Proceedings of the 29th International Conference of the System Dynamics Society, Washington DC, USA, 2011

Analysis of the effects of different complexity factors on the complexity of a simulation game [1]

Onur Özgün and Yaman Barlas

Boğaziçi University

Department of Industrial Engineering, 34342 Bebek, Istanbul, Turkey

Tel: +90 212 3597343, Fax: +90 212 2651800

E-mail: onur.ozgun@boun.edu.tr, ybarlas@boun.edu.tr

Abstract

As a part of a larger ongoing research, this paper presents an experimental study in which the effects of delay, nonlinearity and feedback on game complexity are analyzed. A simple growth management game is designed and the game structure is modified to accommodate complexity factors. Factors are tested in independent Latin square experimental designs, with 4 or 8 levels. As performance measures, subjective difficulty ratings of players are also tested, in addition to the usual game score measure. Results show that individual complexity factors do not necessarily make the game more complex. Only the delay factor yielded a significantly worse game score with respect to the base game. Moreover, increasing delay time and feedback strength even causes improvement in game scores. The effect of repeated trials is only observed in players who played nonlinear games, which happen to be easier than other versions. Players' subjective complexity ratings and game scores show a correlation of -0.6*, which indicates that people tend to rate a game easy if they perform well. As further research, this experimental design will be applied to a stock management game. Next, the interactions of the complexity factors will be tested in a new set of experiments.*

Keywords: simulation games, dynamic complexity, experiments

1 Introduction

The field of system dynamics is built on the fact that many real-life systems are dynamic, closely-coupled, nonlinear, involving delays and feedback loops, which make them difficult to comprehend and analyze. These features make a system *dynamically complex*. Simulation games are interactive models of real world systems that are designed to support decision making and develop understanding about systems involving dynamic complexity. They are referred to by several names in the literature: microworlds, synthetic task environments, high fidelity simulations, interactive learning environments, virtual environments, virtual worlds, scaled worlds, management flight simulators (González et al., 2005). With the improvements in computer technology and increased availability of appropriate tools, the usage of simulation games in education and business has been increased (Faria and Wellington, 2004). In such an environment, it is important to design simulation games that result in effective learning. Although simulation games are useful teaching tools as

[1]Research supported by Boğaziçi University Research Grant no 09HA301D

they enable repetitive trials while compressing the time and space and provide immediate feedback, they still have inherent dynamic complexity due to factors such as accumulation, delay, feedback and nonlinearity in the models.

The effects of dynamic complexity factors in simulation games have been analyzed many times in the literature. For example, studies show that (Booth Sweeney and Sterman, 2000; Cronin et al., 2009) even educated people may be unable to infer the behavior of a system in the presence of a simple stock. Various studies have analyzed the relationship between delay and game performance (Broadbent and Aston, 1978; Diehl, 1989; Sterman, 1989; Paich and Sterman, 1993; Diehl and Sterman, 1995; Brehmer, 1995; Barlas and Özevin, 2004) and many report a negative effect of delay on performance. It is also seen from the research that, strength of feedback is effective on the game performance (Diehl, 1989; Kampmann, 1992; Paich and Sterman, 1993; Diehl and Sterman, 1995; Young et al., 1997; Langley et al., 1998). Likewise, it is known that nonlinearity can deteriorate the performance.

In spite of this wide range studies, most papers in the literature focus on one of the complexity factors with a few noteworthy exceptions. Paich and Sterman (1993) uses a boom-and-bust type of game to test the effects of feedback strength and delay on game performance. The paper tests two levels of each factor in a two-by-two design. Gary and Wood (2005) extends this work by including more cognition-related variables and by using mental model accuracy as another performance measure. Diehl and Sterman (1995) uses a stock management game to tests the effects on feedback strength and delay using more levels. A three-by-five Latin square design is used in this research.

This research aims testing the effects of multiple complexity factors in a consistent environment, at a larger scale. The proposed experimental design is larger in the following dimensions:

- Two types of simulation games are used: a growth management game and a stock management game
- Four dynamic complexity factors are tested: stock, delay, nonlinearity and feedback
- Factors are tested in isolation (first stage), as well as in interaction (second stage)
- Many levels (four or eight) of factors are tested in the first stage
- In addition to game scores, players' subjective difficulty assessments are used as alternative complexity measures

In the first stage of this proposed design, factors are tested in many levels, in isolation. In the second stage, the factors will be tested in interaction, in less levels. This paper briefly first summarizes this two-stage methodology, and then focuses on the first stage of it, where experiments are carried out and initial results obtained. In this set of experiments, a growth management game is designed and used. Different versions of the games are formed by adding delay, nonlinearity and feedback. Many levels are formed by changing their degree and an experiment is designed to assess their effects on game complexity.

Section 2 explains the method used in this research. It introduces the general framework of the overall design and presents the details of the task environment used in the experiments, the experimental design and the game protocol. Section 3 discusses how the benchmark behaviors are obtained. The findings of the experiment are shown in section 4. The paper concludes with discussion of the results and future work.

2 Method

2.1 Overview of Methodology

In this research, we use a two-stage design to test the effects of dynamic complexity factors on overall game complexity. The first stage focuses on the main effects of three dynamic complexity factors: delay, nonlinearity and feedback strength. Delay is analyzed in two dimensions (delay order and time) because both contribute to the complexity at different scales. Stock is not considered as a factor in the first stage because it has natural levels (no stock, 1 stock, 2 stocks) that will be all tested in interaction with other factors in the second stage. In the first stage interactions between factors are omitted, except the interaction between delay order and delay time.

In the second stage, the interactions between the complexity factors will be analyzed. In order to keep the number of experiments required at a reasonable number, in the second stage, only one level of delay, feedback and nonlinearity will be tested. The number of stocks will be added as a fourth factor. All two-, three-, and four-way interactions will be analyzed. Since our objective is to compare the effects of the complexity elements and their interactions, it is important to set the factors at comparable levels so that no factor dominates the others. So one purpose of the first stage is to determine the levels of factors that are neither trivially too weak nor trivially too strong in contributing to the overall game complexity. We will set these levels where factors just start to be significantly complex with respect to the base game, so that no single factor entirely dominates the other factors or is entirely dominated by others.

In both stages, two measures of complexity are used: game performance and players' subjective difficulty assessments. Game performance is measured by total profit relative to the benchmark performance. Players' subjective difficulty assessments are collected on a scale from 1 to 9, where 1 stands for an extremely easy game and 9 stands for an extremely hard game. In order make subjective measures as consistent as possible, all players are given two initial games (one easy and one difficult game) with predefined difficulty measures (1 for the easy game and 7 for the difficult game). Players are asked to assess the difficulty of each game just after playing it, with respect to two reference games.

2.2 The Task

We developed two simulation games called Growth Management Game and Stock Management Game. In this paper, only Growth Management Game will be described since the results obtained until now are related to this game. In the Growth Management Game, the players play the role of a product manager for a certain brand of lotion of a cosmetics company. The time unit of the model is *weeks*. The calculation step, dt, is taken to be 1 week, therefore the model runs in discrete time. The time horizon is 40 weeks. They are allowed to change product *price*, p, and *advertising* minutes aired on radio per week, a. *Sales*, S, is modeled to be directly proportional to the amount of *advertising* and inversely proportional to the *price*. The *sales* amount determines the *revenue*, R, and the *revenue* in turn determines the *weekly profit*, Π. The players' aim is to increase their *cumulative profit* as much as possible, in a sustainable way. They know the general structure of the model but they do not know the parameter values.

2.2.1 Base Game

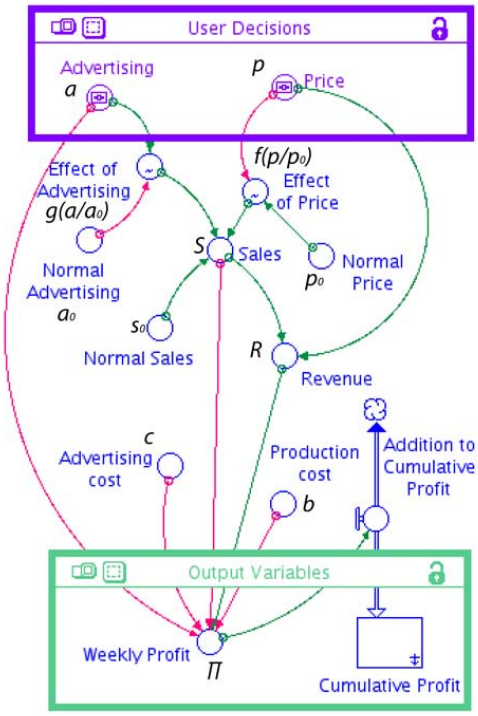
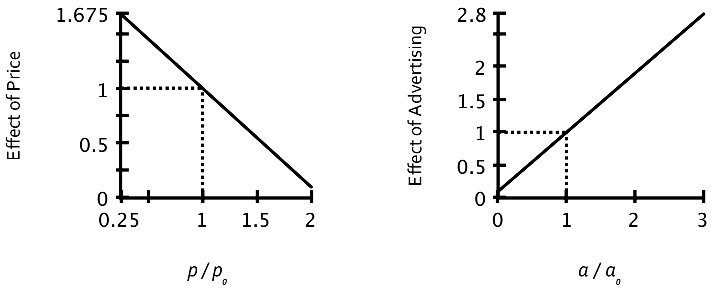


Figure 1: The base structure of the Growth Management Game

The base structure of the growth management game is as shown in Figure 1. When the *price* and *advertising* are set to their normal levels, p_0 and a_0, *sales* remains at a level called *normal sales*, s_0. p_0 and a_0 are also the starting conditions for the game, therefore all the variables stay at equilibrium unless *price* or *advertising* change. The *sales* amount for week t is:

$$S_t = s_0 \; f(p_t) \; g(a_t) \tag{1}$$

where $f(\cdot)$ and $g(\cdot)$ are effect functions of *price* and *advertising*, respectively. Their functional forms are shown in Figure 2.



(a) $f(\cdot)$: Effect of price (b) $g(\cdot)$: Effect of advertising

Figure 2: Linear functions showing the effects of *price* and *advertising* on *sales*

Weekly *revenue* and *profit* are computed as:

$$R_t = p_t \, S_t \tag{2}$$
$$\Pi_t = R_t - b \, S_t - c \, a_t \tag{3}$$

where c is the *cost of advertising* per minute per week and b is the *production cost* per item sold.

Since the base game will constitute a reference point to which the results of other games are compared against, it is kept as simple as possible. The game does not even include a stock. Thus, essentially it is not a dynamic simulation game but a simple trial-and-error task of figuring out the *price-advertising* combination yielding a high profit.

The game has a simple interface as shown in Figure 3. The input devices are two sliders for entering *price* and *advertising* decisions, Start and Advance buttons to start and advance the simulation and an Exit button to quit the game. The output devices are plot of *weekly profit* and *weekly benchmark profit*, the numerical displays of *weekly profit*, *cumulative profit*, *weekly benchmark profit* and *cumulative benchmark profit*.

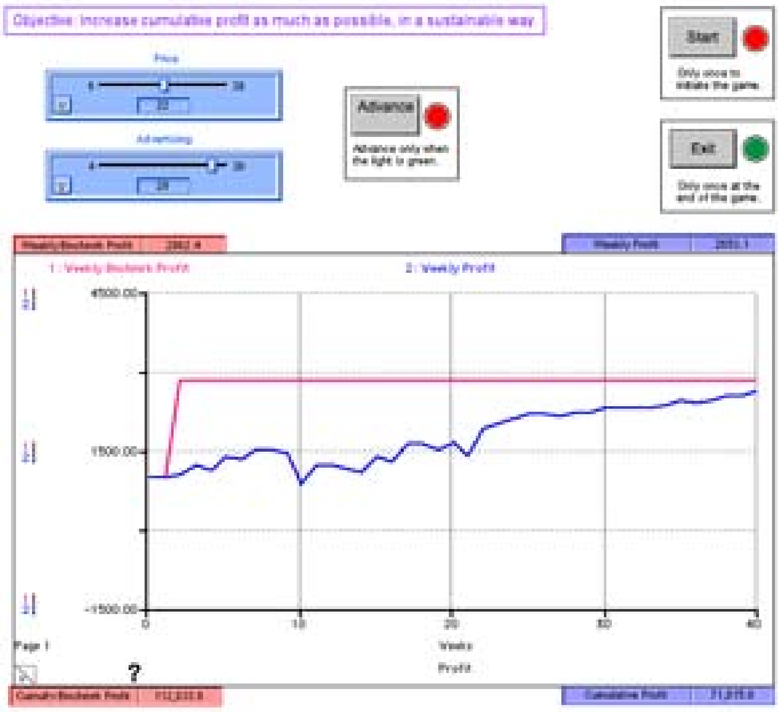
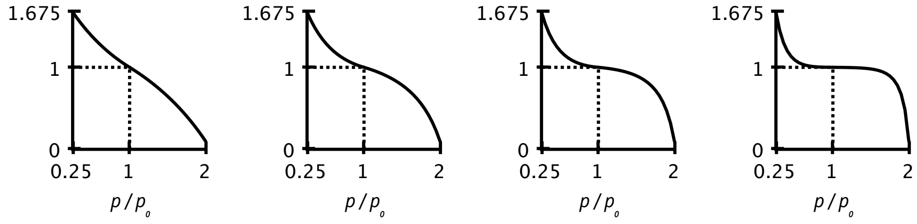


Figure 3: The user interface of the Growth Management Game

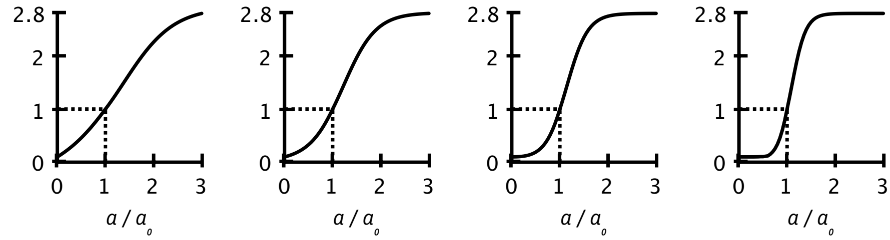
2.2.2 Games Involving Nonlinearity Factor

When nonlinearity is added to the base model, both the effect functions $f(\cdot)$ and $g(\cdot)$ used in Equation 1 are made nonlinear, simultaneously. There are four nonlinear formulations used. Figures 4 and 5 show the final forms of nonlinear functions used for *effect of price* and *effect of advertising*, respectively. Note that neither the base game nor the nonlinear games includes any kind of stock. Thus they do not have any memory, i.e. the players give 40 independent decisions when playing these game versions. Four levels of nonlinearity are be tested as shown in Table 1.



