Counter-Intuitive Managerial Interventions in Complex Systems

Kambiz Maani
Anson Li
The University of Auckland Business School
Private Bag 92019 Auckland
New Zealand
Phone: (64-9) 373-7599, voice 88813
Fax: (64-9) 373-7566
k.maani@auckland.ac.nz
akt.li@auckland.ac.nz

Abstract

Research, as well as decades of working with managers across diverse cultures, nationalities, and industries, has revealed consistent patterns of counter productive decision making in complex systems. In this regard, managers appear to exhibit an unmistakable tendency to "over intervene" in the systems (companies, organizations, communities, etc) for which they are responsible hence generating unnecessary fluctuations and instability in their organizations. Maani, et al (2004), and Sterman, et al (1989; 2000) have studied these phenomena in experimental and simulated environments respectively. Anecdotal evidence, as well as research results, highlights a number of mental models and assumptions commonly held by managers. These are outlined below:

1. Dramatic change should lead to dramatic (positive) results. Our research shows that often the opposite happens.
2. The more change initiatives (interventions), the better the results. Again our research shows that "over-intervention" is counter-productive.
3. Managers often ignore "soft" variables (eg, morale, stress, burnout, loyalty, etc) to the detriment of their organizations. Yet, "soft" variables are powerful lead indicators of performance.
4. Managers are often oblivious to "systems delays". Lack of awareness/attention to delay undermines performance and inhibits system stability.
5. Organizations and managers often judge performance by short-term results. Experience shows that expectation of short-term results is unrealistic and misleading and can lead to counteracting outcomes as performance often declines before it improves.
6. Organizations and managers tend to use too many performance measures (ie, KPIs). Since "what gets measured gets done", excessive and misguided measures can lead to poor results and unexpected consequences.
7. It is not enough to know *what* actions need to be done. *Order* and *timing* of actions are as important as the actions themselves.

The propositions outlined above collectively form the research questions posed in this paper: "Why managerial interventions in complex organizations often produce counterintuitive failures?" In this research, empirical evidence using realistic simulation models of organizations (microworlds) provides supports and sheds light on the above propositions. The results confirm findings from recent longitudinal case studies of organizations by Collins (2001). Research subjects comprise MBA and graduate business students and practicing managers. The paper deals with systems thinking theory and complex decision-making and their implications for transforming managers and organizations towards sustainable performance.

Key words: Complex Decision-Making, Dynamic Behavior, Change Management

Introduction

For every complex question there is a simple answer, and it is wrong.
(Business week 21 April 1980, p.25)

In the past three decades, much research has explored the complexity of decision-making under the 'bounded rationality' of human mind. This includes studies by Simon (1957, 1979, 1987), Morecroft (1983, 1985), Senge (1990), and Sterman (1989, 2000). The latter three have related this dilemma to systems thinking theories. According to Richmond (1994), systems thinking is "the art and science of making reliable inferences about behavior by developing an increasingly deep understanding of underlying structure". By understanding problem situations with a systems perspective, a more holistic understanding can be achieved in terms of the causal relationship between decisions, interventions, and their expected results. Under bounded rationality, it is not realistic to expect that interventions will yield the expected (and only the expected) results. Further, decisions made with good intentions do not always result in the favourable outcomes anticipated by the decision maker. Managers often mistake common causes of problems (variations due to systems) for causes attributable to workers/staff (special causes) creating low morale and costly interventions (Deming, 1982).

Likewise, a substantial amount of research has been carried out in relation to the dynamics of decision making with a systems thinking perspective. This includes "Limits to Growth" (Meadows, 1972), "System Dynamics: Portraying Bounded Rationality" (Morecroft, 1983), "Beyond the Limits" (Meadows, 1992), and the "Improvement Paradox" (Keating et al., 1999). While these studies provide significant insight into the formulation and outcomes of decisions, the empirical work in the area of decision dynamics and interventions in complex systems remains elusive.

Research Objectives

The aim of this study is to address the seven propositions listed below: To our knowledge there is no cohesive body of research that deals with these propositions.

1. Dramatic change leads to dramatic (positive) results
2. The more change initiatives (interventions), the better the results
3. Managers often ignore "soft" variables (eg, morale, stress, burnout, loyalty, etc)
4. Managers are often oblivious to "systems delays"
5. Organizations and managers often judge performance by short-term results
6. Organizations and managers tend to use too many performance measures
7. It is not enough to know *what* actions need to be done

These propositions arise from previous research, summarized in the following section and as our own observations of participant's behaviors in simulated learning environments in executive education programs.

We employ empirical testing with informed participants using simulation microworlds. Research subjects comprise MBA and other graduate business students, practicing managers and in some cases final year high school students.

Through a deeper understanding of the dynamics of decision making the research seeks to identify and derive "patterns of decision-making" which would assist in effective formulation and implementation of policies and strategies.

Bounded Rationality

According to Simon (1957), "bounded rationality is a property of decision making that reflects people's cognitive limitations. Individuals faced with complex choices are unable to make objectively rational decisions". The reasons for this are as follows:

1. They cannot generate all the feasible alternative courses of action;
2. They cannot collect and process all the information that would permit them to predict the consequences of choosing a given alternative; and
3. They cannot evaluate anticipated consequences accurately and select among them.

Morecroft (1983) carried out a study on the philosophy of human decision making expounded by the Carnegie School.[i] "Underlying the work of the School is the powerful notion that there are severe limitations on the information processing and computing abilities of human decision makers. As a result, decision making can never achieve the ideal of perfect (objective) rationality, but is destined to a lower level of intended rationality." (Morecroft, 1983)

Along with the above arguments related to bounded rationality, Morecroft (1982, 1985) identifies six common practices that underlie the shortcomings of the human decision making process. They are:

1. Factored (fragmented) decision making

Complex issues are divided up into pieces (e.g., disciplines, sections, departments, etc) to facilitate decision-making, as "they cannot be handled by an individual".

2. Partial and certain information
Decision makers tend to use "only a small proportion of the information that might be relevant to full consideration of a given situation". They would also "avoid the use of information that is high in uncertainty". This tends to focus the decisions on problem symptoms and locally optimum solutions.

3. Rules of thumb / Routine
This refers to situations where decision makers, under time pressure, resort to "quick fixes" in order to rectify a situation as quickly as possible. Quick fixes often result in "backfire" or unintended outcomes.

4. Goals and incentives
Focus on certain goals and incentives could compromise other areas and undermine the performance of the larger system.

5. Authority and culture
Culture and tradition provide powerful predetermined frameworks for decision makers (i.e. mindset, mental model). Through customary routines and commands, prevailing values and traditions are transmitted to all and hence get reinforced and further ingrained.

6. Basic cognitive processes
"People take time to collect and transmit information. They take still more time to absorb information, process it, and arrive at a judgment. There are limits to the amount of information they can manipulate and retain. These cognitive processes can introduce delay, distortion, and bias into information channels."

