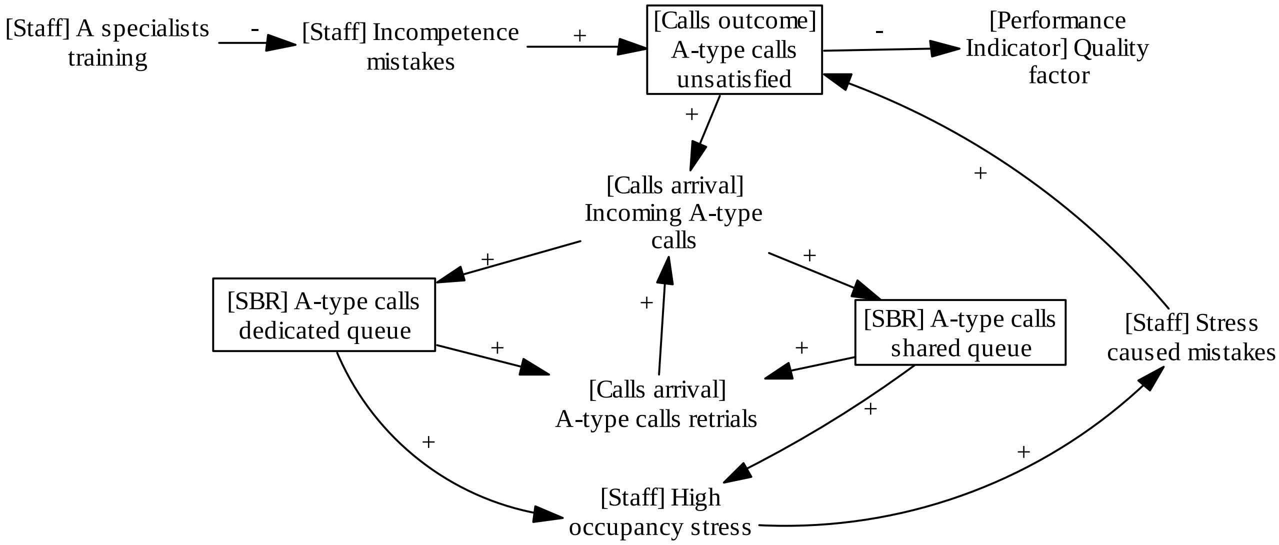

Figure 11 - Training causal loop relationships – 1

The last diagram (**Figure 12**) represents another effect produced by the agents' training, that is on profits. On one hand the increased training increases the agents' cost, thus reducing the profits; on the other hand, it reduces the unsatisfied and abandoned calls, thus increasing the long-term profits. The net result of this two components, determines if the training increase is profitable or less.

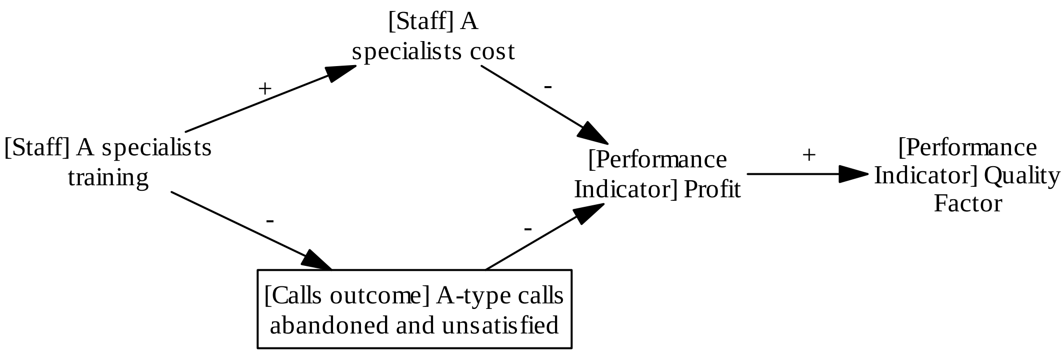

Figure 12 - Training causal loop relationships - 2

The Model

Following the directions given by the causal diagrams, a model was realized (using Powersim 2005 professional, demo version). In the model the flows represented in the diagrams are controlled by logical functions, based on the queue theory, on the cause-effect relations of the causal diagrams and on some real historical data taken from various call center (especially for the Balking and Reneging behaviour, since this parts are out of the management direct control).

Following are some simplified versions of the main parts of the model, to emphasize the connection between the model and the diagrams and causal maps.

The first part represent the arrival process, and the call type selection through the IVR. The incoming calls enter the IVR, where the customers identified themselves and are split in group A and B (**Figure 13**). Note that this, and the following parts represent what happens to the A type calls. A similar diagram is applicable to the B type calls.

In this part, we can note how the blocked, abandoned and unsatisfied calls influence the incoming calls.