(a) N1: $f_1(\cdot)$ (b) N2: $f_2(\cdot)$ (c) N3: $f_3(\cdot)$ (d) N4: $f_4(\cdot)$

Figure 4: Nonlinear price effect functions



(a) N1: $g_1(\cdot)$ (b) N2: $g_2(\cdot)$ (c) N3: $g_3(\cdot)$ (d) N4: $g_4(\cdot)$

Figure 5: Nonlinear advertising effect functions

Table 1: The versions of growth management game involving nonlinearity

Game Version	Delay Order	Delay Time	Nonlinearity	Stock	Feedback
N1			Mild Nonlinearity (Figures 4(a) & 5(a))		
N2	–	–	Moderate Nonlinearity (Figures 4(b) & 5(b))	–	–
N3			High Nonlinearity (Figures 4(c) & 5(c))		
N4			Extreme Nonlinearity (Figures 4(d) & 5(d))		

2.2.3 Games Involving Delay Factor

When the dynamic complexity factor of delay is added to the base game, there are information delays between actions and their results on *sales*. More formally,

$$S_t = s_0 f_d(p_t/p_0) g_d(a_t/a_0) \tag{4}$$

where f_d and g_d are *delayed effect of price* and *delayed effect of advertising*, respectively. The rest of the equations are identical to those of the base game. Since delay structures include stocks, this version of the game has a dynamic component unlike the base game and the nonlinear games. The decisions of a player do not only effect the current period but also the following periods. Delay is analyzed in two components: order of delay and delay time. Order of delay has four levels while delay time has eight levels. There are $4 \times 8 = 32$ possible combinations of all levels of these two variables. The game versions involving delay are given in Table 2.

Table 2: The versions of growth management game involving delay

Game Version	Delay Order	Delay Time	Nonlinearity	Stock	Feedback
O1T2	First Order	2 wk			
O1T4	First Order	4 wk			
O1T6	First Order	6 wk	–	–	–
O1T7	First Order	7 wk			
⋮	⋮	⋮			
O1T11	First Order	11 wk			
O3T4	Third Order	4 wk			
O3T6	Third Order	6 wk			
O3T7	Third Order	7 wk	–	–	–
⋮	⋮	⋮			
O3T11	Third Order	11 wk			
O5T6	Fifth Order	6 wk			
O5T7	Fifth Order	7 wk	–	–	–
⋮	⋮	⋮			
O5T11	Fifth Order	11 wk			
ODT2	Discrete	2 wk			
ODT4	Discrete	4 wk			
ODT6	Discrete	6 wk	–	–	–
ODT7	Discrete	7 wk			
⋮	⋮	⋮			
ODT11	Discrete	11 wk			

2.2.4 Game Involving Stock

In the first stage, stock is not an experimental factor. It would be desirable to keep the number of stocks constant in all versions of the first stage. However, we know that feedback cannot exist without a stock and in order to have game versions with a feedback, we need to add a stock to the game. With the aim of keeping the other game versions (with delay and nonlinearity) as simple as possible, we chose not to add a stock to these versions. In order to be able to measure the isolated effect of having feedback, we need to have a game version with a stock but without any feedback. The game version involving only stock serves as a base version to the players playing games with feedback.

Figure 6 shows the structure of the game involving stock. The *customer* stock, C, increases by a *base recruitment rate*, β, which can be adjusted by *price* and *advertising* decisions. The *customer* stock drains with *attrition rate*. *Attrition rate* is the multiplication of the stock and the *attrition fraction*, δ. Mathematically, the computation of the value of the *customer* stock, C and *sales*, S at any week t are as follows:

$$C_{t+1} = C_t + \beta \; f(p_t/p_0) \; g(a_t/a_0) - \delta \; C_t \quad (5)$$

$$S_t = \sigma \; C_t \quad (6)$$

where σ is the *sales per customer* per week.

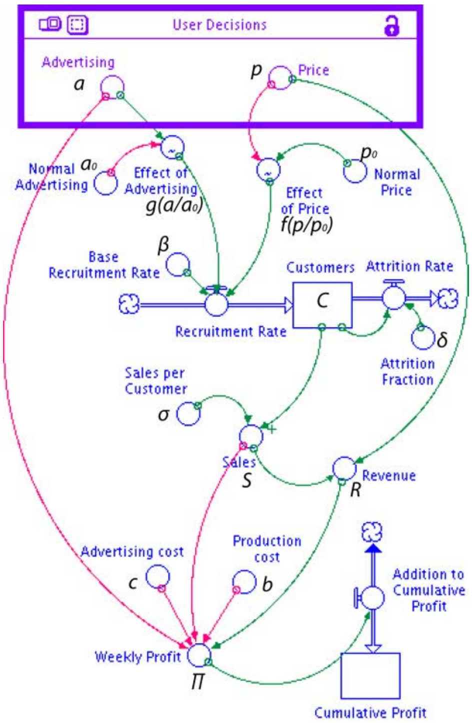


Figure 6: The structure of the growth management game with a stock

2.2.5 Games Involving Feedback Factor

Figure 7(a) shows the model including the feedback. It can be seen from the figure that this model actually includes two balancing feedback loops: one between the stock and the inflow and another between the stock and the outflow. The latter is a simple outflow control loop and included in the model in order to avoid a trivial game. In the experiments, we change the strength of the first feedback loop.

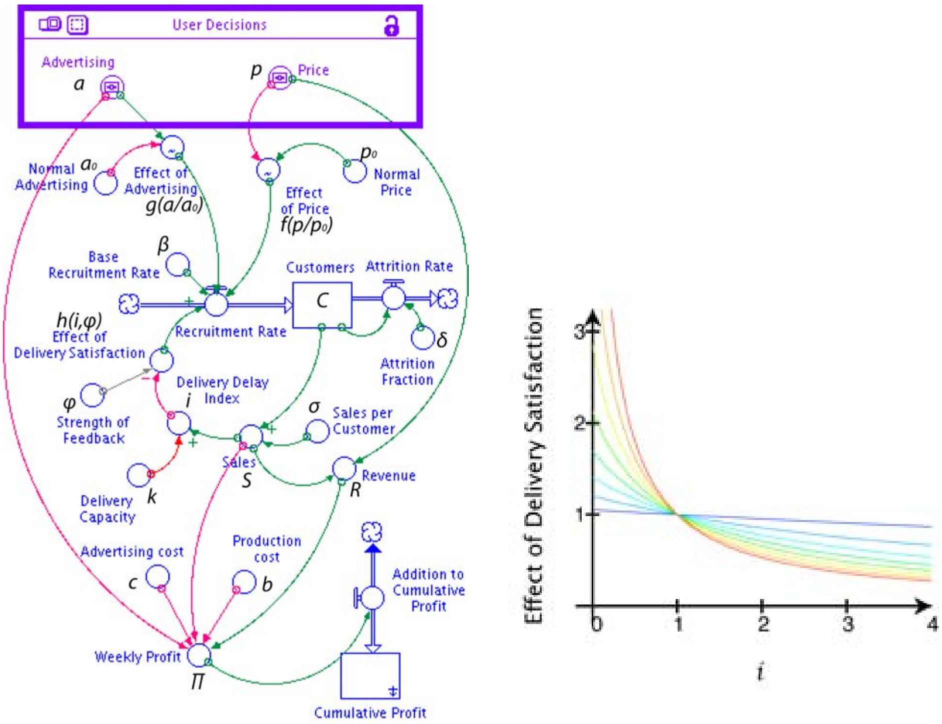
As *sales*, S, increases, the *sales* rate approach the *delivery capacity*, k. This would increase the *delivery delay index*, i, and decrease the value of the *effect of delivery satisfaction* function, $h(\cdot)$, which in turn will decrease the *customer recruitment rate*. Mathematically speaking,

$$i_t = \frac{S_t}{k} \tag{7}$$

$$h(i_t, \varphi) = \frac{1}{1 + \varphi(i_t - 1)} \tag{8}$$

$$C_{t+1} = C_t + \beta\ f(p_t/p_0)\ g(a_t/a_0)\ h(i_t, \varphi) - \delta\ C_t \tag{9}$$

where φ is the *strength of feedback* and $h(i, \varphi)$ is the *effect of delivery satisfaction*, which has the form shown in Figure 7(b) for various values of feedback strength, φ. When *strength of feedback* is small, the effect function, h, is flat. This makes the value of the function close to 1, regardless of the *delivery delay index*. When the feedback is strong, *delivery delay index* has a strong negative influence on the *recruitment rate*. In the first stage, eight levels of feedback strength will be tested as shown in Table 3.



(a) The model involving feedback

(b) The effect of delivery satisfaction, $h(i)$, for various values of feedback strength, φ

Figure 7: The version of growth management game involving feedback

Table 3: The versions of growth management game involving feedback

Game Version	Delay Order	Delay Time	Nonlinearity	Stock	Feedback Strength
F1				1 stock	0.05
F2				1 stock	0.17
F3				1 stock	0.29
F4	–	–	–	1 stock	0.41
F5				1 stock	0.53
F6				1 stock	0.65
F7				1 stock	0.77
F8				1 stock	0.89

2.3 Experimental Design

One of the aims of the first stage of experiments is to determine the minimum level of each complexity element that is significantly different than the base game and set each complexity level at these levels so that in the second stage the complexities of different elements are comparable. In line with that purpose, we took several levels for each complexity factors: 8 for delay time and feedback strength; 4 for delay order and nonlinearity. We use some modified version of Latin square design for the first-stage experiments as shown in Table 4. A total of 20 subjects is used. Delay order and delay time are treated jointly (players 1-8). The delay time has eight levels. With 8 subjects, a Latin square design is possible. On top of the Latin square design for delay time, the delay orders are embedded such that (1) each player plays every delay order, twice, (2) each delay order is played in every trial, at least once, (3) each delay order is combined with each delay time, twice. The only modification to this design was the replacement of O3T2, O5T2, O5T4 combinations with O3T4, O5T6, O5T6 because in a discrete model, when the delay time is less than the order of delay, the model yields numerically incorrect oscillations. Nonlinearity design (players 9-12) is composed of a repeated Latin square design (Trials 3-6 and 7-10). The repetition is done in order to equate the number of trials of each subject. The experimental design of feedback (players 13-20) is a pure Latin square design. While other players play the simplest base game in the beginning, the players testing feedback in the growth management game play the game with stock as the first (and the last) game. Then they play the difficult game as the second game, like everybody else. The aim of the difficult game is to help players in their difficulty assessments by providing another reference point (the other reference is the base game) with a predefined difficulty rating. The difficult game has a first order 8-week delay, mild nonlinearity, one stock and a feedback strength of 0.53. This configuration is determined after pilot experiments.