To deal with the above shortcomings, many authors have suggested ways to improve the effectiveness of human decision-making. These include, among other tools, management and computer frameworks (Gilberto, 1995, Cayer, 2001), computer simulation models (Simon, 1987, Sterman, 1988), and the use of systems thinking in decision-making (Senge, 1990, Maani, et al 2004).

Misperception of Feedback

A classic work in this area is Sterman's research (1989) in relation to the "misperception of feedback". A simulation model, known as the "Beer Game", was used with groups of participants to investigate their interpretation of information feedback and the effects on the interventions derived.

"The decision making task is straightforward: subjects seek to minimize total costs by managing their inventories appropriately in the face of uncertain demand." (Sterman, 1989) In such a "simple" environment, however, things did not always go as planned for most participants, due to the rich simulated environment, which contains "multiple actors, feedbacks, non-linearities, and time delays." (Sterman, 1989) Similar to Morecroft's and Simon's idea about factored decision making, "the interaction of individual decisions with the structure of the simulated firm produces aggregate dynamics which diverge significantly and systematically from optimal behavior" (Sterman, 1989).

The findings of the study are summarised into the following points (Sterman, 1989):

- Subjects failed to account for control actions, which had been initiated but not yet had their effect.
- Subjects were insensitive to feedbacks from their decisions to the environment.
- The majority attributed the dynamics they experienced to external events, when in fact these dynamics were internally generated by their own actions.
- The subjects' open-loop mental model, in which dynamics arise from exogenous events, is hypothesized to hinder learning and retard evolution towards greater efficiency.

The Improvement Paradox

Keating et al. (1999) carried out a study of the effectiveness of improvement programs. The motivation for the study arose from the fact that "most attempts by companies to use them have ended in failure" (Easton and Jarrell, 1998 in Keating et al., 1999) and that even "successful improvement programs have sometimes led to declining business performance, causing layoffs, low morale, and the collapse of commitment to continuous improvement." This dilemma was termed the "Improvement Paradox".

The study was carried out on major companies to understand why improvement programs often fail. The findings suggest that "the inability to manage an improvement program as a dynamic process – one tightly coupled to other processes in the firm and to the firm's customers, suppliers, competitors and capital markets – is the main determinant of program failure. Failure to account for feedback from these tightly coupled activities leads to unanticipated, and often harmful, side effects that can cause the premature collapse and abandonment of otherwise successful improvement programs." The study, however, does not suggest that improvement programs are ineffective in terms of improving organizations. In fact, the authors point out that "firms with developed quality programs significantly outperform their counterparts in profitability, share price and return on assets." The problem lies in the suitability of these programs and the style with which they are implemented.

Good to Great

Jim Collins (2001) together with a research team studied 1,435 "good" companies' performance over 40 years, and identified a mere 11 organizations who had successfully transformed themselves from "good" to "great". In this multi-year study Collins and his research team scrutinized the improvement / change strategies of these companies and identified unique styles of "change" that contributed to the success of the great companies.

From this study, Collins and his team have come up with several "*myths*" about change management:

- That effective change can be facilitated by Change Programs such as launch events and cascading activities.
- That change starts only when there's a crisis that persuades "unmotivated" employees to accept the need for change.
- That stock options, high salaries, and bonuses are incentives that grease the wheels of change.
- That the fear of being left behind, the fear of watching others win, and the fear of presiding over monumental failure are drivers of change.
- That you can buy your way to growth, so it figures that you can buy your way to greatness.
- That the breakthrough that you're looking for can be achieved by using technology to leapfrog the competition.
- That big change has to be wrenching, extreme, and painful.

All of these have proven to be wrong. None of the above "necessary" factors were found in the 11 companies that have managed to transform from good to great. Collins applies two analogies to illustrate how change happens.

The Egg (Transformative change is not visible):
Picture an egg, being hatched. It is sitting there, motionless, no action could be seen. Then all of a sudden, the shell cracks and out pops a chick. Everyone is *surprised* from the sudden change from an egg into a chick. In the egg's perspective, however, that moment of change is just "simply one more step in a long chain of steps that had led up to that moment". Even though it was seemingly dormant, the life form inside has evolved, grown and developed, before the shell cracked open. In organizations, changes are often identified in the wrong places. If a company is focusing on achieving just the "shell cracking" moment, then it is not likely to succeed.

The Flywheel (Slow but persistent action):
Imagine a big steel flywheel of 100 feet in width and 10 feet thick, weighing 25 tons, sitting at a standstill. You are responsible to get it up and spinning about the axle. At the beginning, a tremendous amount of effort is required to start its motion, but through consistent application of force, albeit small, the wheel starts to reinforce its own motion through momentum. In organizations if the correct "force" is applied, then the change and success will reinforce itself, without the requirement of big efforts or dramatic interventions. In contrast, over-hyped change programs often fail, since they lack accountability, they fail to achieve credibility, and they have no authenticity. It's the opposite of the Flywheel Effect; it's the Doom Loop.

Leadership and Decision Making

More recently, Moss Kanter (2003) reports on several companies which have experienced major declines in their fortunes, declines which have been successfully reversed by the interventions of their new CEOs.

These companies, although from different industries and differing in size, experienced similar patterns of decline in their business. Often decisions were made by various functions or divisions (as in "factored decision making" [ii]) to employ quick fixes (as in "rules of thumb") to various problems in order to achieve short-term goals within tight time limits (as in "goals and incentives"). For example, a common practice at Gillette was to offer "discounts to retail customers at the end of a quarter in order to move products and achieve sales targets, thus sacrificing margins and jeopardizing the next quarter's sales".

The author has suggested that the use of common practices as rules of thumb (as in the Gillette case) is very common in troubled companies. These short-run solutions usually make the situation worse in the longer term creating a "Fixes that Fails" archetype. For instance price cuts from discounts, although they would be effective in increasing sales, would also reduce the funds available for marketing, which increases the organization's reliance on the promotional deals. Customers will also know that they can wait until quarter's end to get even better deals..

The resulting deterioration in morale and work culture can be termed "learned helplessness". People in the organization feel that there is little they can do to make a difference in the company and therefore become passive. This in turn reinforces the decline of the organization - a vicious cycle that could lead to ultimate downfall.

Research Methodology

In order to explore the complex dynamics of managerial decision making, an experimental research approach using management flight simulators (MFS) has been used in this study.

The use of simulation models in experimental research as an alternative to laboratory and field experimentation has become common (see for example Bhuiyan, 2004; Thompson, 2004). In such experiments, the subjects participate as decision makers in simulated environments where "manipulation [of independent variables] and control are possible … [and] the course of activities is at least partly governed by the participants' reactions to the various stimuli as they interact among themselves." (Sekaran, 2000)

Experimentation Models

The experimentation models used in this research are two computer simulations: Service Quality Microworld (developed by MIT System Dynamics Group), and the Brand

Management Microworld (developed by Strategy Dynamics Ltd.). These microworlds have been used by the first author in executive courses for several years.

These models, while representing different business situations, employ similar user interface and decision making mechanism. However, the underlying business situations and the nature of the decisions to be made by the participants are significantly different.