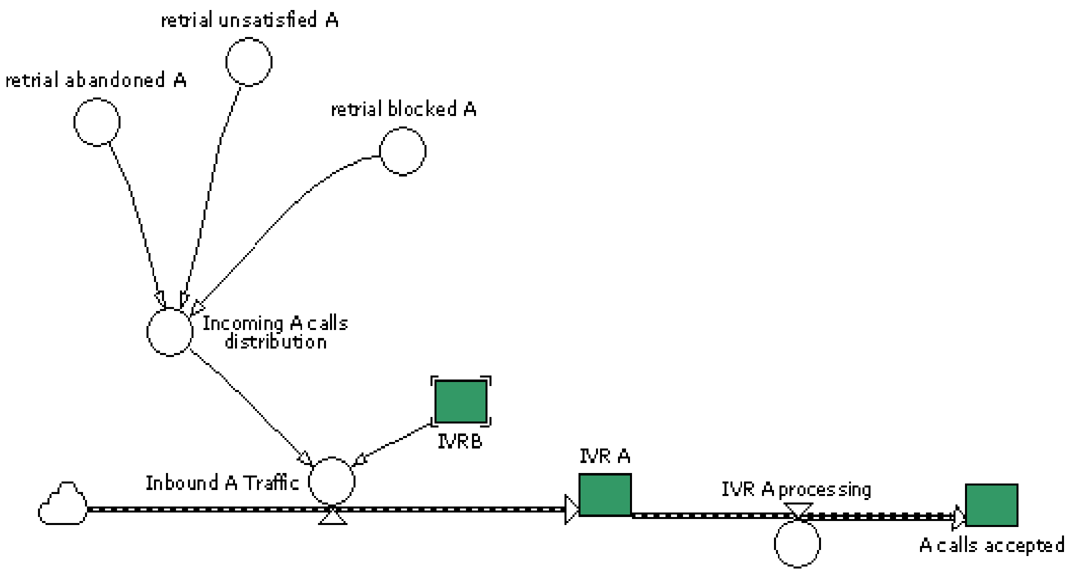

Figure 13 - A type calls arrivals

The second part represents the insertion in queue, and the beginning of the answering process. The calls, from the state of accepted pass to the dedicated queue, shared queue or blocked group, depending on the state of the queues. The calls in queue enter the service center as long as there is an agent available to process them (**Figure 14**). Note that the variables controlling the flows are connected to the variables of the B calls, but this link has been understood for readability sake.

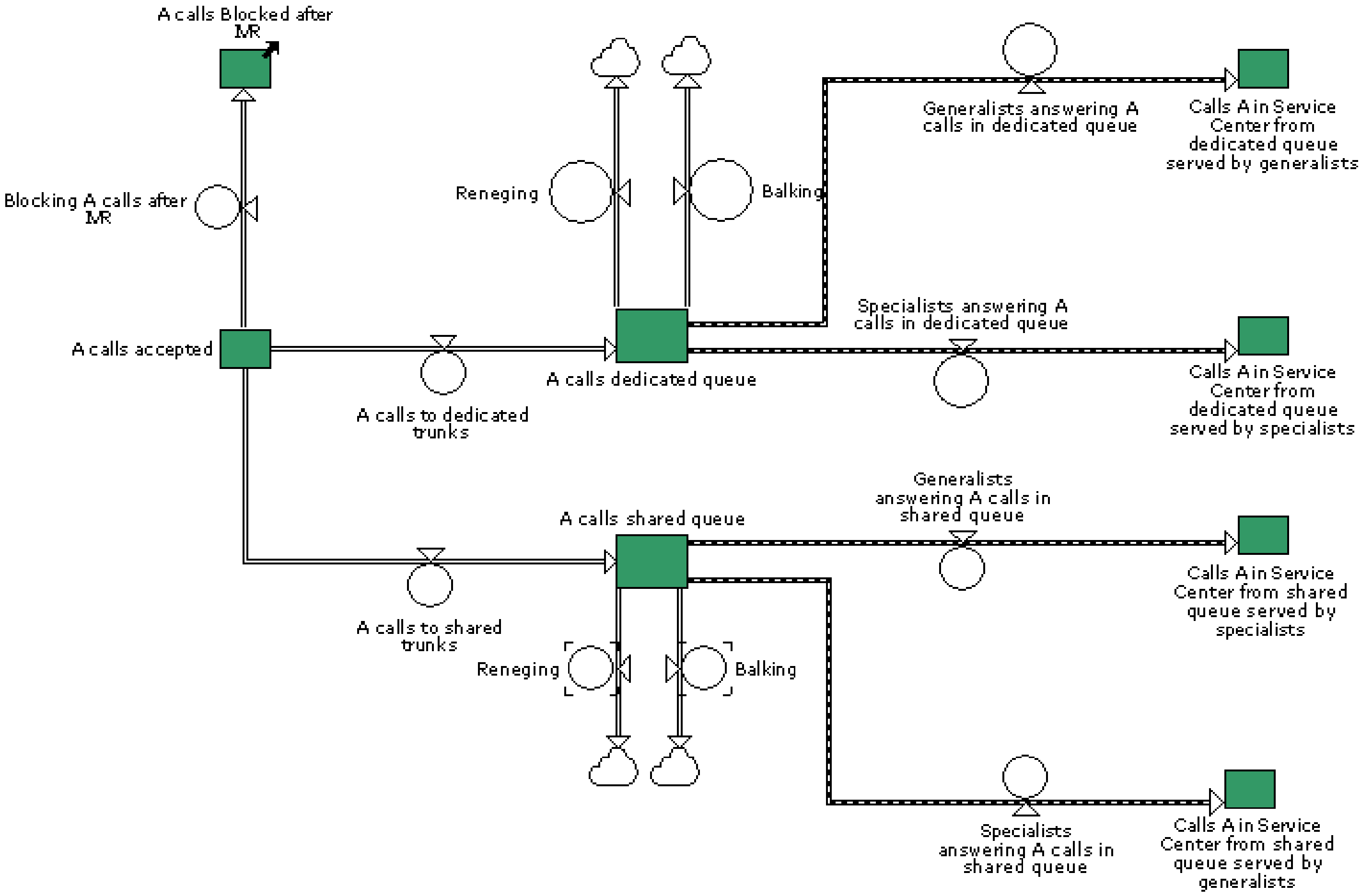

Figure 14 - A type calls Skill Based Routing

The third part flows (**Figure 15**) represent the end of the service, redirecting the calls, as long as they are processed, from the 4 blocks into 2 groups, those handled by specialists and those handled by generalists.