Table 4: The experimental design for the growth management game

Trial	Sbj1	Sbj2	Sbj 3	Sbj4	Sbj5	Sbj6	Sbj7	Sbj8	Sbj9	Sbj10	Sbj11	Sbj12
1	base	base	base	base	base	base	base	base	base	base	base	base
2	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.
3	ODT6	O5T7	O5T6	ODT8	O5T10	O5T11	O1T4	O5T9	N2	N1	N3	N4
4	ODT9	ODT4	O1T10	O5T6	O1T8	O3T4	ODT11	O1T7	N1	N4	N2	N3
5	O1T2	O3T8	O1T4	O5T9	O3T11	ODT10	ODT7	O5T6	N4	N3	N1	N2
6	O3T7	O3T4	ODT11	O3T10	ODT9	O3T8	O3T6	O3T4	N3	N2	N4	N1
7	O1T8	O1T6	O3T9	O1T11	O3T7	ODT4	O5T6	O3T10	N2	N1	N3	N4
8	O5T6	O1T9	O5T8	O1T7	O1T2	O1T6	O1T10	O1T11	N4	N3	N1	N2
9	O3T11	ODT10	O3T6	ODT2	O5T6	O5T7	O3T9	ODT8	N3	N2	N4	N1
10	O5T10	O5T11	ODT7	O3T4	ODT6	O1T9	O5T8	ODT2	N1	N4	N2	N3
11	base	base	base	base	base	base	base	base	base	base	base	base

Trial	Sbj13	Sbj14	Sbj15	Sbj16	Sbj17	Sbj18	Sbj19	Sbj20
1	S	S	S	S	S	S	S	S
2	diff.	diff.	diff.	diff.	diff.	diff.	diff.	diff.
3	F3	F4	F1	F5	F7	F8	F2	F6
4	F6	F2	F7	F3	F5	F1	F8	F4
5	F1	F5	F2	F6	F8	F7	F4	F3
6	F4	F1	F8	F7	F6	F5	F3	F2
7	F5	F3	F6	F8	F4	F2	F1	F7
8	F2	F6	F5	F4	F1	F3	F7	F8
9	F8	F7	F3	F1	F2	F4	F6	F5
10	F7	F8	F4	F2	F3	F6	F5	F1
11	S	S	S	S	S	S	S	S

Our pilot studies and several other studies have shown that as the subjects repeat playing a game, the familiarity increases and scores improve. This improvement can be between trials of identical games as well as different versions of the same games. We believe that this improvement is due to procedural learning and we do not want the effect of repeated trials confound with the effects of dynamic complexity factors. One approach for overcoming this effect could be playing each game in every possible order. In this way the effect of repeated trials could be ruled out. This method is applied in the first stage experimental design by adopting a Latin square design.

Another precaution taken for decreasing the procedural learning is applying minor modifications to the game interface. We slightly change the limits of the sliders between game versions. We carefully make these modifications according to the factor levels of delay time, nonlinearity and feedback so that these modifications do not bring any extra difficulty to the game. Since all players play all the versions of these factors (yet, in different orders), they all face the same interface conditions, in different orders.

2.4 Procedure

The subjects are recruited from undergraduate and graduate engineering students, who has not taken more than one introductory course in system dynamics. They are assigned to the game sets on a first-come-first-served basis. The experiments are carried out on Windows and Macintosh computers using STELLA software, which has essentially the same interface in both computer platforms. Upon arrival the subjects are given a written instruction (see Appendix A for the instructions) and they are told to read it carefully. In the instructions they are given an overview about the underlying model structure, the game objective and instructions about the subjective difficulty assessment. During the experiments, they are asked to rate the difficulty of each game they play on a scale 1 to 9, where 1 corresponds to an extremely easy and 9 corresponds to an extremely hard game as shown below:

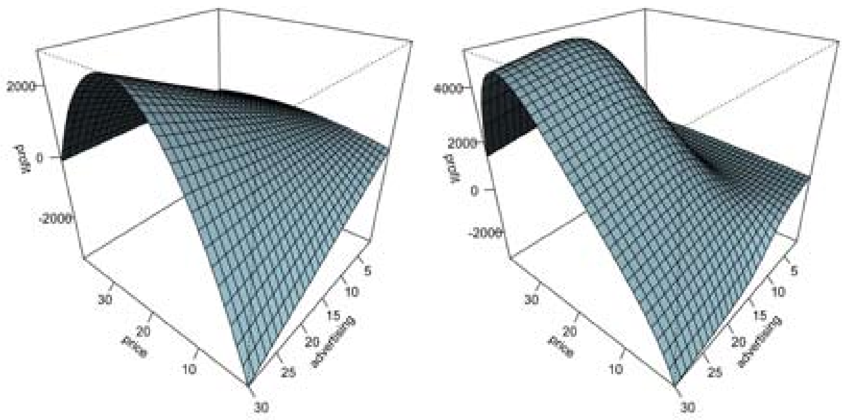
```
EXTREMELY                                   EXTREMELY
  EASY        EASY     AVERAGE     HARD        HARD
   1     2     3     4     5     6     7     8     9
```

They are told that the first (base) game has an pre-assigned difficulty of 1 and the second game has a difficulty of 7. Both the base game and the game with stock has pre-assigned difficulties of 1 (although most probably they are not at the same complexity level). This does not pose a serious problem in the analysis since we do not compare the results of different subjects. After the instructions, they are made to play a trial game for getting familiar with the game interface, the software and the procedure. The trial game has different parameters than other games and is designed not to give hint about the consecutive games. Then they play the base game followed by the difficult game. After that, they play eight games involving one of the complexity factors. At the end of each game, they record their score and rate the difficulty of game they played. In the end, they play the base game once more. After completing all games, they have the chance to revise their difficulty ratings.

The players are given monetary rewards that change according to the performance. The payoff function is *cumulative profit*. Note that the the payoff function is slightly different than the game objective. The subjects are told to increase *cumulative profit* in a sustainable way but they receive payoff based on a measure that does not explicitly punish unsustainability. But the payoff function is not revealed to the subjects. Since the game structures differ among games, it would not be fair to compare players playing games testing different factors. But since all players in a factor group play same games with identical game conditions, but in random order, comparing the overall scores of 8 games would be fair. Therefore, the players are evaluated in groups of 4. The players are paid fixed amounts according to their ranks in a 4-player group.

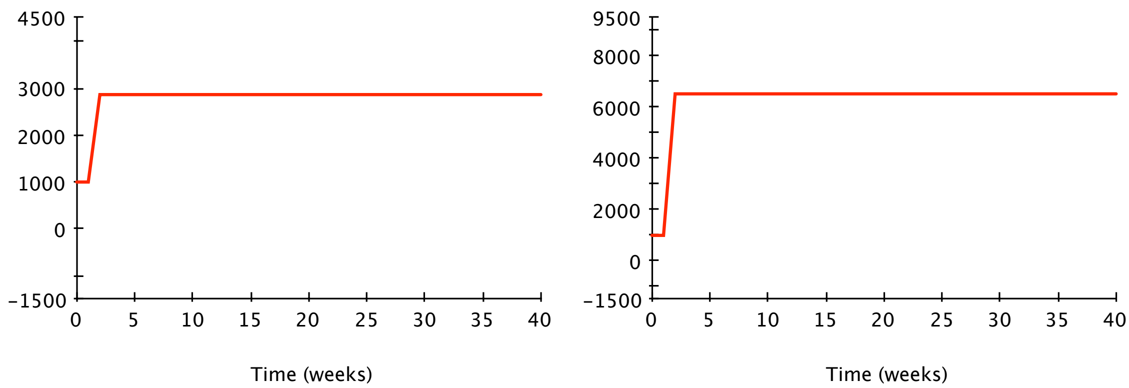
3 Benchmark Behaviors

In line with the players' objective, the benchmark behavior is defined such that long-run profit is maximized. For the base game, the weekly profits corresponding to all *price-advertising* combinations are calculated and the combination yielding the maximum *weekly profit* is applied starting from the second week. For the nonlinear game, a similar method is used. Existence and degree of nonlinearity only changes the *weekly profit* corresponding to every *price-advertising* combination. We can still find the optimum combination using complete enumeration. Figure 8 shows the resulting *weekly profits* for all input values of the base game and for a nonlinear game. Introducing nonlinearity changes the size and location of the region where the player can obtain high profits. Since every week is independent from each other, the strategy of providing inputs yielding maximum *weekly profit* maximizes the *cumulative profit* as well. You can see the benchmark behaviors for the base game and for a game involving nonlinearity in Figure 9. In the base game and the nonlinear games, it is impossible to exceed the weekly benchmark profit and hence, the cumulative benchmark profit. So, benchmark provides a strict upper-limit for these cases.



(a) Base game (b) N3

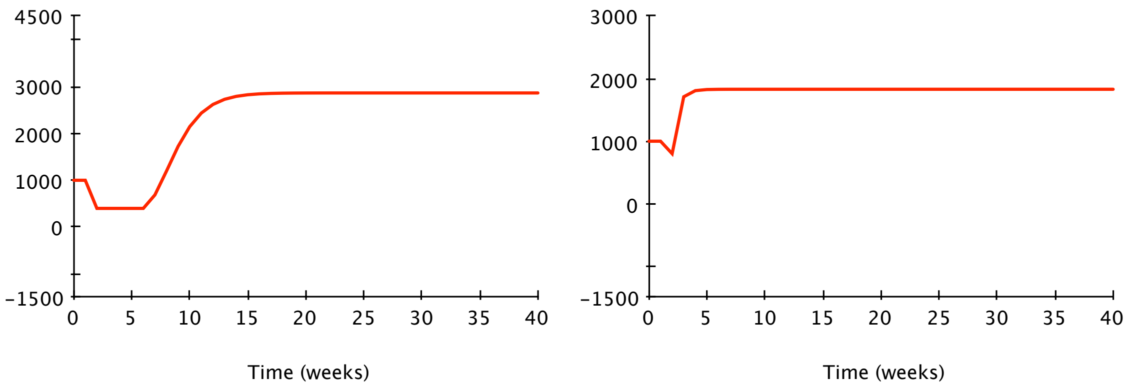
Figure 8: The surface plots showing the resulting *weekly profit* for all combinations of *price* and *advertising*



(a) Base game (b) N4

Figure 9: Benchmark behaviors for the base and nonlinear versions of growth management game

Inclusion of delay brings a dynamic component to the game. Every decision has two effects: one immediate effect and one long-term effect in the opposite direction. In spite of this, the long-term equilibrium points of the profit is not affected from the existence of delay. Thus, the benchmark behavior is found by setting the price and advertising to a combination that maximizes the profit *in the long-term*, starting from second week (See Figure 10(a) for the benchmark behavior of a game with delay). For the game involving stock and feedback, we have to determine the equilibrium level of *customer* stock for every combination of *price* and *advertising*. The corresponding *weekly profit* is easily calculated. Among all the possible *price-advertising* combinations, the one yielding the highest equilibrium profit is selected. Like in all other versions, the benchmark behavior is obtained by setting the *price* and *advertising* to their optimum levels starting from second week (See Figure 10(b) for the benchmark behavior of a game with feedback). The benchmark behaviors of game versions involving delay and feedback maximize the long-term sustainable profit. The *cumulative profit* within the time horizon of the game can be higher than *cumulative benchmark profit*.



(a) O5T8

(b) F5

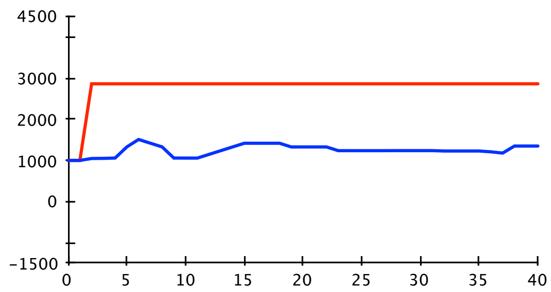
Figure 10: Benchmark behaviors for the versions with delay and feedback of growth management game

4 Results

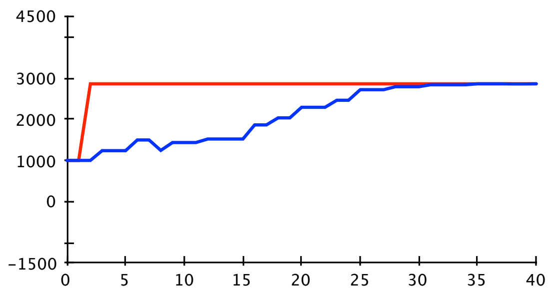
4.1 Qualitative analysis

Before going into a statistical analysis, a qualitative analysis of results is useful. Appendix B presents the behaviors of *weekly profits* for all trials of all players. First fact that is observed is the variation of behaviors between players. Figure 11 shows three example behaviors in the simplest base game. Player 10 was very conservative in moving the sliders controlling the *price* and *advertising* and he/she could not manage to bring the profit to a high level even after 40 trials. Player 8 represents an almost typical behavior: the *weekly profit* steadily increases until the benchmark profit is reached. Player 4 shows another extreme case. This player was simply lucky to discover a combination yielding a profit almost equal to the benchmark in the second week (they are not allowed to change inputs in the first week). At week 14, he/she tries to beat the benchmark and realizes that it is not possible.

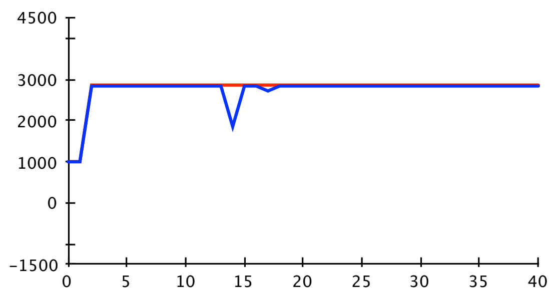
This big variance in the first game is only attributable to factors related to the players themselves.



(a) Player 10: base



(b) Player 8: base



(c) Player 4: base

Figure 11: Three example behaviors in the base game. Blue line: *weekly profit*, Red line: *weekly benchmark profit*

The player characteristics also seem to have influence on the experience effect. While some players improve their performance over trials regardless of the game versions played (e.g. Players 4, 12, 20), many of them does not show a clearly identifiable progress.

Recall that the game objective is to increase profit in a sustainable way. The sustainability requirement is added to discourage any behavior that creates temporary high profit, toward the end of the game. Normally almost any action that creates sudden jumps in the profit results in a decrease in the long term. However toward the end of the game, the long-term effect of temporary jumps will not be observed because of the termination effect. The sustainability requirement is further explained in the instructions. Players are told to "increase the weekly profit toward a maximum level, while avoiding excessive up and down movements toward the end of the game". Most players obeyed the warning. There are a few exceptions like the one shown in Figure 12.