Service Quality Microworld (SQM)

The Service Quality Microworld simulates the operations of a generic service company. The simulation starts at a "steady state" where "output variables" such as incoming orders, orders completed, work backlog, rework, hiring, personnel turnover, time pressure (employee), monthly profit, and monthly expenses are held at a constant rate. Appendix A shows the partial Causal Loop Diagram showing the key dynamics in the model.

During the experiments, participants can manipulate the values of three "input variables" (along the course of 60 months) in order to achieve certain goals, such as maximizing cumulative profits, minimizing rework, or maximizing production. The decision variables are monthly "Net Hiring", "Production Goal", and "Quality Goal". By intervening any/all of these three input variables, various output variables will be affected through complex and dynamic relationships among them. The simulator generates a number of KPIs in the forms of graphs and reports.

Brand Management Microworld (BMW)

The Brand Management Microworld simulates a business organization at the start up, introducing a new brand of drinks in an established market. The participants are given a "launch budget" (£20 million by default) at the beginning which they may utilize throughout the course of the product's launch (12 years). Output variables include consumer awareness of the brand, sales, stores stocking the brand, advertising campaign reach, monthly profits. Since the model represents an organization at the start up, unlike SQM, the model does not begin at the steady state. Appendix B shows a partial causal loop diagram for the BMW model.

During the simulation, participants can manipulate the values of three "input variables" (along the course of 12 years (144 months) in order to achieve certain goals, such as maximizing profits, maximizing sales, maximizing the number of stores stocking the brand. These input variables are "Wholesale Price", "Advertising per Month", and "Size of Sales Force". By changing any/all of these three input variables, various output variables will be affected through the complex and dynamic relationships among them. A large selection of KPIs is also available in the forms of graphs, tables, and reports.

Decision Variables

Table 1 summarizes the decision variables of the two microworlds.

Table 1- Comparison of Decision Variables in Two Microworlds

Service Quality Microworld		**Brand Management Microworld**	
Input Variable	Type	Input Variable	Type
Production Goal	Target	Wholesale Price	User Defined Value
Quality Goal	Target	Advertising Spending / month	User Defined Value
Net Hiring	Flow	Sales Force Size	Stock

The two "target" variables in SQM allow users to set goals for these variables. However, the "actual" performance of these variables does not often match their targets, as can be seen in the examples shown below. As shown in Figure 1, the *actual* Production and Quality are consistently below their corresponding goals. This discrepancy is a consequence of the endogenous decision polices of the subjects and their interactions with model variable. In these examples, this discrepancy is generated by an under-capacity of available HR leading to high pressure for the employees, with further attritions exacerbating the staff shortage (a reinforcing vicious cycle).

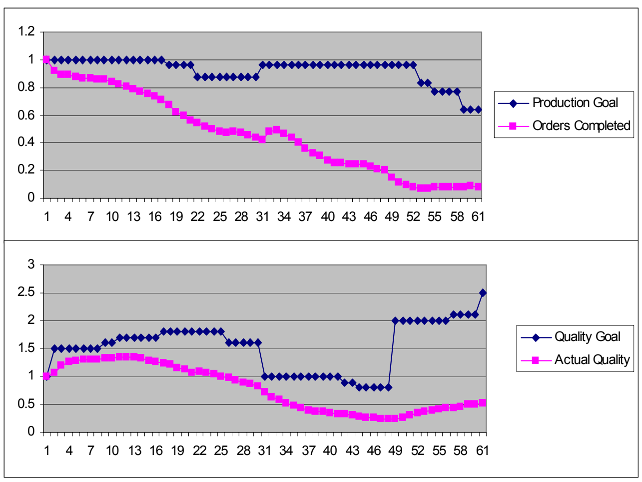


Figure 1 - Sample output from SQM Microworld

In contrast, the "user defined" variables in BMW, are the actual values defined by the participants. That is, the values for "Wholesale Price" and "Advertising Spending / Month" are set by the participants and directly impact the model.

Likewise, the workforce variables for both models ("Net Hiring" for SQM and "Sales Force Size" for BMW) while related to human resources, are of different nature and have different meanings. The "employees" in SQM directly influence production level and quality whereas the "Sales Force" in BMW is responsible for persuading stores to stock their brand of drinks, which influences sales indirectly. Further, "Net Hiring" is a "flow" variable which affects the rate of change in personnel level, whereas "Sales Force Size" is a "stock" variable which represents the level of employees.

The above difference in structure and the decisions variables of these Microworlds will enhance the generalizability of the findings of this study.

The Experiments

The data represents 223 experiments (118 for SQM and 105 for BMW). Subjects comprise undergraduate, graduate, MBA and executive students as well as some final year high school students. Each participant had worked on either one of the simulation models (not both). Each experiment involved two sets of exercises (see Appendix C for details).

In the experiment sessions, the participants were required to perform certain tasks to achieve the stated goals using the respective Microworlds. The subjects worked individually during the experiments with no breaks so no information exchange and "interaction effects" were expected to occur. Data were collected on:

- Demographical information about the participants;
- Strategies devised by participants for carrying out the task(s) and/or achieving the goal(s) in the simulation model;
- Actual decision making carried out in the experiment on the simulation model;
- Outcomes and results on the simulation model; and
- Participants' interpretation and comments relating to the decision making and outcomes/results.

The above include both quantitative and qualitative data, which allows a triangulated perspective of the research questions.

Measurement Variables

Following the experiments, the strategies of the participants were examined by scrutinizing the graphic outputs of their KPIs, the three input variables, and other measures computed in the simulation.

For comparison purpose, the values of all decision variables used during the simulation, and were recorded by the subjects themselves, were converted into indices (with a base value of 1 for the initial default value). For example, the Magnitude variable is calculated as the average change in the index during the simulated period. Hence, the larger the index, the larger the degree of intervention made by the subject in the system.

Furthermore, the outcome patterns of each experiment were compared against the original strategy developed by the participants to examine whether:

1. They had adhered to their original strategy throughout the experiment; and
2. To what extend the simulation results were consistent with their anticipated outcomes.

Any discrepancies from the original strategy and their expectation were noted and studied closely to find out more about the mental model of the participant, and how the information 'feedback' influenced the participants' decisions during the course of the simulation.

The Findings

Magnitude and Frequency of Change

To start with, a cursory examination of the experiment results revealed some consistent and compelling patterns of causal relationships. In particular, strong and consistent patterns was observed between the dependent variable (Cumulative Profit) and the frequency of decisions/interventions; the extent of change (the "Magnitude" variable), and the number of decision variables used in the interventions ("#Variables").

In order to test these observations statistically, two multiple regression models (for SQM and for BMW respectively) using Magnitude, Frequency, and #Variables as independent variables against Cumulative Profit as dependent variable were run. The multiple regression results are summarized in Table 2.