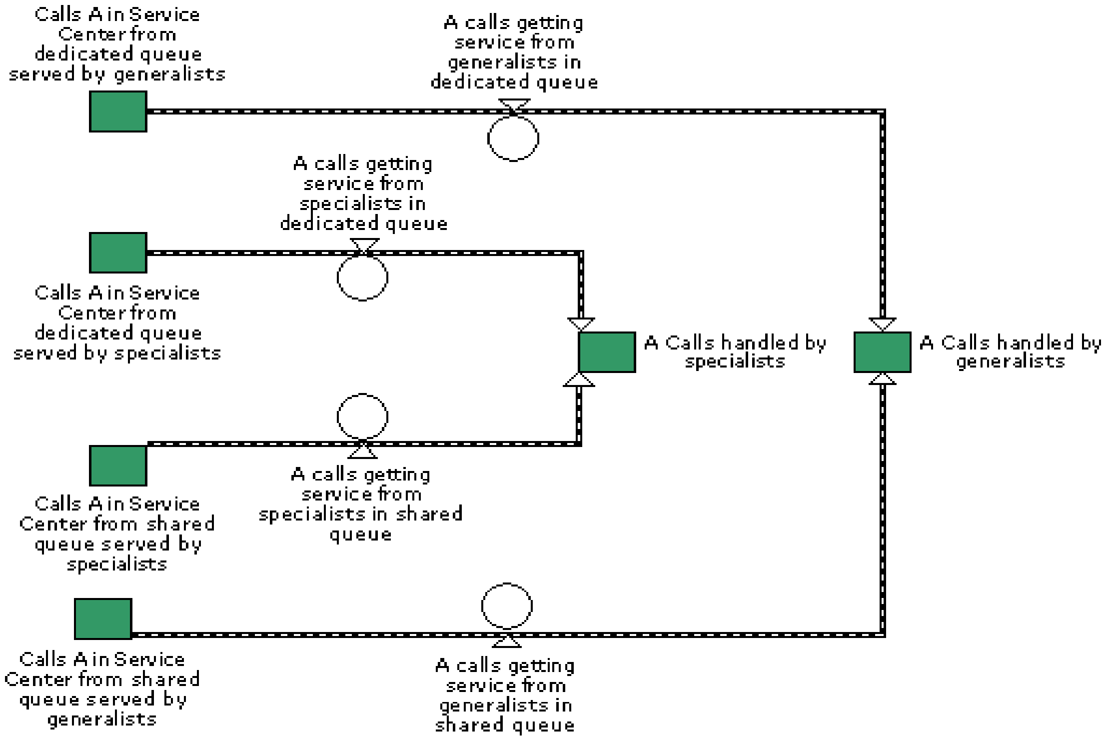

Figure 15 - A type calls completing

The part of the model showed in **Figure 16** represents the behaviour of the agents. It closely resembles the block diagram, representing the flows that from the central state of free agents go to and come from the peripheral conditions of On Pause, Working and Reserve (that is the agents assigned to other tasks).

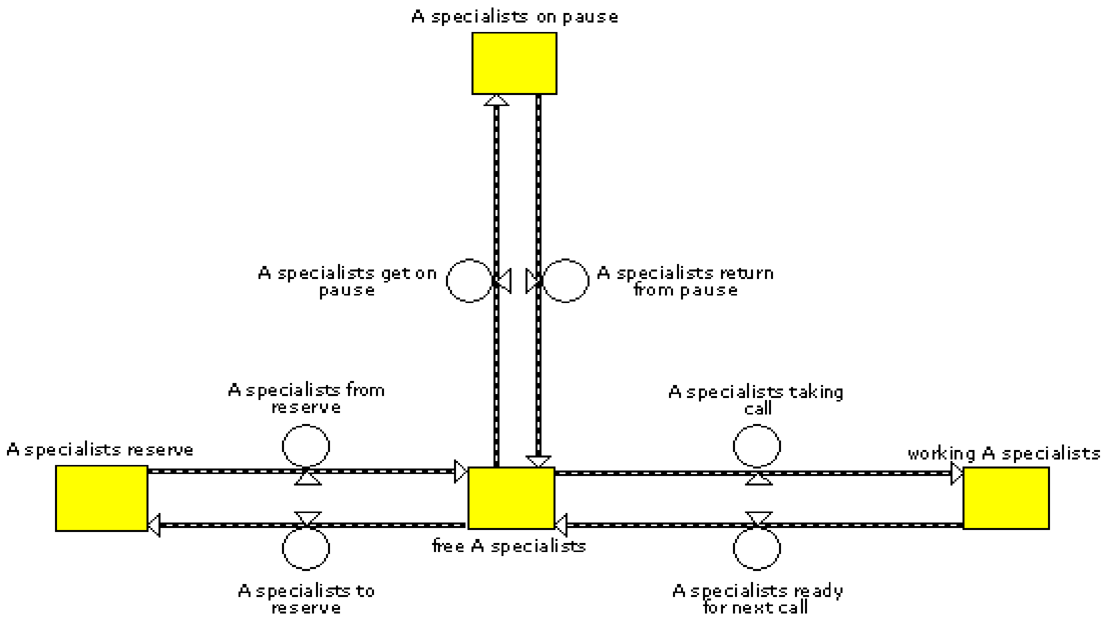

Figure 16 - A specialists behaviour

Model Application

Starting from a real case study, we used the model to show how the performance of the CC can be improved implementing the Dynamic Skill Based Routing. In the analyzed case, the CC had to process two different types of call, a "premium" type (that we will call type A), and a standard call (that we will call type B). Premium successful calls are worth 10€ each for the Call Center, while standard successful calls are worth 6€; every other characteristic (ATT, ACW) is the same. Hence, whenever possible, Premium calls should have the priority.