Another interesting observation is that some players (e.g. Players 5 and 16) seem to perform worse in their last game, although it is exactly identical to the first (and presumably easiest, base) game. A statistical test on the significance of this observation is explained in the next subsection.

A visual analysis of the results also reveals a distinctive difference between performances of Players 9-12 and the remaining ones. Based on this observation, nonlinearity seems not to bring much complexity to this game. More will be said on this after statistical analysis of the results.

Figure 12: Example of an unsustainable growth. Player 20, Trial 3: N4

4.2 Quantitative Analysis

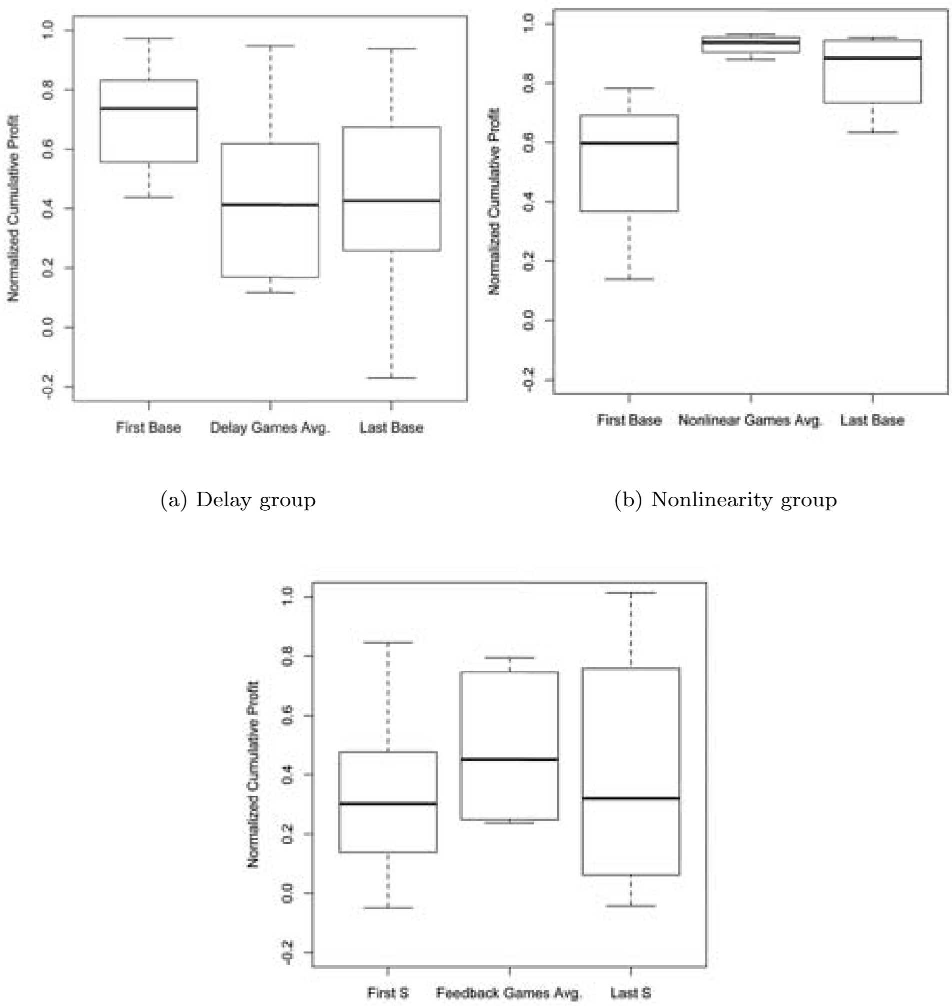
The different game versions have different orders of scales. So in order to make a comparison between results of different versions, they should be brought to a comparable scale. The performance measure used to quantify the experiment results is called *normalized cumulative profit*. It is defined as follows:

$$\frac{\textit{Cumulative profit} - \textit{Cumulative profit of do-nothing strategy}}{\textit{Cumulative benchmark profit} - \textit{Cumulative profit of do-nothing strategy}} \tag{10}$$

Normalized cumulative profit removes the scale effect between games and results in consistent scores that make sense. For example if the player does exactly same as the benchmark, she get a score of 1, in all games. If she does nothing, she gets zero, again in all games. Theoretically, it is possible to get scores as high as 4.3 and as low as -2.3. But the typical scores rarely exceed 1 or become negative.

There are three groups of players: delay group (players 1-8), nonlinearity group (players 9-12) and feedback group (players 13-20). Analysis is carried out based on these three groups.

Figure 13 shows box plots comparing the scores of two base games and the average of trials in which complexity factors are present. One-sided paired t-tests at significance level of 0.05 show that there is significant difference between scores of the first base game of the delay group and the games involving delay. The game scores worsen when delay is introduced, in spite of a possible experience effect. Interestingly, games scores do not improve in the last base game, although the last game is a simpler game *and* identical to the first game in which higher scores are obtained. Moreover, the *price-advertising* combination yielding maximum profit is identical in base games and all delay games. That is, if the subjects applied the same winning strategy in both base games and the delay games, they would perform well in both. However, players may be misconditioned in the intervening games by long-term effects of their actions, and may have been intimidated in the final base game although they are told that it is the base game. As shown in Figure 13(b), there is also a significant difference between the scores of the first base game and the average scores of the games involving nonlinearity. The difference is positive, meaning that the players performed better in the nonlinear games! On the other hand, there is no significant difference between the last base game and the nonlinear games. These two observations bring the possibility of an experience effect in this group of games. Another point that should be kept in mind in analyzing these results is the game structures. As the nonlinearity increases, the region yielding a high profit gets larger. Also, the distance between the starting point and the point of maximum profit gets closer. In theory, these two properties could indeed make the game easier. On the other hand, unlike the delay games, the input combination resulting in the maximum profit changes as the nonlinearity changes. This is a property that is expected to make the game difficult. As Figure 13(c) demonstrates, there is no significant difference between the scores of two base games and the average scores of the games involving feedback. Note that, the base games of this group involve a stock by definition.



(a) Delay group

(b) Nonlinearity group

(c) Feedback group

Figure 13: Comparison of first base game, average of trials 3-10 (games involving complexity elements) and the last base game scores.

4.2.1 Delay Group

Note the the Latin square experimental design allow us to extract the effects of players, trials and experimental complexity factor for the trials between 3 and 10 of players and providing us enough replications (eight replications for all factors except for the unbalanced delay order situation). Figure 14 shows boxplots of *normalized cumulative profits* for the levels of player, trial, delay order and delay time factors. Table 5 shows the analysis of variance table of the same data group. These results indicate a strong variation between players' scores. The variance caused by the differences between players has a dominating effect on the results. According to the statistical analysis, there is no effect of repeated trials. Thus, the effect of experience between delay games can be ignored. On the other hand, the results suggest a significant delay time effect at the significance level of 0.05. Interestingly, the *normalized cumulative profits* increase as delay time increases! A more careful look indicate that, this increasing effect is due to large scores in 9 week or longer delays. Indeed, if we only consider delay times at the 2-8 wk range, the p-value of delay time factor rises to 0.09717, which is not significant at 0.05 significance level. This is somewhat counterintuitive but may be explained by the very nature of the delay in this game. In the existence of delay, actions have two consequences: one immediate consequence and one delayed consequence. Thus, in the presence of long delays, players can enjoy the short-term benefits of this effect. In addition to this, there is a similar effect on the benchmark behaviors. When there is longer delay, it takes more time for the benchmark to reach the equilibrium. Thus, the *cumulative benchmark profit* becomes smaller when the delay time gets larger. Referring to Equation 10, we see that larger delays can result in large scores by the combined effects of increased numerator and decreased denominator.

Table 5: ANOVA table for *normalized cumulative profits* in the delay group

```
              Df  Sum Sq Mean Sq F value    Pr(>F)
Player         8 16.9549 2.11937 39.7879 < 2.2e-16 ***
Trial          1  0.0082 0.00820  0.1539  0.696396
Delay Order    1  0.0364 0.03636  0.6826  0.412404
Delay Time     1  0.4936 0.49363  9.2672  0.003628 **
Residuals     53  2.8231 0.05327
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is a very interesting property of the growth management game. A deeper analysis is conducted to understand how and when this effect is influential on the game results. We needed to know, whether it is really the temporary profit gains that created higher scores in the longer-delay games. To answer that question, we defined another variable: *weekly equilibrium profit*, and its cumulative version, *cumulative equilibrium profit*. *Weekly equilibrium profit* is the *weekly profit* that would be realized as a result of the player's action, if there were no delay. In other words, it is the long-term, equilibrium level of profit that will be reached if no further action is taken. We computed the *cumulative equilibrium profits* for all games of all players and normalized it with respect to the benchmark as we did in Equation 10. The results show that there is no delay time effect when we use *normalized cumulative equilibrium profit* as a performance measure. This supports our hypothesis that the temporary profit gains are responsible for higher scores in the existence of long delays.

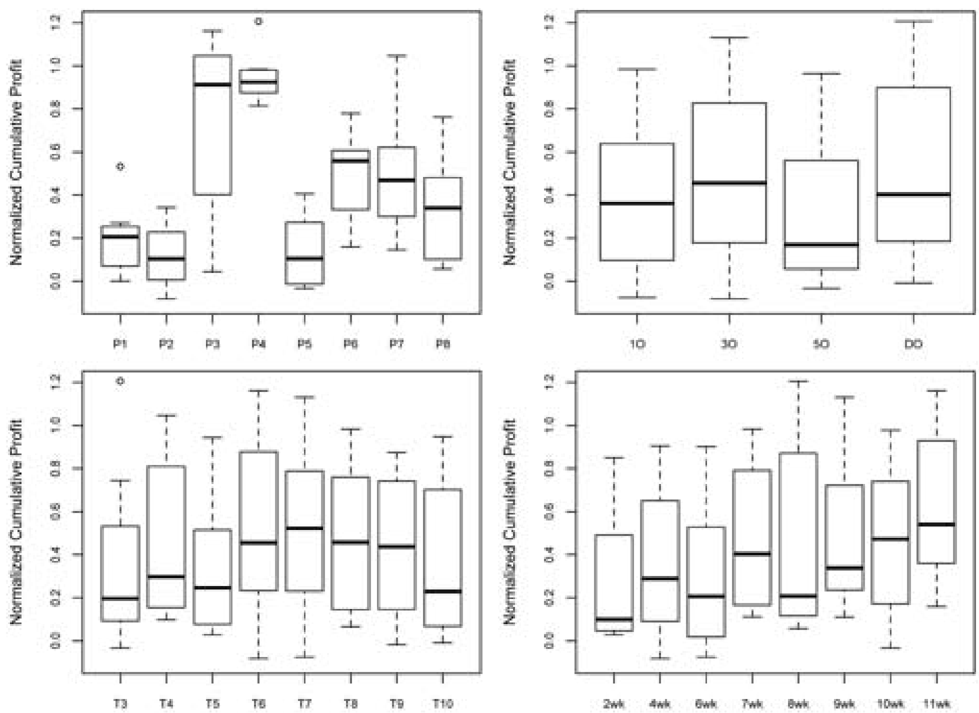


Figure 14: *Normalized cumulative profits* for different levels of: players (top left), trials (bottom left), delay order (top right) and delay time (bottom right), in the delay group

Figure 15 and Table 6 show the analysis of variance for the subjective difficulty assessments of players in the delay group. Similar to the case in game scores, there is a strong player effect. The remaining factors are not significant although there is a weak trial effect. Delay time, which is a significant factor in terms of *normalized cumulative profit*, is not significant in terms of subjective difficulty.

Table 6: ANOVA table for subjective difficulty assessments of players in the delay group

```
              Df  Sum Sq Mean Sq F value  Pr(>F)
Player         8 2490.62 311.328 99.5201 < 2e-16 ***
Trial          1   11.26  11.257  3.5983 0.06329 .
Delay Order    1    0.22   0.224  0.0716 0.79004
Delay Time     1    1.09   1.095  0.3499 0.55667
Residuals     53  165.80   3.128
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

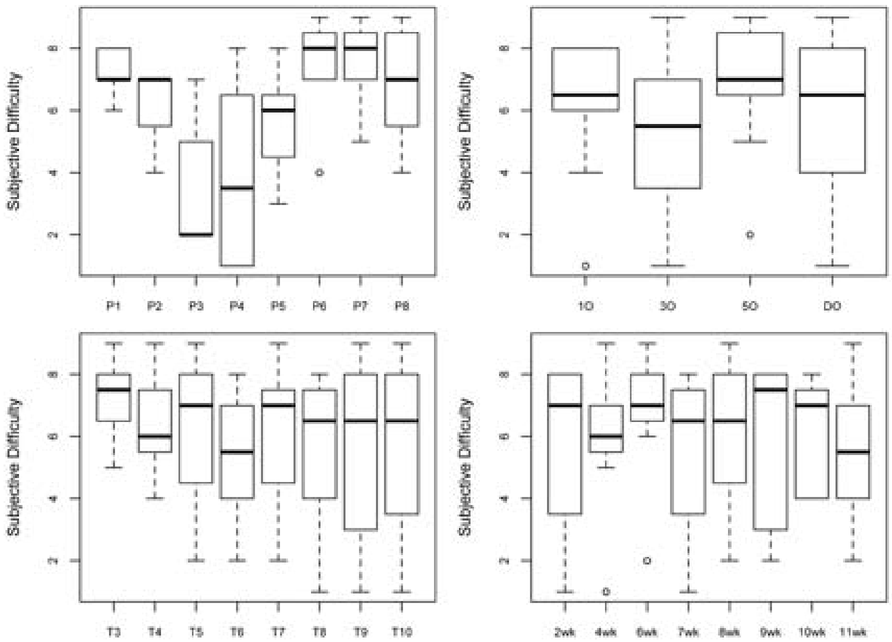


Figure 15: Subjective difficulty assessments of players in the delay group for different levels of: players (top left), trials (bottom left), delay order (top right) and delay time (bottom right)

4.2.2 Nonlinearity Group

Figure 16 and Table 7 show the analysis outputs for *normalized cumulative profits* of nonlinearity group. Still there is a strong player effect. There is no significant difference between levels of nonlinearity. However, the effect of trial is significant only for this group. As players gain experience, their scores improve. This learning is despite changing location of input combination yielding maximum profit. This result may be explained by the lack of dynamic component in the nonlinear games. The games are so easy that players can learn quickly and transfer their experience between trials.