Table 2 – Multiple Regression Results

Service Quality Microworld

n	R^2	F	Sig. F	Variable	Coefficient	t-stat	p-value
118	0.20	9.791	0.0000	Intercept	-661,780	-0.19	0.849
				Magnitude	-73,457,756	-3.11	0.002
				Frequency	-240,648	-3.33	0.001
				#Variables	6110409.463	4.00	0.000

Brand Management Microworld[iii]

n	R^2	F	Sig. F	Variable	Coefficient	t-stat	p-value
105	0.36	19.097	0.0000	Intercept	-17,431	-4.69	0.000
				Magnitude	8,343	4.13	0.000
				Frequency	-411	-3.48	0.000
				#Variables	7,756	5.23	0.000

The regression results confirmed the above observations as the models as well as the three independent variables are statistically significant with high levels of confidence. It must be noted that, in the above experiments, variations in performance are primarily due to the interaction dynamics between the input variables and endogenous model variables and do not arise from stochastic effects in the model. [iv]

In general, the above results refute both research propositions 1 and 2 - arising from the conventional wisdom and managerial truism. For Proposition 1, the results indicate that frequent interventions strongly correlate with inferior outcomes. As for the second Proposition, even though the results are statistically significant, curiously, the "Magnitude" variable appears to work in opposite directions for the two microworlds. That is, while the "direction" of the Magnitude coefficient is negative for the SQM experiments, it is positive for the BMW microworlds. That is in the case of SQM, the results show that the greater the magnitudes of changes, the worse the result, while for the Brand Management Microworld, the greater magnitudes of interventions were in fact beneficial to the outcome.

The above finding is both surprising and illuminating and provides further insights. In order to explain this 'contradiction', one needs to better understand the different nature of these microworlds. For instance, the Service Quality Microworld simulates a company operating at a *steady state*, whereas, the Brand Management Microworld is based on a newly start up business. Viewing these dynamics in light of Collins (2001) "Flywheel" analogy, the steady state model (SQM) is a "self-propelling" flywheel at the start, while the newly established organization is a flywheel at *stand still*, which requires a significant amount of effort to kick start. Thus, in the BMW case, initially, much greater effort/intervention is required to "fill" the advertising pipeline, raise public awareness, and stock the stores in order to make sales happen. Once this happens, the system can be sustained with far less effort/interventions. This is further evident in the intercept of the BMW regression where the negative sign indicates lack of intervention leads to a small *loss*. However, as Collins observes, in order to start up the flywheel, effort should be applied "gradually [and] consistently", rather than using some "over hyped" changes that "lack accountability, fail to achieve credibility, and have no authenticity".

Counter to the BMW's aggregate results shown in Table 2, in a number of experiments dramatic change did result in great failures. In these cases, though large changes would have been appropriate, the subjects utilized 'wrong' levers (variables) for intervention – hence the choice of the decision variables does matter. Further statistical analysis using ANOVA as well as detailed examination of individual simulation experiments confirms that the *choice* of decision variables does indeed make a difference in the outcomes of interventions.

In summary, the above results present counterintuitive insights and lend support for the following observations.

1) High frequency of change (Frequency) could have negative impact on the outcomes of intervention strategies.

2) Dramatic (over-hyped) change/intervention (Magnitude) does not necessarily lead to positive results and, in fact, it could be counterproductive.

The Awareness of "Soft" Variables in Performance Measure

Scrutinizing participants' strategies and reflections before and after the game play, reveals remarkable patterns in relation to the awareness of "soft" variables in decision making in complex systems.

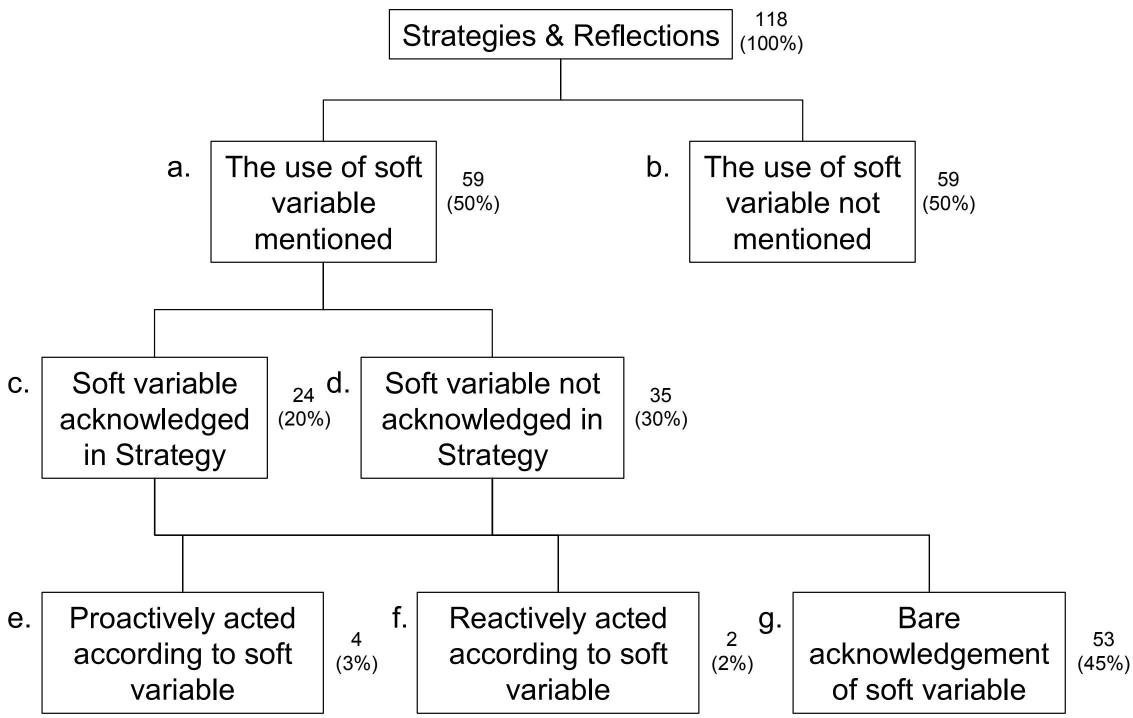
"Soft" variables are defined (Maani et al., 2000) as the "subtle, 'invisible', and powerful factors" that influence the behaviour and performance of people and organisations alike. They include such things as morale, burnout, commitment, loyalty, confidence, and care for customers and learning capacity, and can be regarded as the measures of the internal health and vitality of an organisation.

Among the vast number of variables at play in the two simulators, there exists only one specific variable that can be used as proxy for a soft variable, namely, "time pressure" in SQM. According to the structure of the model, time pressure plays an important role in the behaviour of the system. It is directly related to all three of the decision variables. Net hiring directly affects the capacity of the company, and thus, the workload of the workers. Production goal changes directly affect the work pressure, and so does the quality goal. In the SQM model, time pressure is quantified as a numerical value for the purpose of simulation, and thus, cannot be considered entirely as a soft variable. Nevertheless, this applies to all variables found in computer simulations.

Even though the time pressure variable is presented in the SQM model as a quantitative value, the nature of this variable remains qualitative. That is, it is safe to assume that in most participants' mental models, the nature of time pressure is qualitative, as is reflected in their strategies and reflections.