In our simulations we assumed the following parameters for the both call types:

Average Talk Time (seconds):	*180*
Number of Calls per half hour:	*250*
Average time (seconds) due to After Call Work:	*30*
Service Level objective:	*80% calls answered in 20 sec.*
Simulation Duration Time (seconds):	*2700*

First, we considered the performance of the CC while using Static SBR, and then we verified the improvements deriving from the implementation of Dynamic SBR, keeping the staff at the same level. For the both cases we did a set of 30 simulations.

The results are shown in the following table:

	A type calls answered	B type calls answered	Service Level	Quality Factor	Profit (€)
StaticSBR	356	371	0,76	0,83	2554
DynamicSBR	372	349	0,85	0,89	3497

As we said before, the Quality Factor represents an indicator that considers both speed and quality of service, and the quality of the system on economical terms.

The results show that while implementing the Dynamic SBR we have an increase in the A type calls answered, while B type calls answered decrease. This is not to be considered negative, since the A calls have a higher value than the B calls, and the Dynamic SBR allows to manage priorities among the calls in a better way than the Static SBR (in this case the model processes more low-value calls than high value ones).

Having part of the staff that can answer both types of calls means that we will be able to optimize the staff work, shortening the average time to answer and then improving the service level. In our case, we even manage to pass from a service level under the objective to a service level much higher than what we wanted to achieve. The profit rises, too, meaning that not only we can manage to answer quicker, but we can give also better quality answer (due to the fact that quicker answers mean less abandoning, less retrials, less call volume and hence less occupancy for the agents); the quality factor rises, too, as a further indicator that the performance of the call center is improving.

Here are some graphs that show the outcome of the two call types with both the policies during one of those simulations; the x axis is the time axis.

Figure 17 and 18 show what happens to A type calls, with Static SBR first, and Dynamic SBR second. We see that in the first case we have a rather constant percentage of unsatisfied customers throughout the simulation, while there is a consistent increase in abandoned calls in the second half. This means that the agents begin to slow down their answer rate, due to an excessive fatigue.

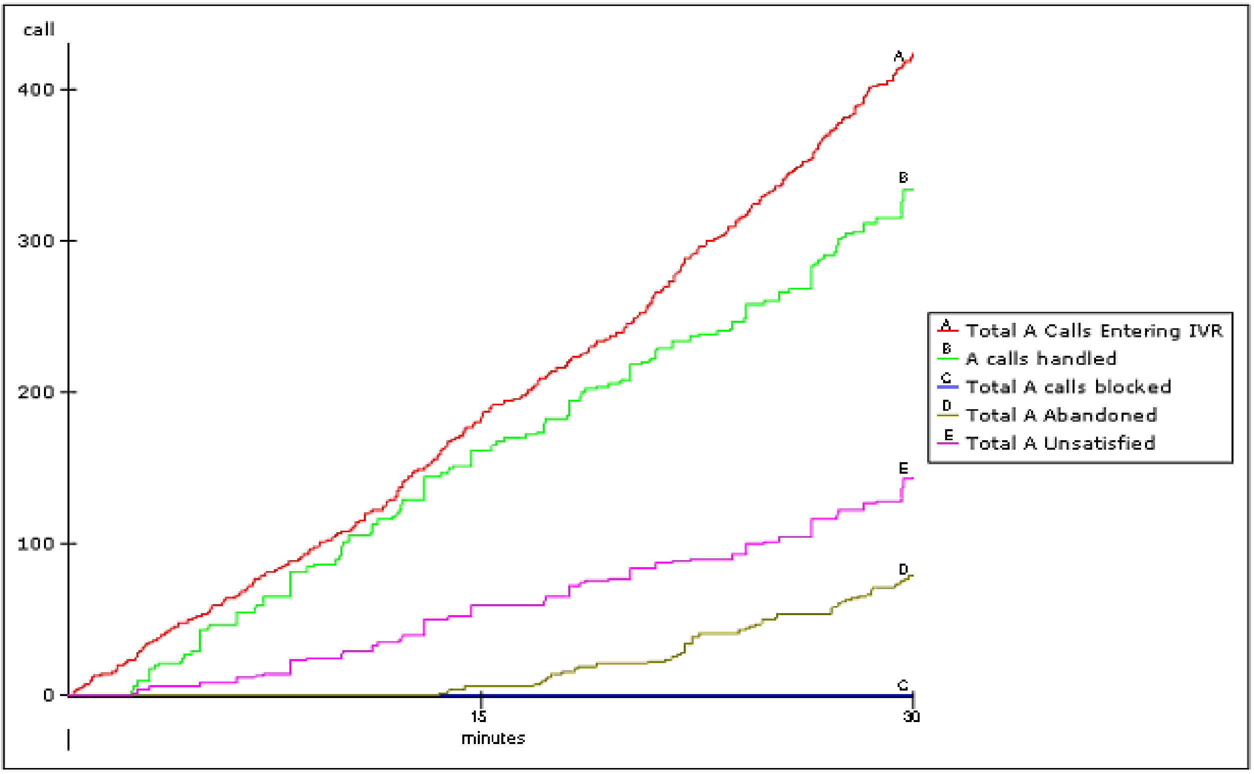

Figure 17 - A type calls outcome with Static SBR

Applying the Dynamic SBR we have a good decrease in the unsatisfied calls, and what's more interesting, we see that the abandoned calls have almost completely disappeared (**Figure 18**). This is a good sign, since it denotes the fact that the staff can now manage the call load more easily, without suffering from excessive stress after a warm up time.