Table 7: ANOVA table for *normalized cumulative profits* in the nonlinearity group

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Player	4	27.6750	6.9187	3620.7052	< 2.2e-16	***
Trial	1	0.0459	0.0459	24.0291	4.353e-05	***
Nonlinearity	1	0.0003	0.0003	0.1712	0.6824	
Residuals	26	0.0497	0.0019			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The subjective difficulty assessments of players in nonlinearity group is summarized in Figure 17. In this case, none of the factors are significant enough to explain the variability in the data so no ANOVA table is given.

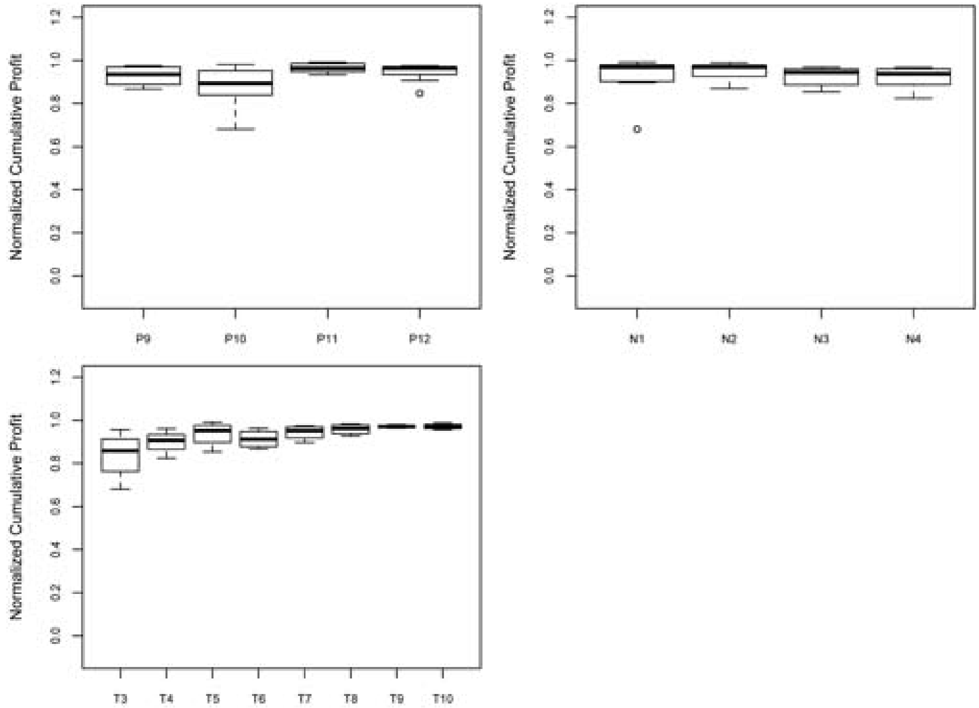


Figure 16: *Normalized cumulative profits* for different levels of: players (top left), trials (bottom left) and nonlinearity (top right), in the nonlinearity group

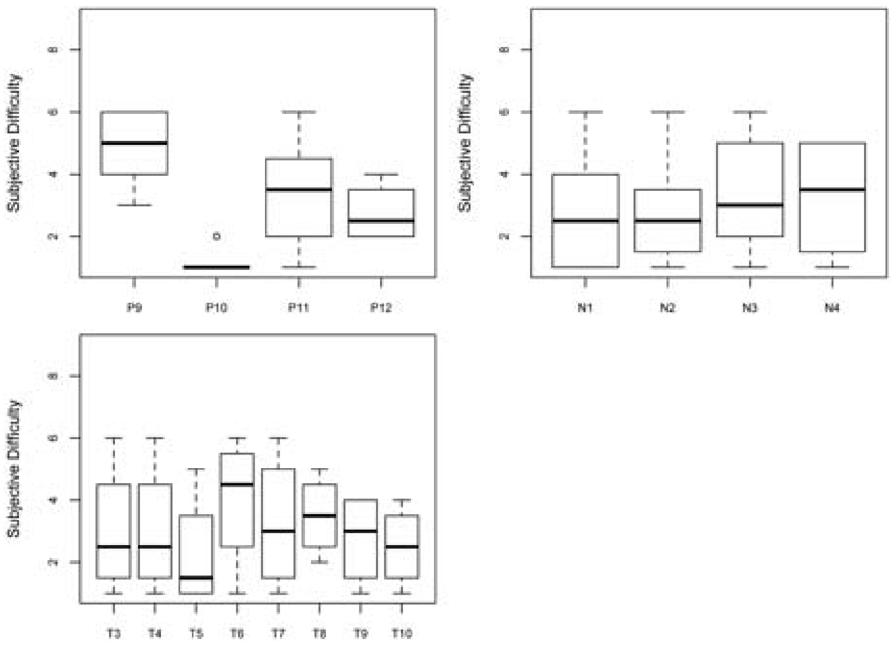


Figure 17: Subjective difficulty assessments of players in the nonlinearity group for different levels of: players (top left), trials (bottom left) and nonlinearity (top right)

4.2.3 Feedback Group

Figure 18 and Table 8 summarize the analysis results of *normalized cumulative profits* of the feedback group. The player and feedback effects are capable of explaining the variance in the data. Increasing feedback causes the game scores to improve. This observation is surprising as well, because the feedback loop is a negative feedback loop and should suppress growth of the profit. We believe that the change in the resulting *weekly profit* response surface as a result of changing feedback strength, is responsible for the increasing scores.

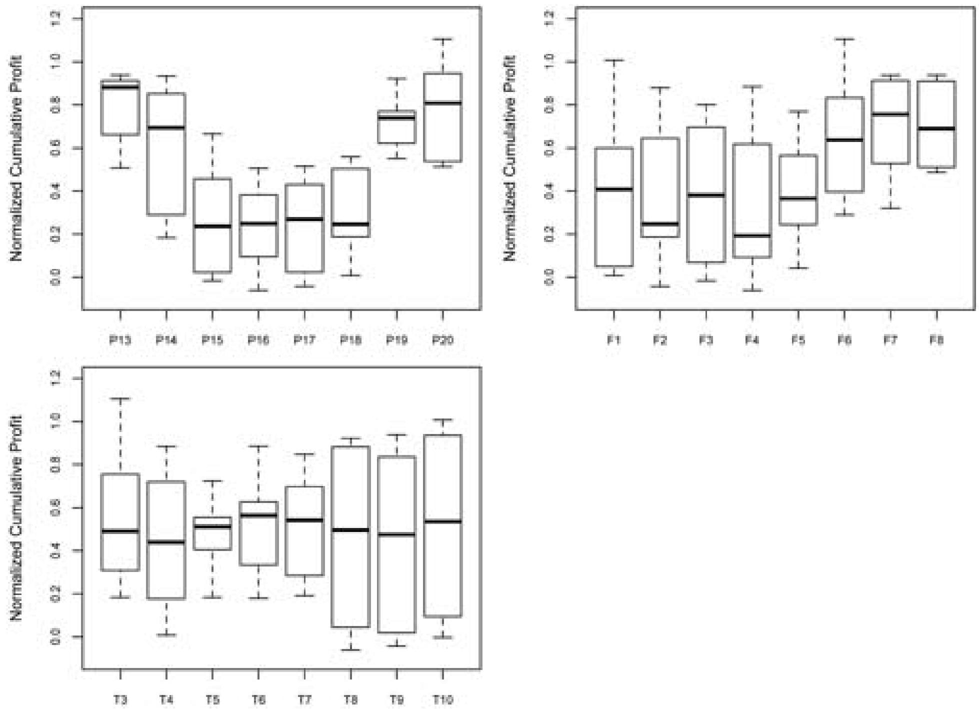


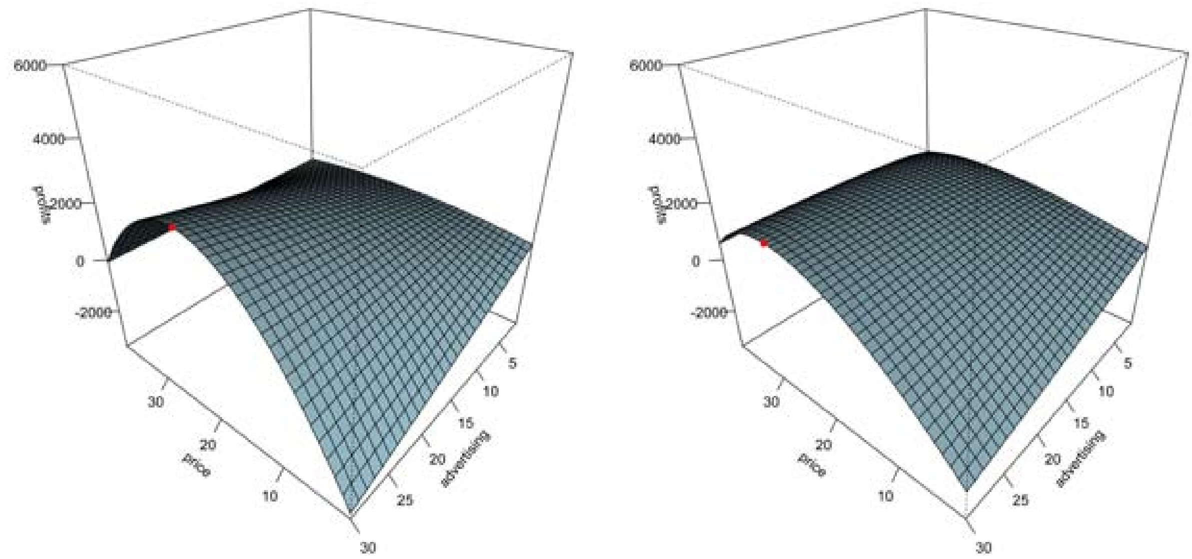
Figure 18: *Normalized cumulative profits* for different levels of: players (top left), trials (bottom left) and feedback (top right), in the feedback group

Table 8: ANOVA table for *normalized cumulative profits* in the feedback group

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Player	8	19.0352	2.37940	79.3824	< 2.2e-16	***
Trial	1	0.0040	0.00401	0.1338	0.716	
Feedback	1	1.0506	1.05060	35.0504	2.288e-07	***
Residuals	54	1.6186	0.02997			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As Figure 19 illustrates, as the feedback strength increases, the maximum obtainable *weekly profit* is reduced. This suppresses the *weekly profit* surface and the surface becomes flatter. This makes the relation between the action and reaction more linear. Since the duration of delay created as a result of the added stock is not affected by the strength of the feedback, increasing feedback strength makes it easier for the player to discover a winning strategy.



(a) F1 (b) F8

Figure 19: The *weekly profit* surfaces for two levels of feedback strength.

Figure 20 and Table 9 show the subjective difficulty assessments of the players in the feedback group. Their assessments are parallel to their game scores; higher scores made people to assign a lower difficulty rating. The results are also in agreement with our interpretation. The player perceived the games with stronger feedback to be easier.