The analysis for the *awareness* of soft variables was carried out by observing the participants' strategies and reflections for each experiment. Participants who did/did not acknowledged the use of time pressure as an indicator of performance, and whether they used it as a proactive or reactive measure was recorded.

The coding scheme and outcomes are summarized in the following diagram:



As the above diagram shows, in a total of 118 SQM experiments, 50% of the subjects indicated an acknowledgement of the variable "time pressure" (i.e. by mentioning its effect, quoting its value, and/or trying to influence it during the simulation). Twenty percent of the sample did acknowledged time pressure in their strategies, while the other 30% only stated its usage in their reflections.

Overall, in all the experiments, only a mere 3% *proactively* managed time pressure as a critical performance measure (that the maintenance of low time pressure was stated in their strategy, and efforts were exerted towards this goal during the game play). Another 2% of all participants reactively managed it (they discovered the criticality of this variable during the simulation, usually as a result of their own poor performance during the game play, and therefore, tried to rectify the situation by easing off time pressure). Only 45% of the responses barely mentioned the existence of time pressure - quoted its value, or commented that time pressure had an effect on performance, without showing any evidence of effort towards maintaining it.

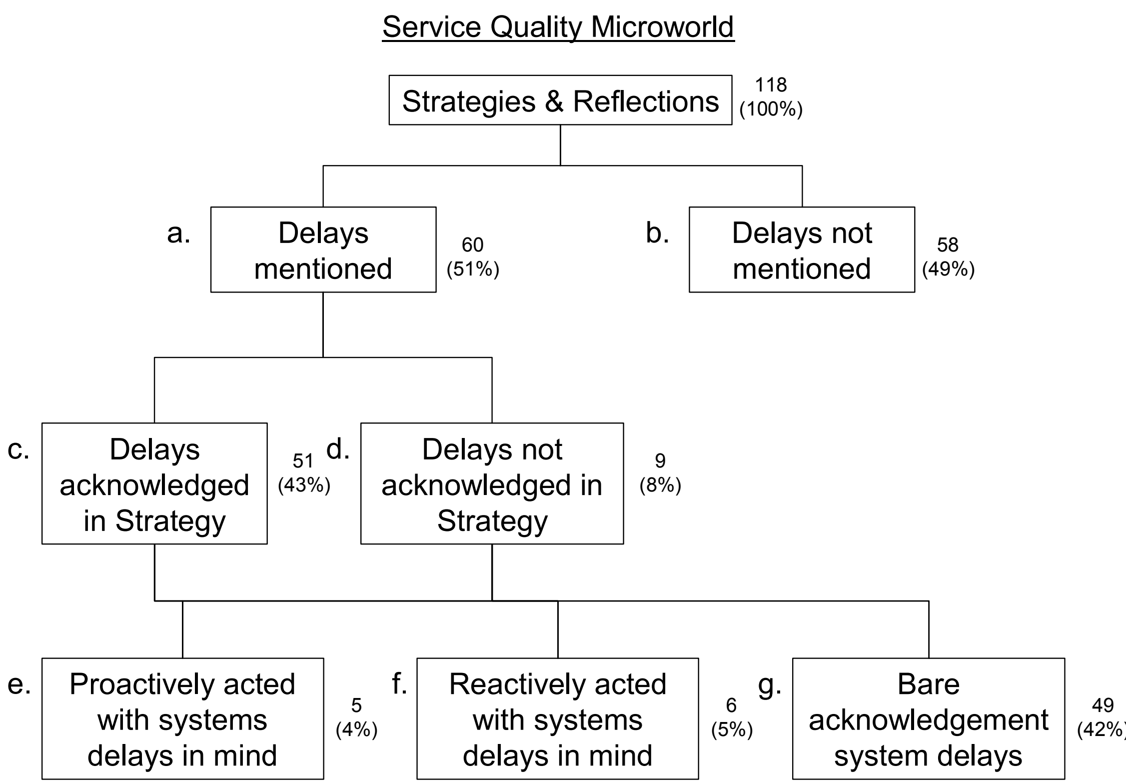
Acknowledgement of Systems Delays in Complex Environments

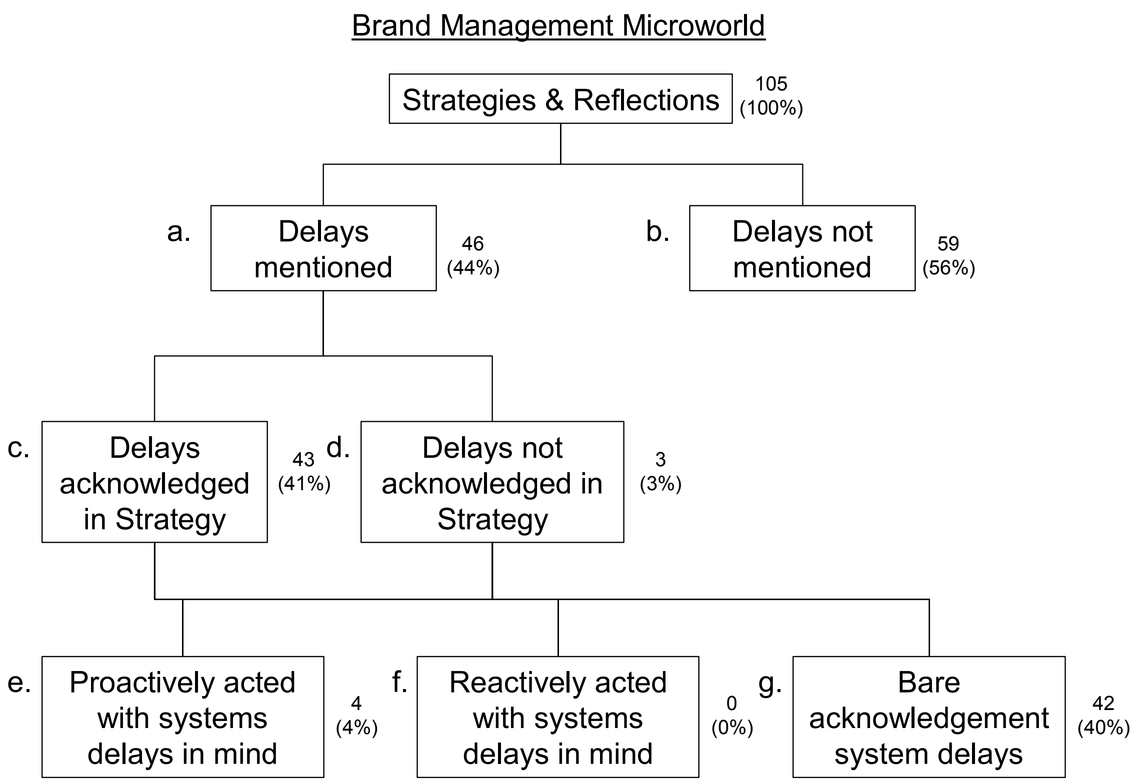
Systems delays are embedded in both the SQM and BMW models (refer to the appendix). Each of them is unique and is related to different variables within the system. For instance, in the SQM model, all three decision variables, net hiring, production goal, and quality goal have an immediate impact upon total personnel (capacity), actual production, and actual quality, respectively, as well as on time pressure. These impacts, however, are not always proportional to the magnitude of the intervention. That is, the changes in the target variables unfold over a period of time, instead of being instantaneously noticeable.