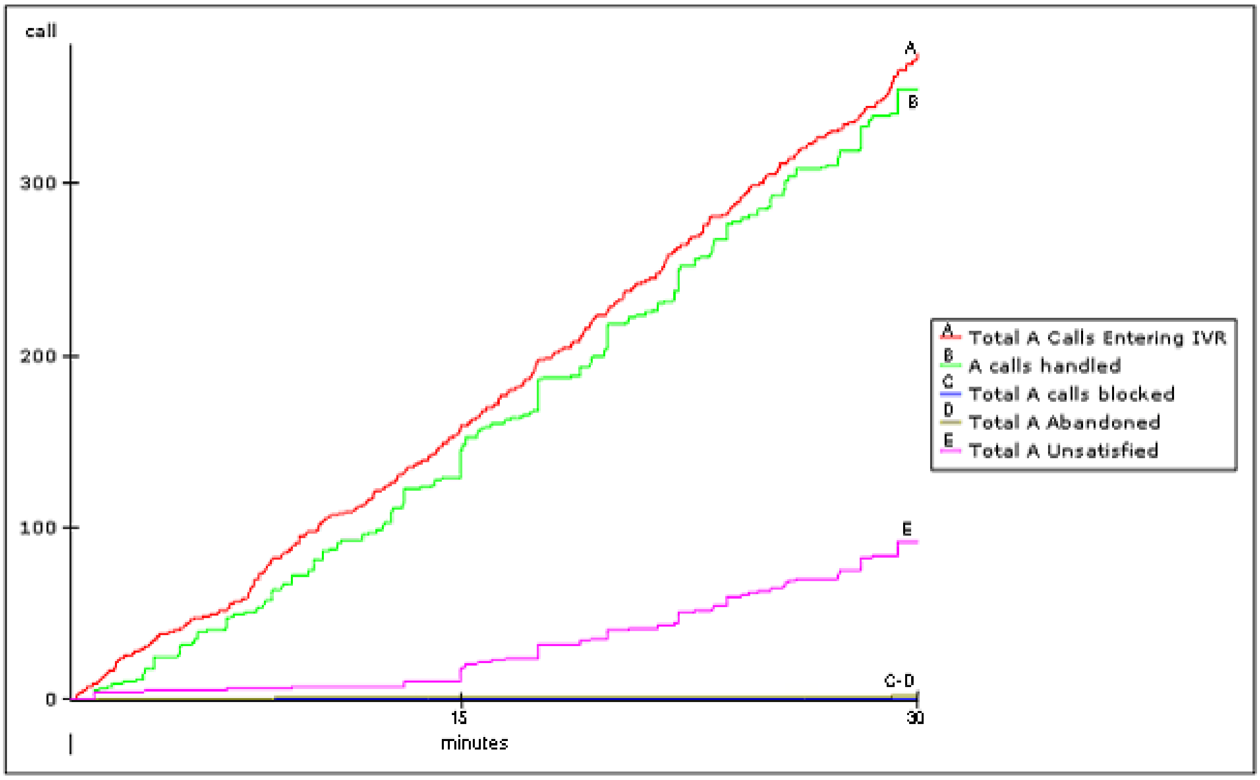

Figure 18 - A type calls outcome with Dynamic SBR

In the Static SBR B type calls follow almost the same pattern as the A type calls, with abandoning raising in the second half of the simulation (**Figure 19**). Note that even if the B type calls are less important than the A type calls, their abandoning process begins later. This is because in Static SBR the two groups of agents are completely separated, hence it is not possible to focus all the staff efforts on the most important calls, should the need arise.