Table 9: ANOVA table for subjective difficulty assessments of players in the feedback group

```
           Df  Sum Sq Mean Sq F value  Pr(>F)
Player      8 1747.25 218.406 97.9064 < 2e-16 ***
Trial       1    0.86   0.860  0.3856 0.53725
Feedback    1   15.43  15.429  6.9163 0.01110 *
Residuals  54  120.46   2.231
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

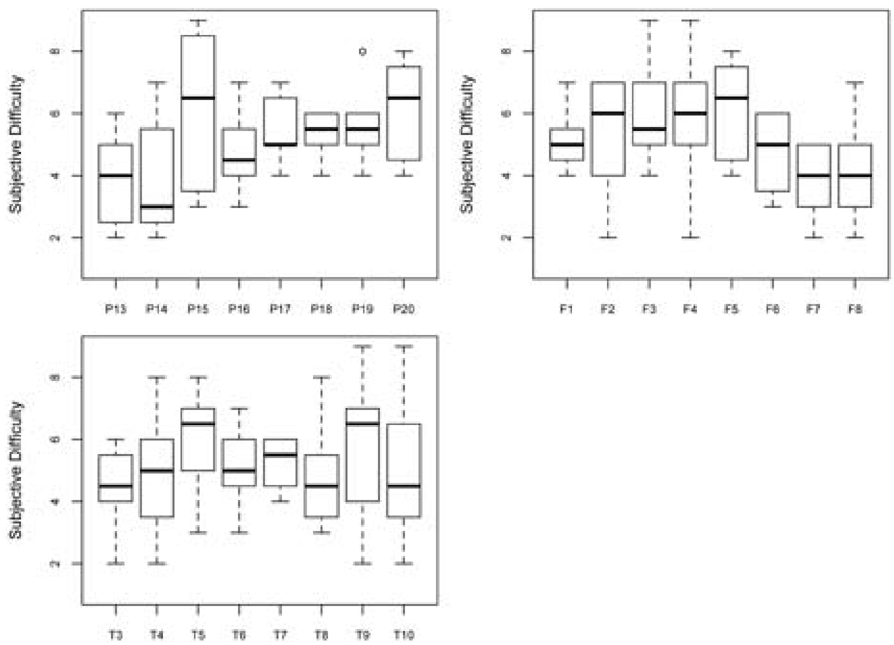


Figure 20: Subjective difficulty assessments of players in the feedback group for different levels of: players (top left), trials (bottom left) and feedback (top right)

5 Discussion

The statistical analysis of the outputs of this set of experiments indicated surprising results.

In terms of dynamic complexity factors, this study shows results that would not be guessed beforehand. As a complexity factor, only delay caused lower game scores, on the average. But although delay brings difficulty in the game with respect to the first base game, within the delay group, as delay time increases the scores tend to improve. The other dynamic complexity factors do not cause worsening in scores either. While increasing feedback strength actually improves the scores, increasing nonlinearity does not have any significant effect. On the other hand, the average scores of the nonlinear games is higher than the first base games. While learning could be responsible for the improvement, the structure of the game in nonlinear versions makes it easier to get high scores.

Before the experiments are carried out, an improvement was expected as people get experienced in the game. However, this experience is only observed in nonlinearity group. Both the trial effect within nonlinear games and the difference between first and last base games confirm improvement (in nonlinear experiments). Surprisingly, players of the delay group performed significantly worse in their last base game with respect to their first base game although they are identical games and the strategy to have the highest sustainable profit is the same with all the games in-between. Thus, it could be a reasonable conjecture to think that different dynamic complexity elements can bring different psychological side effects.

One important originality of this study was to ask players' subjective difficulty ratings. It turned out to be that *normalized cumulative profits* and subjective difficulty assessments of the players are usually in the opposite directions: higher the score, the lower the difficulty rating. In fact, there is a correlation of -0.6 between two measures. Thus, it seems that people tend to think of a game difficult, if they cannot get a high score.

6 Conclusion and Future Research

As a part of a larger research, this paper presented a game-based experimental design to test the effects of delay, feedback and nonlinearity on game performances and players' subjective difficulty ratings, which is a new measure of complexity that was not used before this research. A very simple growth management game is designed and used in the experiments. The base game does not have any dynamic complexity element. Delay, nonlinearity and feedback (naturally with a stock) is introduced to the model one by one. Their levels are changed in the experiments by changing delay order, delay time, shape of nonlinear functions and gain of feedback loop in the model. Three factors are tested in isolation, using 4 to 8 levels in Latin square design with 20 subjects. Players are provided one simple and one rather difficult game at the beginning with predefined difficulty levels. They are asked to assess the difficulties of remaining games on a discrete scale from 1 to 9. In the very end, they are given the simple game once more.

Analysis of the results indicated surprising findings in terms of effects of dynamic complexity factors, their effects on experience learning, as summarized above. Comparing the averages with the base games, no complexity factor other than the delay has any worsening effect on game performance. The delay factor on the average yields poorer performance compared to the no-delay base game. But as the delay level is increased within the delay games, the subjects surprisingly perform better. This observation is explained by the fact that each action has two-fold results when there is delay: one immediate effect and one long-term effect. While the immediate effect is instantly observed, the long-term effect shows itself with a delay. When there is longer delay, it is possible to obtain short-term profit gains that are not completely balanced out with the long-term loss, which makes a net positive gain. Also, when the delay time is large, it takes more time for the benchmark to reach its maximum level. Since the player's profit is divided by the benchmark to find the game score, it becomes clear why longer delays creates higher scores. On the average, nonlinear games resulted in higher game scores than the base linear games. Although this improvement may be explained by the significant experience effect, the very structure of the nonlinear games could be responsible on the improvement as well; as the degree of nonlinearity of the effect functions is increased, the resulting profit surface changes shape producing a larger area with high profits. The statistical results of the feedback group were not as anticipated as well. Increasing the strength of the negative feedback loop generated higher scores in a growth game. Further analysis revealed that increasing feedback strength makes the profit surface flatter and the game easier to control. Another interesting observation is that the effect of repeated trials turned out to be significant only in nonlinearity group, where players do not play any game involving any stock or delay. The inclusion of any sort of delay may be preventing procedural learning between games. On the other hand, there are also indications of false learning; the scores of the last base game of the delay group is not superior to the average of games involving delay despite the last base game being an easier game.

The next step in this research will be carrying out almost the same experiment setting on a stock management game. The initial results are mostly in line with out expectations. After these experiments are analyzed, the aforementioned dynamic complexity factors and stock being an additional factor will be tested in interaction on the two types of games. In this phase, one level from each factor will be used and all two-, three- and four-way interactions will be tested. A larger subject set (24 to 48) will be used. In addition to game performance and subjective difficulty assessments, we are planning to test players' conceptual learning of the underlying system. When completed, this research aims to provide a procedure that will help improve learning from simulation games.

References

Barlas, Y. and M. G. Özevin, 2004, "Analysis of Stock Management Gaming Experiments and Alternative Ordering Formulations," *Systems Research and Behavioral Science*, Vol. 21, pp. 439–470.

Booth Sweeney, L. and J. D. Sterman, 2000, "Bathtub dynamics: initial results of a systems thinking inventory," *System Dynamics Review*, Vol. 16, No. 4, pp. 249–286.

Brehmer, B., 1995, "Feedback Delays in Complex Dynamic Decision Tasks," in Frensch, P. and J. Funke (eds.), "Complex Problem Solving: The European Perspective," pp. 103–130, Lawrence Erlbaum, Hillsdale, NJ, U.S.A.

Broadbent, B. and B. Aston, 1978, "Human Control of a Simulated Economic System," *Ergonomics*, Vol. 21, pp. 1035–1043.