In the SQM model, there are two variables with an explicit delay in their relationship - Net hiring and Personnel mix. Any newly hired worker will only be half productive that an experienced worker, until they have worked in the organisation for a period of 2 years.

In the BMW model, delays are embedded with the application of the three decision variables, wholesale price, advertising spending, and sales force size. Wholesale price has an immediate, full scale impact on retail price, which in turn impacts the number of active customers at a non-linear rate. Advertising spending and sales force size have non-linear impacts upon consumer awareness and number of new/lost stores respectively. In this case, the effects of interventions in these three variables will have an inherent delay on the target variables, unlike the SQM case where the impact is incremental (except for the net hiring's effect on personnel expertise).

The awareness of such delays by the participants are observed and coded from the strategies and reflections in their experiments (118 experiments from SQM and 105 experiments from BMW). The coding scheme and findings are presented in the following diagrams:





As summarized below, across the two different models, the awareness of systems delays is relatively consistent.

	SQM	BMW
Acknowledgement of delays	51%	44%
Acknowledgement of delays in strategy	43%	41%
Proactively acted with systems delays in mind	4%	4%
Reactively acted with systems delays in mind	5%	0%

About half of the experiments for both models have shown that the participants are aware of systems delays. The majority of this group have anticipated this before the start of the simulation by noting possible delays that they are going to encounter in the course of their strategies. However, even though half of the experiments have shown awareness of delays, only a very small proportion (4% of the entire sample) have shown proactive actions with consideration of the delays - for example, hiring more workers early on to off set the delay required for them to become experienced, while keeping quality and production goals at a lower level to ease off the time pressure. In SQM experiments only

5% have shown effective learning from the game play, by noticing the effects of delays and, therefore, reactively adjusted their strategies with consideration of systems delays.

The Use of Performance Measures

The key variables available in each model for use as performance measure are listed in the table below. Any of these variables that were mentioned by the participants in their strategies and/or reflections for tracking their performance are considered as a performance measure in this analysis.

SQM	**BMW**
Profits	Profits
Rework	Number of aware consumers
Time pressure	Number of active consumers
Work backlog	Costs
Actual production	Sales
Personnel turnover	New stores approached
Actual quality	Number of stores lost
Personnel expertise	Number of stores
Total personnel	Brand image
Total costs	Market share
Market attractiveness	Advertising campaign reach
Waiting time	

For the purpose of the analysis, the number of performance measures used by the subjects in each experiment was recorded. A regression model was used l to statistically test the relationship between the number of variables used for performance measure as the independent variable and the Cumulative Profit as the dependent variable. The outputs of these models are presented in the following tables:

Service Quality Microworld

n	**R^2**	**F**	**Sig. F**	**Variable**	**Coefficient**	**t-stat**	**p-value**
118	0.00	0.00	0.9957	Intercept	3,140,564	0.78	0.437
				Number of Performance Measures	5,089	0.01	0.9957

Brand Management Microworld[v]

n	**R^2**	**F**	**Sig. F**	**Variable**	**Coefficient**	**t-stat**	**p-value**
105	0.05	5.986	0.0161	Intercept	6,363	1.72	0.0889
				Number of Performance Measures	-3,311	-2.45	0.0161

As can be seen while the BMW model indicates statistical significance (i.e. the presence of correlation) the results for the SQM model are not statistically significant.

The lack of consistency in these findings betrays conclusive statement for the effect number of performance indicators used on performance. It is however plausible that the number of performance measures may have a material effect on performance but that it cannot be used as a dependable predictor for performance.

Order and Timing of Change

The notion of the sequence and timing of change was suggested by Keating et al., (1999), in improvement projects (paradox) where the authors assert that in order to facilitate sustainable improvements, a build up of capacity before the direct improvement initiative is necessary. For example, before increasing the production target, an increase of the production capacity needs to be carried out, otherwise the performance in the long run will be jeopardised by increased pressure due to overwork associated with improvement initiative.

In the SQM environment, the net hiring decision directly influences the production capacity, and in the BMW environment, advertising spending directly influences the pool of aware consumers, which in turn, can facilitate sales. Based on these characteristics, it is reasonable to assume that superior performance in cumulative profits will result from an increase in capacity as an initial change. That is, an initial increase in net hiring in SQM or an initial increase in advertising spending in BMW.

A t-test was used to test this hypothesis for each models. The outputs are summarised below:

SQM Microworld t-test

	Increased Net Hiring as Initial Change	Other Interventions as Initial Change
Mean	-2,194,773	-16,961,721
Variance	3.34776E+14	2.62142E+14
Observations	45	13
Pooled Variance	3.19212E+14	
Hypothesized Mean Difference	0	
df	56	
t Stat	2.625	
P(T<=t) one-tail	0.0056	

BMW Microworld t-test

	Increased Ad Spending as Initial Change	Other Interventions as Initial Change
Mean	25,584	10,689
Variance	335,809,360	149,862,272
Observations	41	12
Pooled Variance	295,703,125	
Hypothesized Mean Difference	0	
df	51	
t Stat	2.639	
P(T<=t) one-tail	0.0055	

The above tests indicate that in both models, capacity building decisions which showed an increase in net hiring (for SQM) or advertising spending (for BMW) were significantly better than other decision sequences. In other words, sequence and timing of decisions/actions do matter. This finding concurs with the findings in other studies such as the improvement paradox (Keating et al., 1999).

These findings, while statistically significant and compelling are being scrutinized for individual patterns and outlying behaviors. Further, real life cases are utilized to further validate the results.

Conclusions

This research addresses seven pitfalls in managerial decision making in complex systems. To our knowledge there is no cohesive body of literature that deals with these questions. Yet, practice, observations and anecdotal evidence point to endemic and habitual patterns of decision making with costly consequences for individuals and organizations. These decision pitfalls are summarized below.

- Managers tend to over-intervene. This is manifested in frequency and magnitude of their decisions. This 'over-intervention' behavior is caused and amplified by a lack of systemic thinking and misperception of dynamics of change. In day to day work, these behaviors are referred to variously as micro management, over reaction, tampering, etc.

- Managers and decision makers commonly ignore soft variables to the detriment of the staff and their organizations. Yet, soft indicators are powerful *lead* indicators of individual and organizational performance.

- Managers are often oblivious to "systems delays". Lack of awareness/attention to delay undermines performance and inhibits system stability.

- Organizations and managers tend to use too many performance measures (i.e., KPIs). Since the choice and number of KPIs impacts performance, excessive and inappropriate performance measures can lead to poor results and unintended consequences. In our experiments, this proposition was only statistically supported in the case of the BMW model. Our research continues to shed further light on this proposition.

- Managers generally focus on actions. Our research shows that *sequence* and *timing* of actions could be as important as the actions themselves.