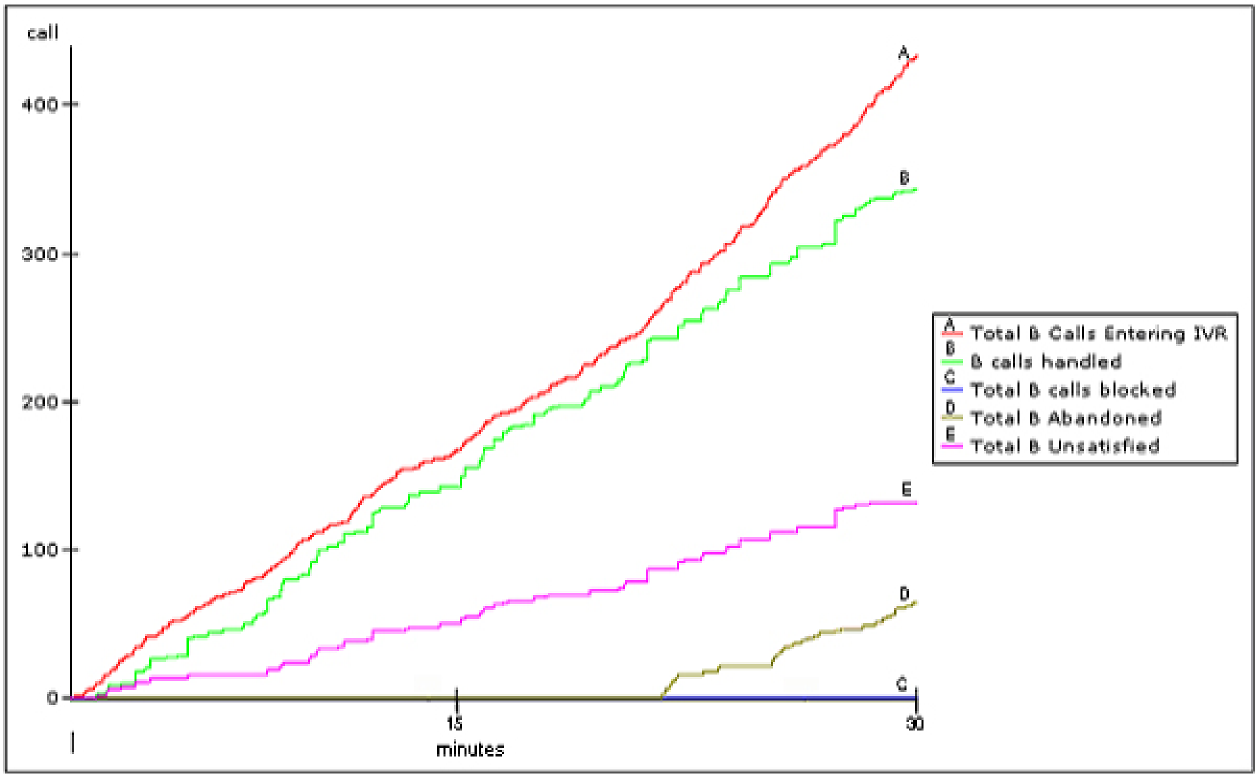

Figure 19 - B type calls outcome with Static SBR

As happened for the A type calls, applying the Dynamic SBR has brought an improvement on the calls' outcome. The unsatisfied calls have been reduced, ant the abandoned calls have almost disappeared (**Figure 20**). We still have a very little percentage of abandoning for B type calls, but this is explained, since, being part of the staff resource in common between the two call types, some priority is given to the most profitable A type calls, to the expenses of the less important B calls.

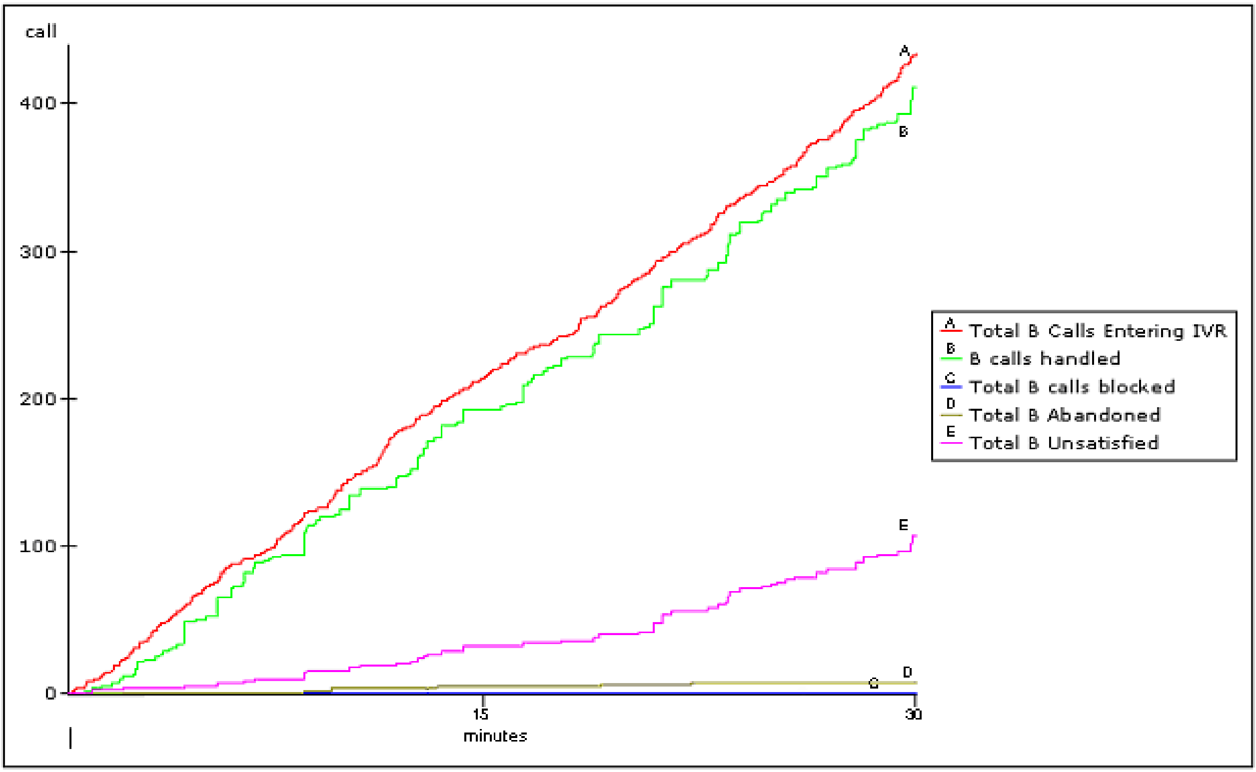

Figure 20 - B type calls outcome with Dynamic SBR

Figures 21 and 22 show the calls receiving an answer that end at a given time, respectively with Static and Dynamic SBR. We can see that in Static SBR the number of completed calls slowly decreases; this is another consequence of what we saw in **Figure 17 and 19**, since agents at the end of the simulation become slower, due to the increasing occupancy. On the other hand, with Dynamic SBR calls keep getting an answer at a steady rate, and in our case they even increase in time.