Cronin, M., C. González and J. D. Sterman, 2009, "Why dont well-educated adults understand accumulation? A challenge to researchers, educators, and citizens," *Organizational Behavior and Human Decision Processes*, Vol. 108, No. 1, pp. 116–130.

Diehl, E. and J. D. Sterman, 1995, "Effects of Feedback Complexity in Dynamic Decision Making," *Organizational Behavior and Human Decision Processes*, Vol. 62, No. 2, pp. 198–215.

Diehl, E. W., 1989, "A study on human control in stock adjustment tasks," *Proceedings of the 1989 International Conference of the System Dynamics Society*, Stuttgart, Germany, pp. 205–212, The System Dynamics Society.

Faria, A. J. and W. J. Wellington, 2004, "A survey of simulation game users, former-users, and never-users," *Simulation & Gaming*, Vol. 35, No. 2, p. 178.

Gary, M. S. and R. E. Wood, 2005, "Mental Models, Decision Making and Performance in Complex Tasks," *Proceedings of 23rd International Conference of the System Dynamics Society*, Boston, MA, U.S.A., The System Dynamics Society.

González, C., P. Vanyukov and M. K. Martin, 2005, "The use of microworlds to study dynamic decision making," *Computers in Human Behavior*, Vol. 21, No. 2, pp. 273–286.

Kampmann, C. E., 1992, *Feedback Complexity and Market Adjustment: An Experimental Approach*, Ph.D. Dissertation, Massachusetts Institute of Technology.

Langley, P., M. Paich and J. D. Sterman, 1998, "Explaining capacity overshoot and price war: misperceptions of feedback in competitive growth markets," *Proceedings of 16th International Conference of the System Dynamics Society*, Québec City, Canada, The System Dynamics Society.

Paich, M. and J. D. Sterman, 1993, "Boom, Bust, and Failures to Learn in Experimental Markets," *Management Science*, Vol. 39, No. 12, pp. 1439–1458.

Sterman, J. D., 1989, "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment," *Management Science*, Vol. 35, No. 3, pp. 321–339.

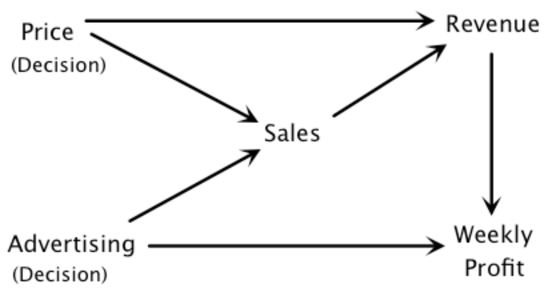
Young, S. H., C. P. Chen, S. W. Wang and C. H. Chen, 1997, "The landmine structure and the degrees of freedom of decision-making," *Proceedings of 15th International Conference of the System Dynamics Society*, Istanbul, Turkey, pp. 15–20, The System Dynamics Society.

A Game Instructions

A.1 Delay Group

This interactive simulator is about a company producing personal care products that aims to increase its profit. The company operates in a hypothetical world and the general rules of economy may not work as they do in the real world. Among various products, the company focuses on its new brand lotion. At the moment, each bottle is sold for 20 $. However, the executive board allows the product managers to change the price. Also, there is an ongoing advertising campaign for the product. The company only uses radio commercials as their advertising tool. Currently, 10 min/week of advertising is aired on radio. The product manager can increase this if necessary. The advertising cost per minute is constant whatever the total advertising amount is. The production capacity is assumed unlimited.

You are the product manager of this new lotion. Your aim is to increase your profit. As the product manager, you have two tools to control the profit: minutes of radio advertising duration per week and the unit price of the product. The figure below gives a broad representation of the causal relationships between key variables. Note that this is a general overview and there may be some other causal links in particular game versions you play.



As the Advertising duration increases, or as the Price decreases, the Sales volume increases. Other things being equal, higher the Sales, higher the Revenue, and higher the Revenue, higher the Weekly Profit. However, there are tradeoffs: (1) decreasing the Price also decreases the Revenue obtained from the Sales since Revenue is calculated by multiplying Sales volume by unit Price; (2) increasing the Advertising may decrease the Profit because the Profit is calculated by deducing the Advertising expenditure and the Production Cost of the lotion from the Revenue. Mathematically, the Weekly Profit is calculated as:

Revenue − (Minutes of Advertising × Advertising cost per minute) − (Production Cost × Sales Volume)

Advertising cost per minute and Production Cost are constant but unknown. You will decide on the Advertising duration and the unit Price for 40 weeks and your objective is to increase your cumulative profit as much as possible, in a sustainable way. This means increasing the weekly profit toward a maximum level, while avoiding excessive up and down movements toward the end of the game.

You will play 11 different games. Each game will be independent from each other. The first game will be the base game and will have the simple underlying structure explained above. The second game will be a much difficult game and will be a modified version of the first game. The remaining eight games will be different from the base game only in one aspect: there will be a delay between the input decisions and their effects on the Sales, and the duration and order of these delays will be varied from game to game. Since there is delay, your decisions will not show their full effects on Profit immediately. The games may have different initial conditions and parameter values, so a specific strategy that works in one game may not automatically work in another game. Finally, at the very end of the experiments you will play the simplest base game once again.

You will be asked to assess the difficulty of achieving success in each game on a scale from 1 to 9 as shown below.

EXTREMELY EASY		EASY		AVERAGE		HARD		EXTREMELY HARD
1	2	3	4	5	6	7	8	9

The base game (the first and the last game) is already assigned a difficulty of 1. The second game with a much higher difficulty has an assigned difficulty measure of 7 as a reference. There is no correct answer in the difficulty assessments. You may assign the same difficulty to two or more games, if you think they are at the same difficulty level. You do not have to utilize the entire scale up to 9. If you think the games are not that hard, you may assign all of them difficulties lower than 9. There will be no systematic playing order with respect to difficulty. You have to rate the difficulty of the remaining eight games. After each game, circle your difficulty assessment of the game on the sheet provided. You can revise your previous assessments after playing and observing the difficulties of the succeeding games. At the end of the 11th game, you will be asked to finalize your rating list and return it to the facilitator.

Do not open or play the games before you are told to do so. You will play the games in a specific order as indicated by numbers. For opening a game, double click the file. The game screen is as shown above. When you open the game file click the Start button once to start the game. This will initialize the game and advance you to the first week. You cannot change the first weeks profit so do not move the sliders before clicking the Start button. Each week, you have to set a Price and Advertising value using the sliders and click the Advance button once. You will observe the behavior of Weekly Profit on the graph in blue and see its numerical value in a blue box above the graph. You will also observe the benchmark behavior of weekly Profit on the same graph in red and see its numerical value on a red box. The benchmark behavior represents the ideal case where the decisions are set throughout the game so that the long-run profit is maximized. The purpose of the benchmark is to give you an idea about a good performance. When you complete 40 weeks, a warning box will appear. When you finish the game you should (1) write down your Cumulative Profit and your difficulty assessment on the sheet provided, (2) click the Exit button and (3) save the game when you are asked. Do not play any game more than once, pass to the next game. If you did something by error that you did not intend to do, please stop immediately and inform the facilitator. You will have a trial game at the beginning for you to familiarize with the game interface. Please take your time to experiment with the controls and understand how controls work.