- Managers (and organizations) often judge performance by short-term results. As performance often declines before it improves, expectation of short-term results is unrealistic and misleading and can lead to counteracting outcomes. This proposition is yet to be analyzed in light of the research experiments.

This research addresses seven propositions in relation to common pitfalls in managerial decision making. While the findings have shed new light on these propositions, the conclusions are far from final. Research is going on in this area.

References

Allison, G. T. (1971). *Essence of Decision*, Boston, Little, Brown and Company

BHUIYAN N., GERWIN D., THOMSON V., (2004) "SIMULATION OF THE NEW PRODUCT DEVELOPMENT PROCESS FOR PERFORMANCE IMPROVEMENT", *MANAGEMENT SCIENCE*, LINTHICUM: VOL. 50, ISS. 12; DEC P. 1690

Booth Sweeney, L, Sterman, J "Bathtub Dynamics: initial results of a systems thinking inventory", *System Dynamics Review*, Vol. 16, No. 4, winter 2000

Cavaleri, S. and J. D. Sterman (1997) "Towards evaluation of systems thinking interventions: a case study." *System Dynamics Review* 13(2): 171-186.

Cayer, C. (2001), "Choosing appropriate business improvement tools." Quality Congress, ASQ's ... *Annual Quality Congress Proceedings*: 282

Churchman, C. W. (1969) *The Systems Approach*, New York, Delacorte Press

Collins, J. (2001), *Good to Great*, Harper Business.

Collins, J. (2001), *Level 5 Leadership: The Triumph of Humility and Fierce Resolve*, Harvard Business Review, January 2001.

Deming, W.E. (1982) *Quality, Productivity, and Competitive Position*, MIT Press, P.116

Doyle, J. K. (1997). "The cognitive psychology of systems thinking" *System Dynamics Review* 13: 253-265

Easton, G. S. and S. L. Jarrell (1998) "The effects of total quality management on corporate performance: An empirical investigation." *The Journal of Business* 71(2): 253.

Elmuti, D. and Y. Kathawala (2000) "Business reengineering: Revolutionary management tool, or fading fad?" *Business Forum* 25(1/2): 29

Gilberto, P. A. (1995). "The road to business process improvement - Can you get there from here?" *Hospital Materiel Management* Quarterly 17(2): 12

Hall, R. I. (1984). "The Natural Logic of Management Policy Making: Its Implications for the Survival of an Organization." *Management Science* 30: 905-927.

Kanter, R. M. (2003). "Leadership and the psychology of turnarounds" *Harvard Business Review* 81(6): 58

Keating, E. K., R. Oliva, et al. (1999), "Overcoming the Improvement Paradox." *European Management Journal* 17(2): 120-134.

Maani, K, Cavana, R, *Systems Thinking and Modelling – Understanding Change and Complexity*, Prentice Hall, 2000

Maani, K, Maharaj, V "Links between Systems Thinking and Complex Decision Making", System *Dynamics Review*, Vol. 20. No. 1, 2004

Maani, K. and Li, A. "Dynamics of Managerial Intervention in Complex Systems" proceedings of the *22nd International Conference of System Dynamic Society*, Oxford, July 25 - 29, 2004

MacKinnon, A. J. and A. J. Wearing (1984) "Systems Analysis and Dynamic" *Acta Psychologica* 58: 159-172

Meadows, D. H. (1972), *The Limits to Growth: a report for the Club of Rome's Project on the Predicament of Mankind*. London, Potomac Associates

Meadows, D. H., D. L. Meadows, et al. (1992), *Beyond the Limits - Global Collapse or a Sustainable Future*, England, Earthscan

Morecroft, J. D. W. (1983). "System Dynamics: Portraying Bounded Rationality." *OMEGA* 11(2): 131-142.

Morecroft, J. D. W. (1985). "Rationality in the Analysis of Behavioral Simulation Models" *Management Science* 31(7): 900-916.

Richmond, B. (1994). "Systems Thinking/Systems Dynamics: Let's just get on with it." *System Dynamics Review* 10(2-3): 135-157.

Sekaran, U. (2000). *Research Methods for Business - A Skill Building Approach*, John Wiley & Sons, Inc, pp.166-167

Senge, P. M. (1990) *The Fifth Discipline - The Art & Practice of The Learning Organization*, Adelaide, Random House

Service Quality Microworld (SQM) manual (1994), *The Organizational Learning Centre*, Massachusetts Institute of Technology , 125 Cambridge Park Drive, Cambridge, Mass, 02140, USA

Simon, H. A. (1979). "Rational Decision Making in Business Organizations" *The American Economic Review* 69(4): 493

Simon, H. A. (1987). "Decision Making and Problem Solving" *INTERFACES* 17(5): 11-31.

Simon, H. A., Ed. (1957). *Models of Man - Social and Rational*, New York, John Wiley & Sons, Inc

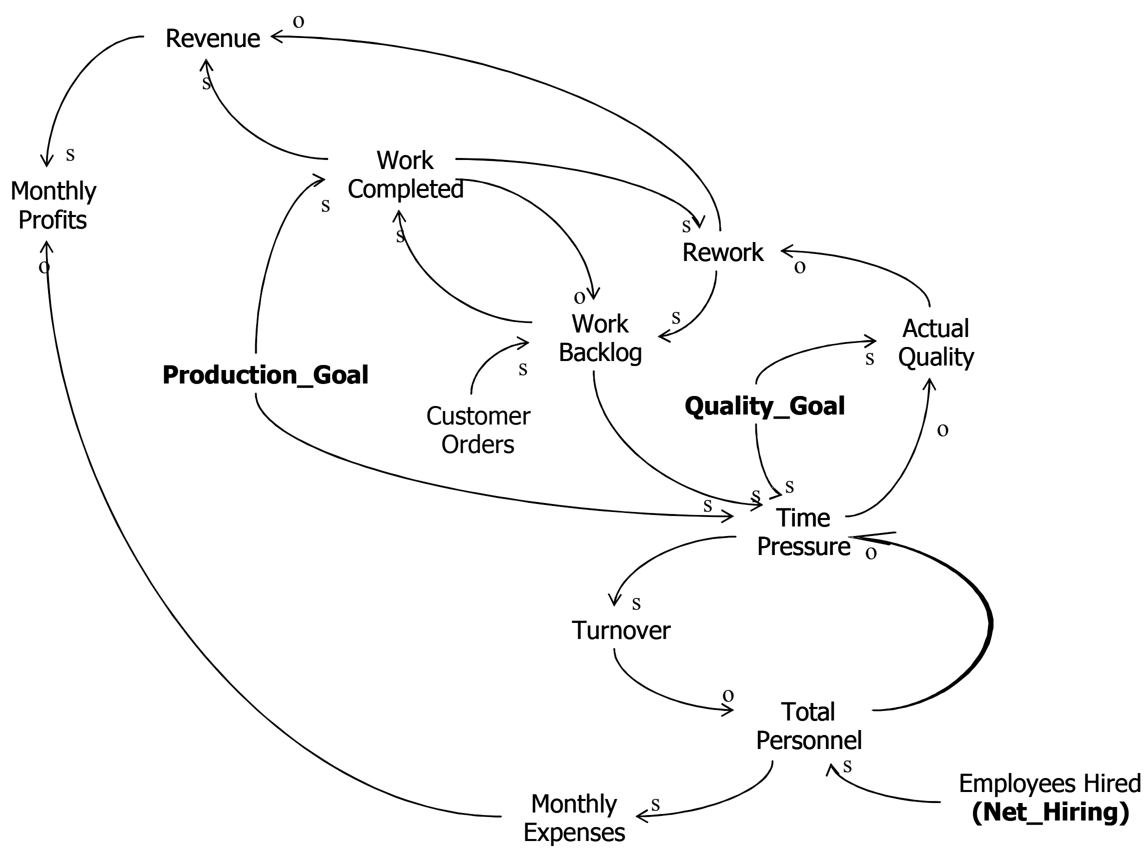
Sterman, J, (1989) "Modeling Managerial Behaviour: Misperceptions of Feedback in a Dynamic Decision Making Experiment", *Management Science*, Vol. 35, No. 3, March