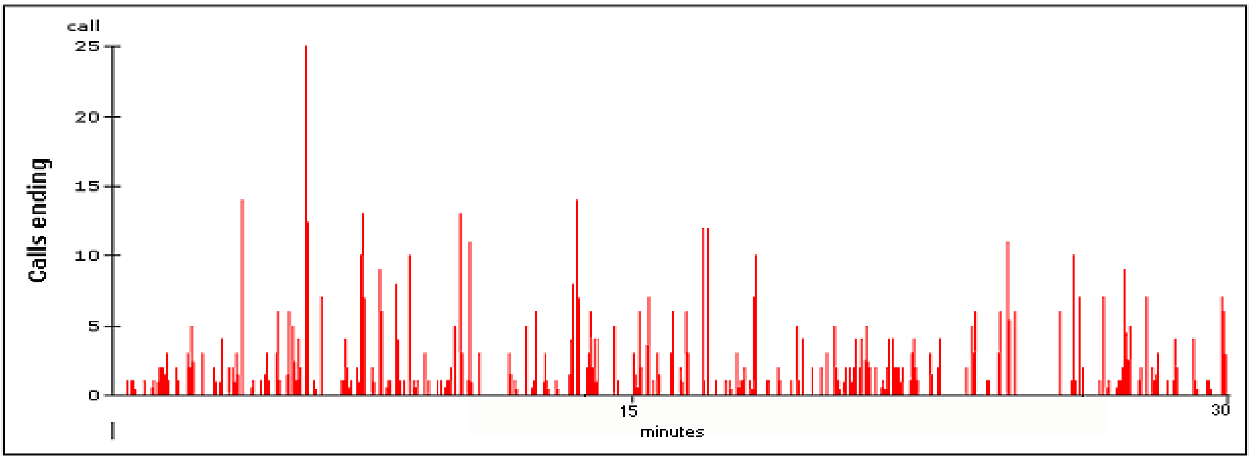

Figure 21 - Calls' ending progression with Static SBR

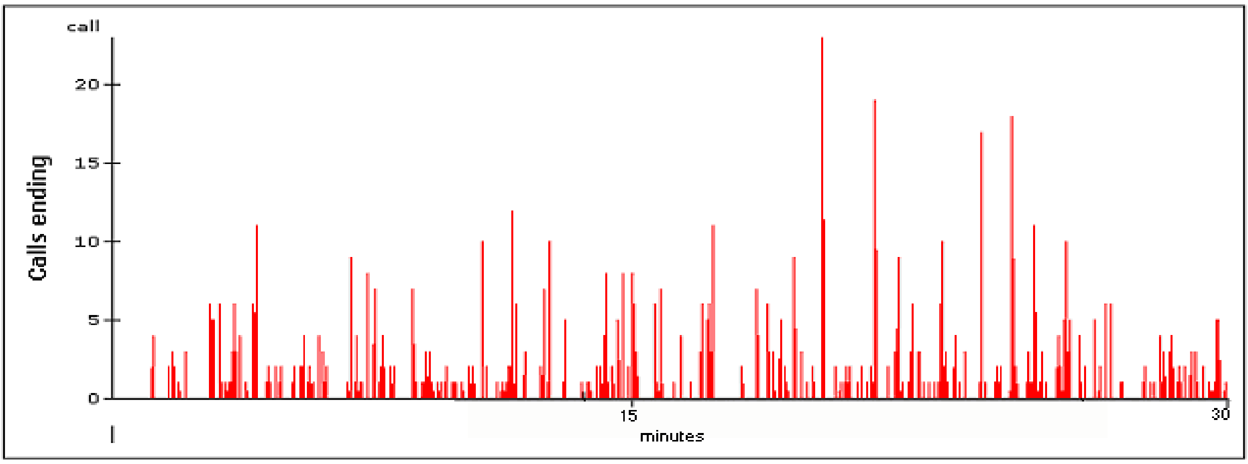

Figure 22 - Calls ending with Dynamic SBR

Figures 23 and 24 show the number of customers in contact with an agent in every moment, with Static and Dynamic SBR. With Static SBR the calls barely get over 40; this means that a big part of the agents' working time is wasted in delays caused by the high level of occupancy (**Figure 23**). With Dynamic SBR, on the contrary, the number of calls receiving answer keeps rising, meaning that using Dynamic SBR we allow the same number of agents to process a greater number of calls.