Make sure that you understand the instructions completely before you start the experiments. If there is anything you do not understand, please ask your questions before you start playing. It is important that you know what you have to do in the experiments. For the validity of the results, it is necessary that the experiment be carried out as intended. Work on your own and do not talk to the other subjects.

You will get a base payment of 5 Liras for participating and completing the experiment properly. If you forget to save the game files or do not fill out the game sheet, you will not receive any payment. Depending on your performance in the games that you play, you will earn an additional payment. If you show the best performance among four players playing identical games with you, you will earn an extra payment of 13 TL (hence a total of 18 TL). Your extra payment will be 5 TL and 2 TL if you get the second or the third place. Only performance scores will be used in determining payment amounts. Your difficulty assessments will not have any effect on your payment. Thank you for your participation.

A.2 Nonlinearity Group

Only fifth paragraph is modified as follows:

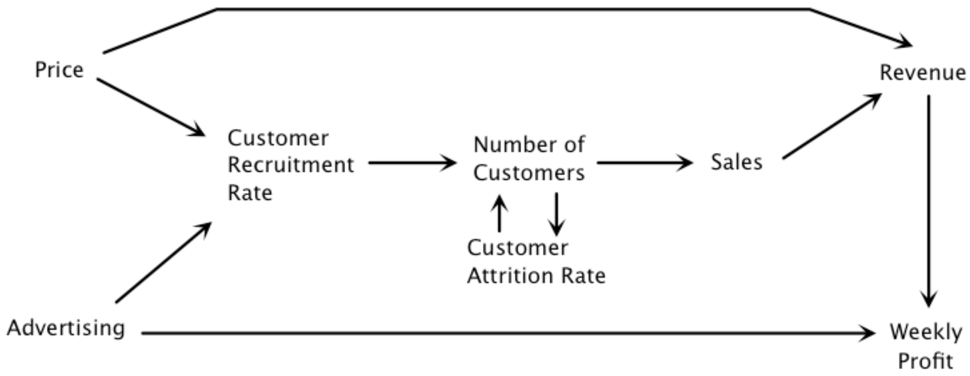
You will play 11 different games. Each game will be independent from each other. The first game will be the base game and will have the simple underlying structure explained above. The second game will be a much difficult game and will be a modified version of the first game. The remaining eight games will be different from the base game only in one aspect: the functions representing the effects of the decisions on Sales will be nonlinear and the degree (the specific shape) of these nonlinearities will be varied from game to game. The games may have different initial conditions and parameter values, so a specific strategy that works in one game may not automatically work in another game. Finally, at the very end of the experiments you will play the simplest base game once again.

A.3 Feedback Group

This interactive simulator is about a company producing personal care products that aims to increase its profit. The company operates in a hypothetical world and the general rules of economy may not work as they do in the real world. Among various products, the company focuses on its new brand lotion. At the moment, each bottle is sold for 20 $. However, the executive board allows the product managers to change the price. Also, there is an ongoing advertising campaign for the product. The company only uses radio commercials as their advertising tool. Currently, 10 min/week of advertising is aired on radio. The product manager can increase this if necessary. The advertising cost per minute is constant whatever the total advertising amount is. The production capacity is assumed unlimited.

You are the product manager of this new lotion. Your aim is to increase your profit. As the product manager, you have two tools to control the profit: minutes of radio advertising duration per week and the unit price of the product. The figure below gives a broad representation of the causal relationships between key variables. Note that this is a general overview and there may be some other causal links in particular game versions you play.

The Number of Customers increases with Customer Recruitment Rate and decreases with Customer Attrition Rate. As the Advertising duration increases, or as the Price decreases, the Customer Recruitment Rate increases. Customer Attrition Rate is simply a fraction of Number of Customers. Sales is directly proportional to the Number of Customers (Two customers create



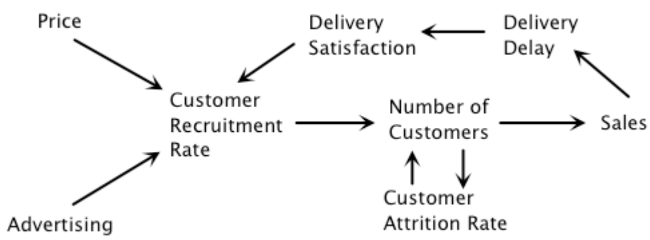
one unit of sale per week). Other things being equal, higher the Sales, higher the Revenue, and higher the Revenue, higher the Weekly Profit. However, there are tradeoffs: (1) decreasing the Price also decreases the Revenue obtained from the Sales since Revenue is calculated by multiplying Sales volume by unit Price; (2) increasing the Advertising may decrease the Profit because the Profit is calculated by deducing the Advertising expenditure and the Production Cost of the lotion from the Revenue. Mathematically, the Weekly Profit is calculated as:

Revenue − (Minutes of Advertising × Advertising cost per minute) − (Production Cost × Sales Volume)

Advertising cost per minute and Production Cost are constant but unknown. You will decide on the Advertising duration and the unit Price for 40 weeks and your objective is to increase your cumulative profit as much as possible, in a sustainable way. This means increasing the weekly profit toward a maximum level, while avoiding excessive up and down movements toward the end of the game.

You will play 11 different games. Each game will be independent from each other. The first game will be the base game and will have the simple underlying structure explained above. The second game will be a much difficult game and will be a modified version of the first game.

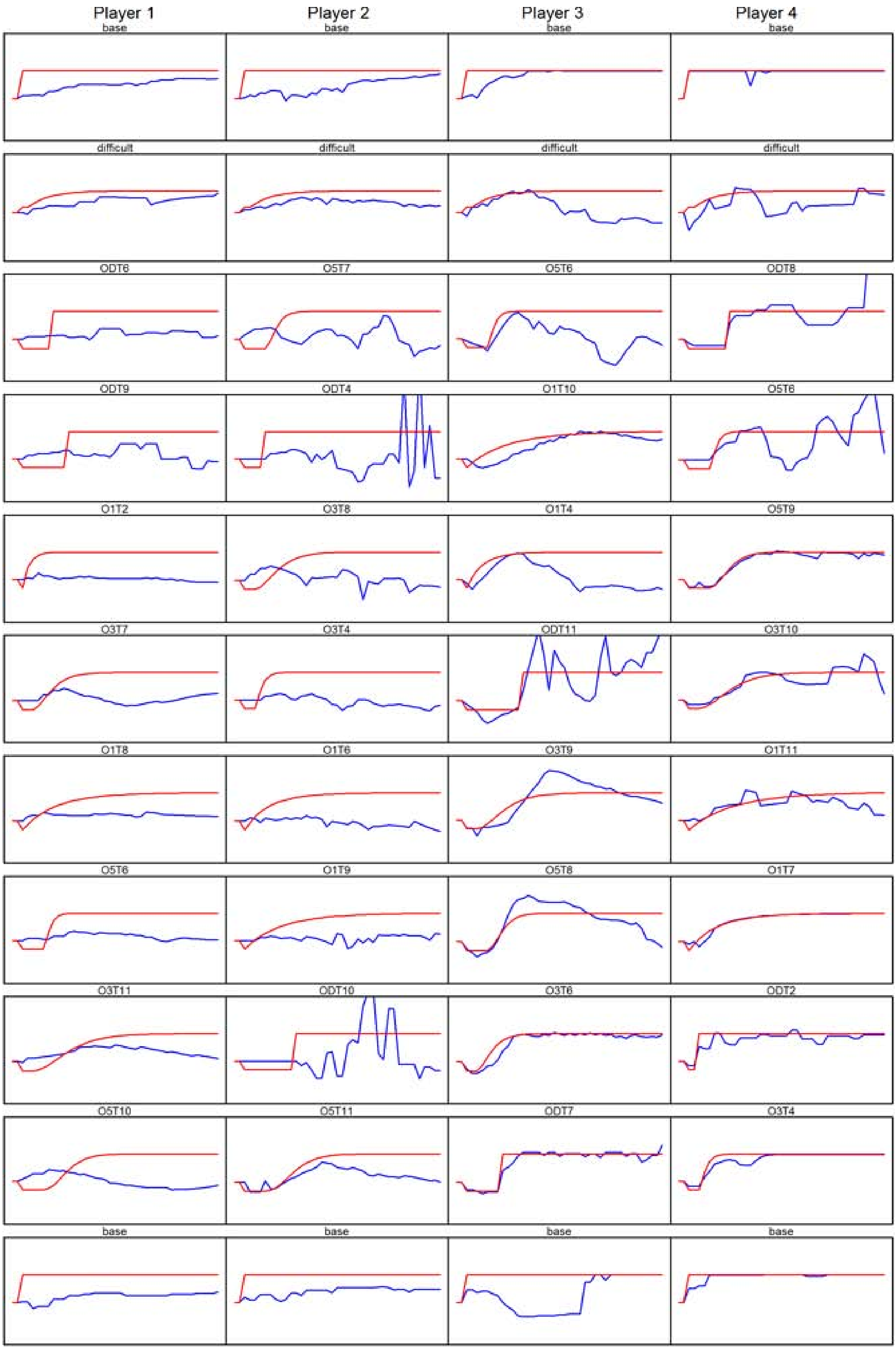
The remaining eight games will have a slightly different structure involving a new feedback loop, as shown below. Increased Customer Recruitment Rate increases Number of Customers, which results in increased Sales, leading to a rise in Delivery Delay due to insufficient delivery capacity. This in turn creates a decline in Delivery Satisfaction of customers, which has a negative effect on the Customer Recruitment Rate, completing a negative feedback loop. This loop works in a balancing way and the system will resist to indefinite increase in Sales. The strength of this feedback loop is varied from game to game.

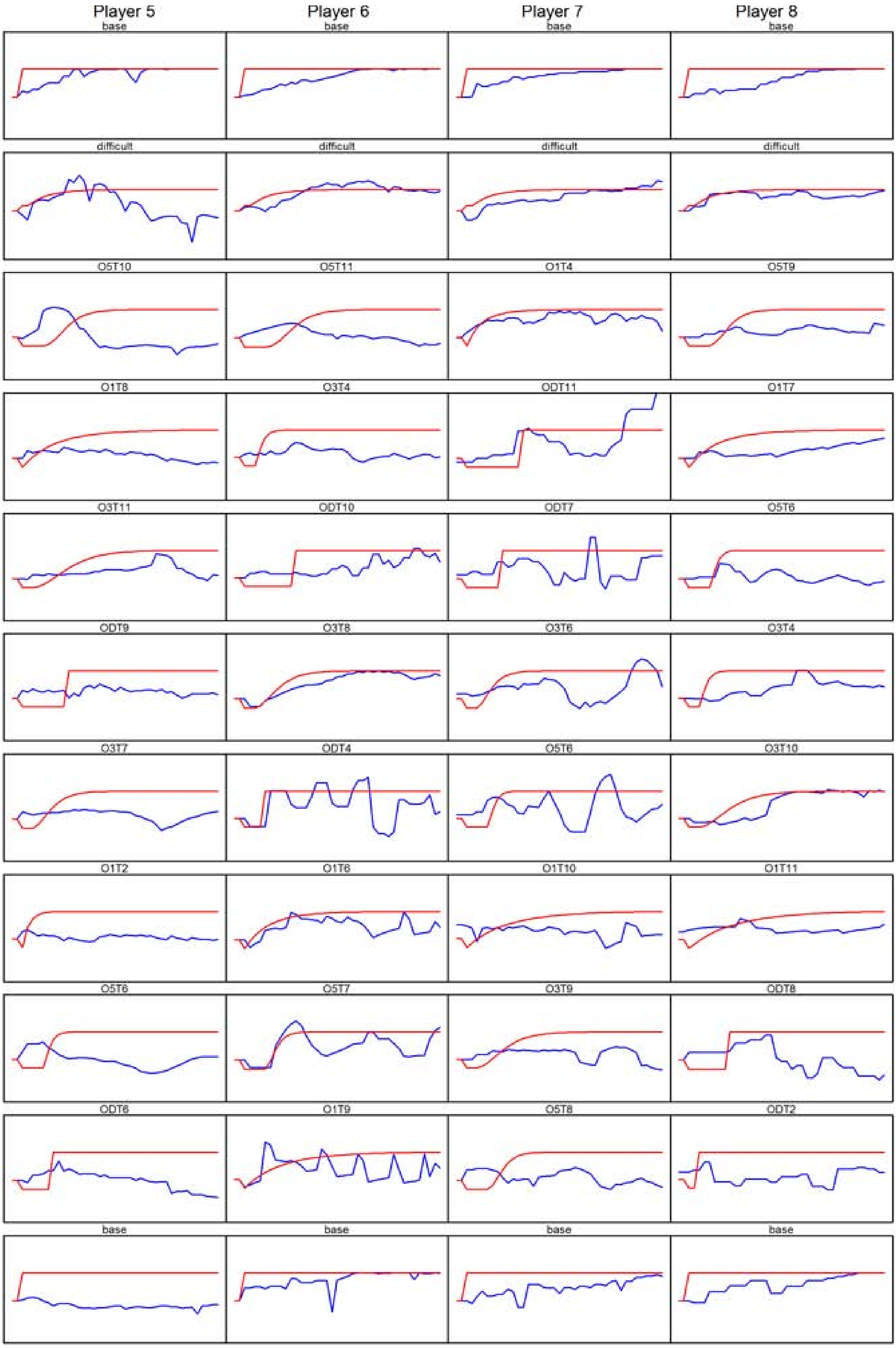


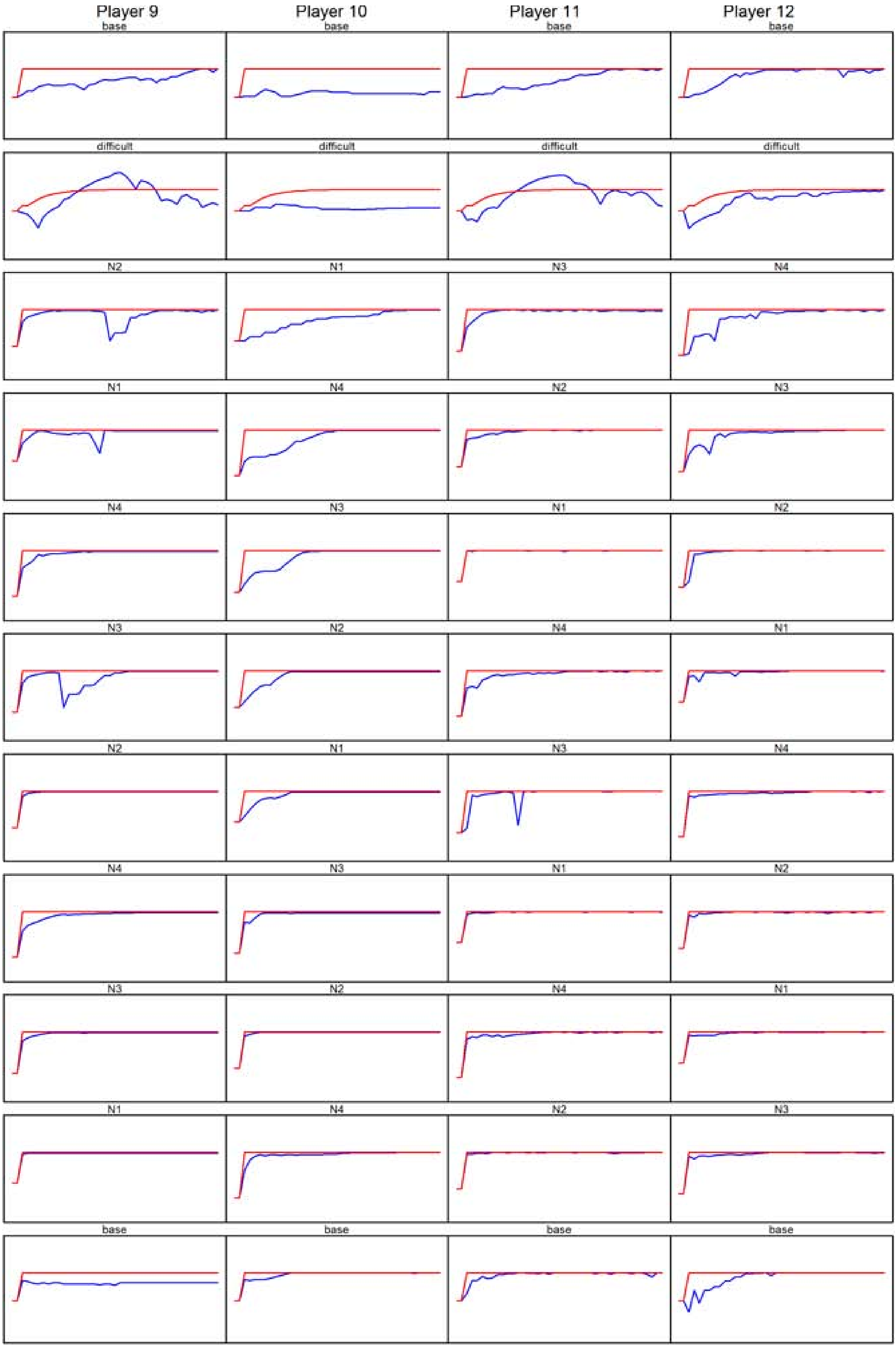
The rest of the instructions is the same as the delay group instructions.

B Game Behaviors

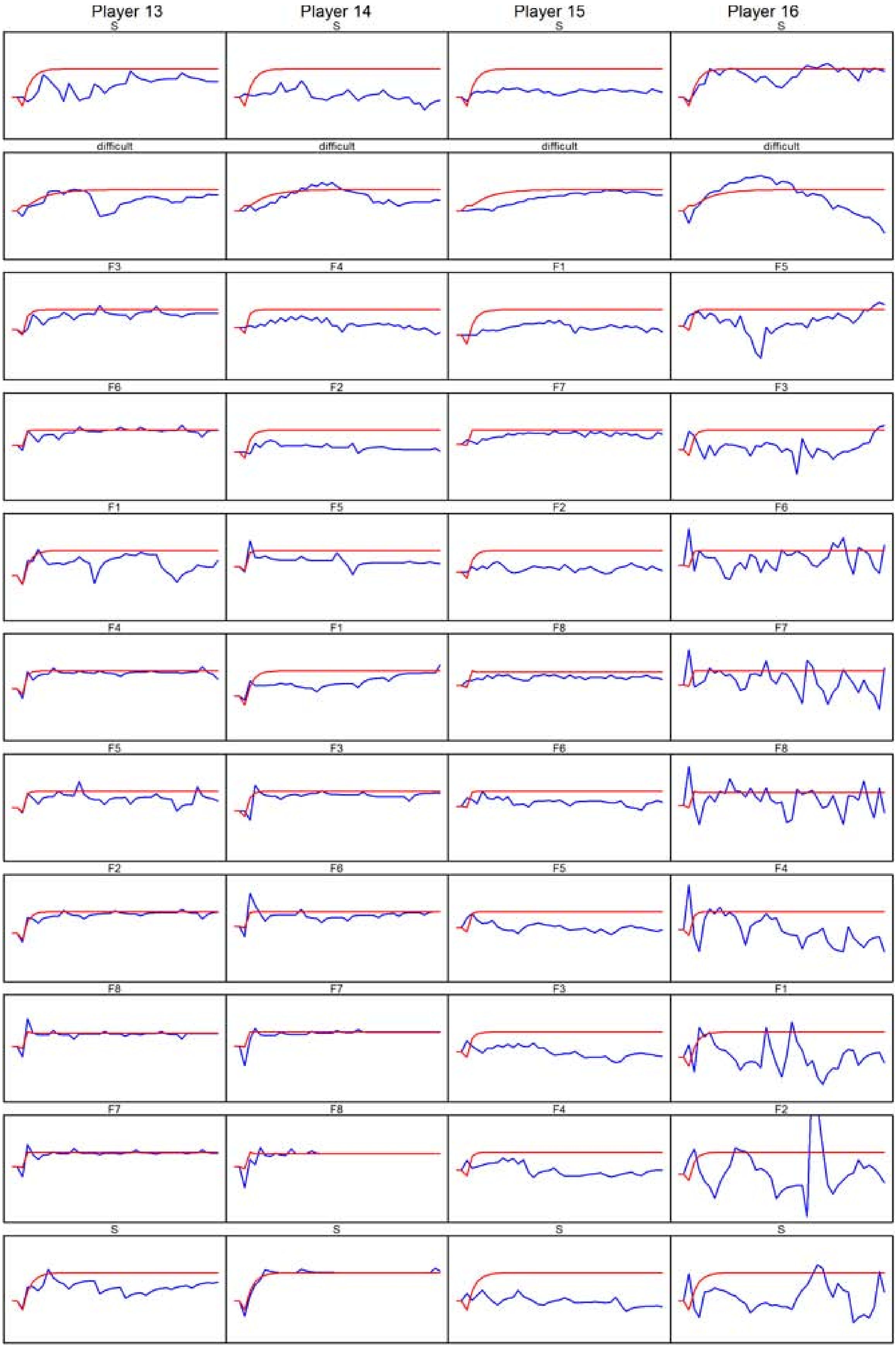
The following figures show the behaviors of *weekly profit* (blue) and *weekly benchmark profit* (red). The columns show players and the rows show playing order. The game versions are indicated above each plot.

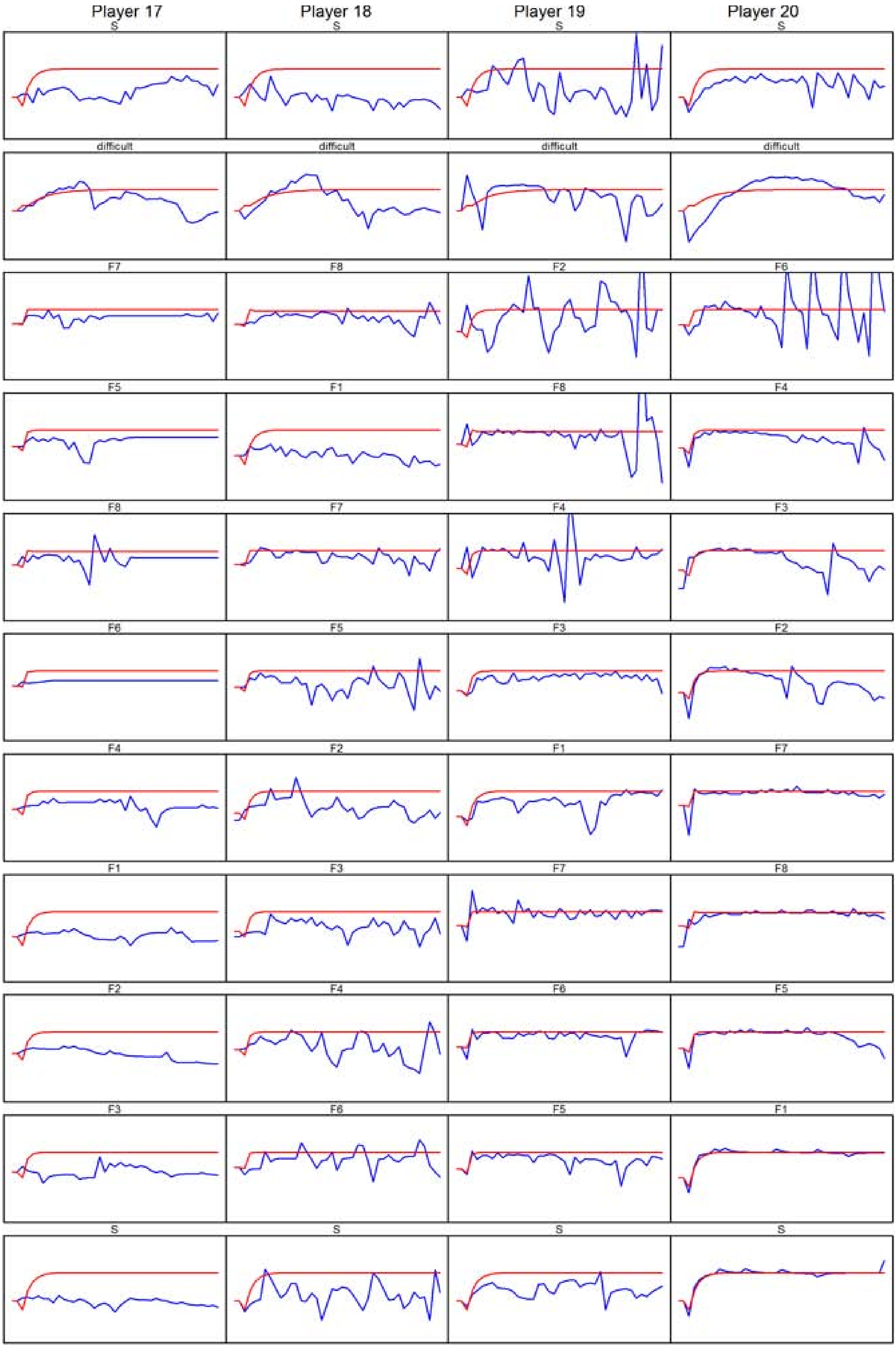



Player 5
Player 6
Player 7
Player 8
base
base
base
base
difficult
difficult
difficult
difficult
O5T10
O5T11
O1T4
O5T9
O1T8
O3T4
ODT11
O1T7
O3T11
ODT10
ODT7
O5T6
ODT9
O3T8
O3T6
O3T4
O3T7
ODT4
O5T6
O3T10
O1T2
O1T6
O1T10
O1T11
O5T6
O5T7
O3T9
ODT8
ODT6
O1T9
O5T8
ODT2
base
base
base
base



Player 9
Player 10
Player 11
Player 12
base
base
base
base
difficult
difficult
difficult
difficult
N2
N1
N3
N4
N1
N4
N2
N3
N4
N3
N1
N2
N3
N2
N4
N1
N2
N1
N3
N4
N4
N3
N1
N2
N3
N2
N4
N1
N1
N4
N2
N3
base
base
base
base


Player 13
Player 14
Player 15
Player 16
S
difficult
F3
F4
F1
F5
F6
F2
F7
F8


Player 17
Player 18
Player 19
Player 20
S
difficult
F7
F8
F2
F6
F5
F1
F4
F3
S