Sterman, J. D., Ed. (1988). A Skeptic's Guide to Computer Models Managing a Nation: *The Microcomputer Software Catalog*, Boulder, CO, Westview Press

Thompson D., Waisanen L., Wolfe R., Merion R., et al. (2004) "Simulating the Allocation of Organs for Transplantation", *Health Care Management Science*, Bussum:. Vol. 7, Iss. 4; Nov, p. 331

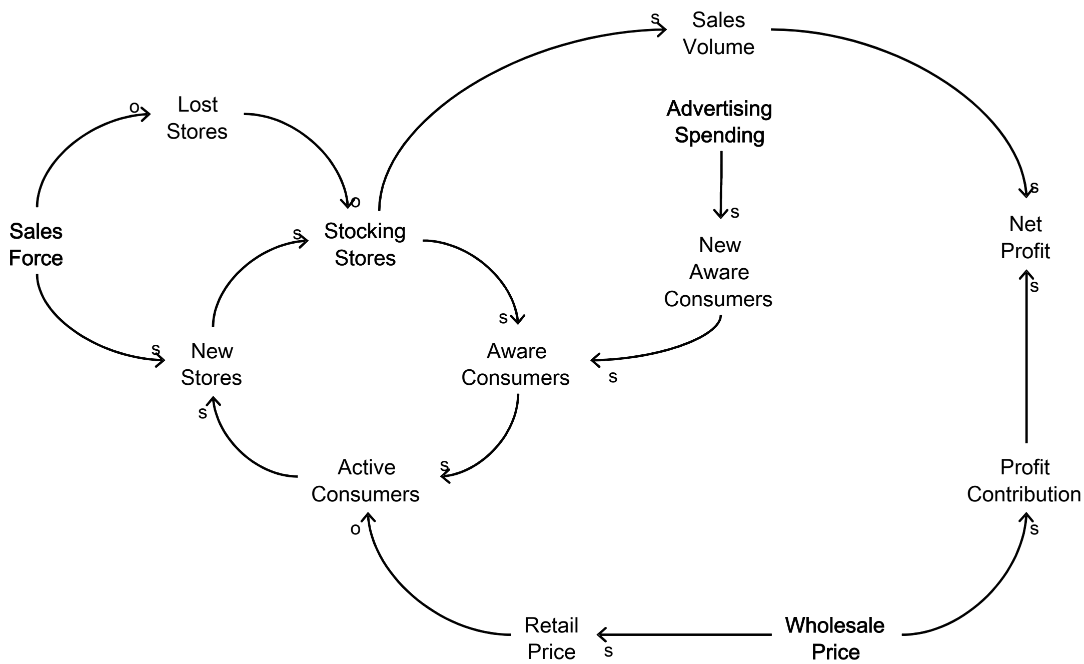
APPENDIX A

Partial Causal Loop Diagram involving the input variables for SQM



APPENDIX B

Partial Causal Loop for BMW Microworld



APPENDIX C

RESEARCH EXPERIMENTS

During the experiment sessions, the participants were required to achieve the stated goal of maximizing cumulative profits over the simulated period (5 years for Service Quality Microworld, and 12 years for Brand Management Microworld). This was to be accomplished by implementing various interventions with the three input variables (discussed under Decision Variables).

There were two separate exercises involved in each experiment session:

Exercise One: Participants were asked to achieve the goal by intervening with *any one* of the three variables over the simulated period. That is, the participants could only choose ONE VARIABLE to intervene in the model.

Exercise Two: Participants were asked to achieve the goal by using *any combination* of up to THREE VARIABLES over the simulated period. That is, the participants could intervene in the model using 1, 2 or 3 decision variables.

In both exercises, participants were asked to develop a strategy before starting the simulation using the Learning Cycle: "Conceptualize – Experiment – Reflect" (Maani et al., 2000) in developing their decisions and interventions.The subjects were asked to record their strategies on the worksheets provided, along with a detailed log of their decisions, actions and results. Further, they were asked to *predict* the likely behaviour pattern of their chosen KPIs over the course of the simulation. Subjects were monitored inconspicuously during the session.

Once the planning step was completed, the participants were asked to record a schedule of their interventions on a time line. That is, *when* and *how much* change in the chosen input variables they were planning to implement.

At the end of the simulation run, they were required to record the stated outcome of the experiment (i.e., the cumulative profits at the end of the simulation), and comment on the result as well as the process. This information was also recorded on the worksheet.

Endnotes

i A school of thought named in recognition of Carnegie-Mellon University, where much pioneering work on human decision making was done in the 1950s and 1960s by such well-known figures as Simon, Cyert, March and Williamson

ii Refer to the modes of decision making in the Bounded Rationality section

[iii] The simulated period of the results from the Brand Management Microworld has been restricted to 63 months, rather than the full 144 months. This is due to the fact that many of the participants have run out of the initial launch budget of £20 million soon after period 63 (which resulted in the termination of the simulation). This has created an inconsistency in the Frequency and Magnitude measures. Therefore, the simulation outputs from all participants' results after period 63 are truncated. The Frequency, Magnitude, and Cumulated Profits are taken at period 63. Note that this does not affect the #Variables variable.

[iv] The reported results are based on nil market feedback and zero noise options of the SQM model. Therefore there are no stochastic influences in these experiments. The SQM Microworld, however, allows random influences through the market feedback (a measure of input growth) and noise variable ("a measure of the amplitude of the random effect over customer orders") as well as seven growth scenarios (SQM Manual, 1994, p.7).

[v] The simulated period of the results from the Brand Management Microworld has been restricted to 63 months, rather than the full 144 months. This is due to the fact that many of the participants have run out of the initial launch budget of £20 million soon after period 63 (which resulted in the termination of the simulation). This has created an inconsistency in the Frequency and Magnitude measures. Therefore, the simulation outputs from all participants' results after period 63 are truncated. The Frequency, Magnitude, and Cumulated Profits are taken at period 63. Note that this does not affect the #Variables variable.