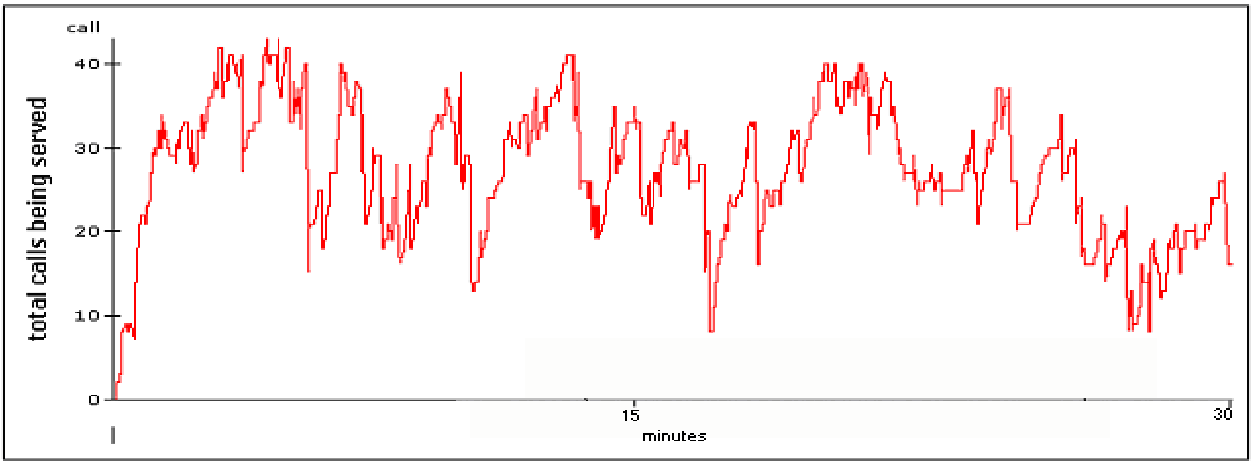

Figure 23 - Calls being processed with Static SBR

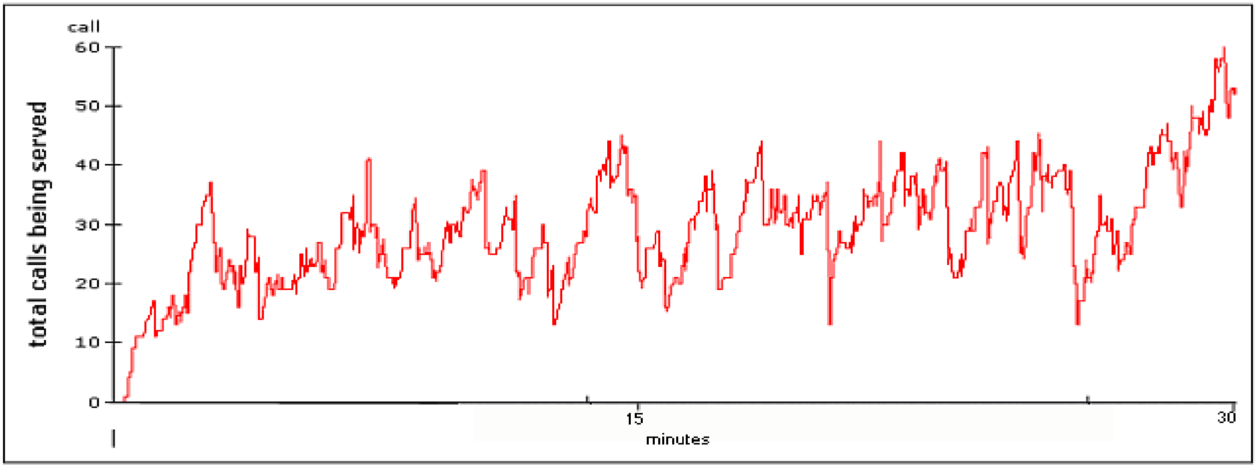

Figure 24 - Calls being processed with Dynamic SBR

Conclusions and future work

In this paper we have shown how a simulation model can be a useful support tool to a service system manager, particularly a CC manager, and can be used to examine the over time behavior of e rather complex dynamic systems. We also tried to show how the use of Dynamic Skill Based Routing can really improve the performance of a Call Center. It is important to note that this improvements haven't been achieved increasing the number of agents, but simply optimizing the staff management, and thus are accessible to most, if not all, the Call Center managers.

We have achieved some of the objectives we pointed out in the past article [1], but there is still potential work to be done.

First of all, one useful development would be to focus on a strategic level, extending the simulation length and introducing aspects like the long term burnout, or schedule management, making it possible to take into account such long term aspects as the influence of the learned experience in the Skill Based Routing, aspects that are not considered in our model, since it is oriented on an interval of 30 minutes.

Secondly, further work could be developed in order to allow the model to communicate with other computer systems making it possible to exchange automatically inputs and results.

Also, it would be useful to make the model more user-friendly, making it usable by non-specifically trained users, too; in fact, the main inputs (size and distribution of the staff, trunk sizes and so on) can be modified intuitively through a series of sliders, and the most important parameters are visualized through some graphs, but should someone operate on the inner parameters of the model (i.e. the balking or reneging distribution, among others), a good understanding of the program would be necessary.

References

[1] Caramia Massimiliano, Armenia Stefano, Onori Riccardo, Giannunzio Valerio, *Service Quality and Customer Abandonment: a System Dynamics approach to Call Center Management* (2003)

[2] Cleveland Brad, Mayben Julia, *Call Center Management on Fast Forward* (1997-1999), Call Center Press - A division of ICMI, Annapolis, Maryland

[3] Anton Jon, Bapat Vivek, Hall Bill, *Call Center Performance Enhancement – Using Simulation and Modeling* (1999), Ichor Business Books – Purdue University Press, West Lafayette, Indiana

[4] Shumsky Robert A., *Approximation and analysis of a call center with flexible and specialized servers* (2004), OR Spectrum, Springer-Verlag

[5] Stolletz Raik and Helber Stefan, *Performance analysis of an inbound call center with skill-based routing* (2004), OR Spectrum, Springer-Verlag

[6] Wallace Rodney B. and Whitt Ward, *Resource pooling and staffing in call-centers with skill based routing* (2004)

[7] Adria Marco and Chowdhury Shamsud D., *Centralization as a design consideration for the management of the call centers* (2003), Information & Management, Elsevier

[8] Busacca B., Valdani E., *Customer based view* (1999), Finanza, Marketing e Produzione, Anno XVII

[9] Sterman J.D., *Business Dynamics. System Thinking and Modelling for a Complex World* (2000), Mc Graw Hill