SYSTEM DYNAMICS AND LABORATORY EXPERIMENTS

Santiago Arango Aramburo, PhD
Universidad Nacional de Colombia, Medellín, Colombia
Centro de Complejidad – CeiBA
Carrera 80 No. 65 – 223, oficina M8A-211, Escuela de Sistemas
Phone/fax: +57 (4) 4255371 / +57 (4) 2341002
saarango@unalmed.edu.co

Jaime Andrés Castañeda Acevedo, MSc (c)
Universidad Nacional de Colombia, Medellín, Colombia
Centro de Complejidad – CeiBA
Carrera 82 A No. 21 – 159
Phone/fax: +57 (4) 3430773 / +57 (4) 3430773
jacasta2@unalmed.edu.co

Yris Olaya Morales, PhD
Universidad Nacional de Colombia, Medellín, Colombia
Centro de Complejidad – CeiBA
Carrera 80 No. 65 – 223, oficina M8A-209, Escuela de Sistemas
Phone: +57 (4) 4255352
yoyalam@unalmed.edu.co

ABSTRACT
This paper is a review of research on the application of laboratory methods to System Dynamics (SD). Although laboratory methods have been used in psychology for many years, our review focuses on the laboratory experiments developed from the experimental economics field and on their contributions to SD and the management sciences. In particular, we examine the use of experimental methods for estimating the decisions used by SD models, and the intersection of SD and laboratory experiments in theory testing and theory building. We also discuss methodological issues that experiments in SD should address to improve the value of experimental results and we remark the main findings of the reviewed works.

Keywords: System Dynamics, Laboratory Experiments, Dynamic Decision-Making, Bounded Rationality.

1. INTRODUCTION

In economics, a laboratory experiment is defined as a formal method that allows researchers to test economic theories or generate new hypothesis in a controlled environment (Roth, 1983). The basic principles are the Induced Value theory (Smith, 1976 and 1982) and the precept of Parallelism (Smith, 1982). The Induced Value theory states that, to control the economic environment of the experiment, experimental subjects should be rewarded contingent to their performance, while the precept of Parallelism deals with the external validity of the experimental data. The main purpose of performing experimental research in economics is to test the validity of theoretical results.

Experimentation in SD started at MIT in the late 80s with Sterman's experiments on capital investment (Sterman, 1987 and 1989a) and multi-stage supply chain management (Sterman, 1989b). Initially, Sterman (1987) applied the methodology for testing a behavioral simulation model for capital investment in a simple macroeconomic setting. Later, Sterman used experimentation for developing the Misperceptions of Feedback hypothesis from another capital investment game (Sterman, 1989a) and a multi-stage-supply chain game known as the *Beer Game* (Sterman, 1989b). These experiments showed serious failures in the decision makers' ability to understand the interactions between their decisions and the environment and gave rise to a body of research that have took advantage of SD tools for improving experimental designs.

SD sets experiments up in complex environments exhibiting feedback structures, delays and non-linearities, approximating to real decision environments more closely than experiments in economics and psychology and enhancing this way the study of Dynamic Decision-Making (DDM) (Edwards, 1962). Some of the issues studied by experimental decision-making in dynamic environments are the understanding of bio economics (*e.g.*, Moxnes, 2000), behavioral factors involved in the cooperation dilemma beyond Ostrom's theory (Ostrom, 1998) of collective action (*e.g.*, Castillo & Saysel, 2005), and other economic phenomena such as commodity cycles (*e.g.*, Arango, 2006). The experiments in DDM indicate that decision makers systematically misperceive environments characterized by interacting feedback loops, time delays and nonlinearities (*e.g.*, Sterman, 1989a and 1989b). Moreover, these misperceptions are not limited to dynamically complex environments; laboratory experiments in simple dynamic environments also show serious failures in decision makers' ability to understand basic systems thinking concepts (*e.g.*, Sweeney & Sterman, 2000; Cronin *et al.*, 2009). The design improvements offered by SD tools to experimentation and the variety of problems addressed by these experimental studies show that laboratory experiments in SD can greatly contribute to the analysis, testing and construction of theories.

The rest of the paper is organized as follows: First, we provide a brief background on laboratory experiments and its principles. Then, we discuss the use of laboratory experiments in SD, pointing out the experimental designs and some experimental results. We conclude with a discussion on results and some methodological issues.

2. PRINCIPLES OF LABORATORY EXPERIMENTS

Laboratory experiments take place in a controlled environment composed by three elements: first the goal, which corresponds to the objective pursued by the individual participants of the experiment. Second the system, which describes the decision-making environment and behavioral rules. And third the behavior, which corresponds to the decisions made by the individual participants of the experiment (Friedman & Sunder, 1994; Friedman & Cassar, 2004).

For example, the goal could be that the agents of a given market maximize their profits, the system could be a particular type of market where buyers can only buy a fixed number of units from sellers, and the behavior would be the purchasing and selling decisions. In this environment, the experimenter controls the goal and the system and observes subjects' behavior (Smith, 1994). Figure 1 shows how the basic components of a laboratory experiment on decision-making can be applied to the SD method.

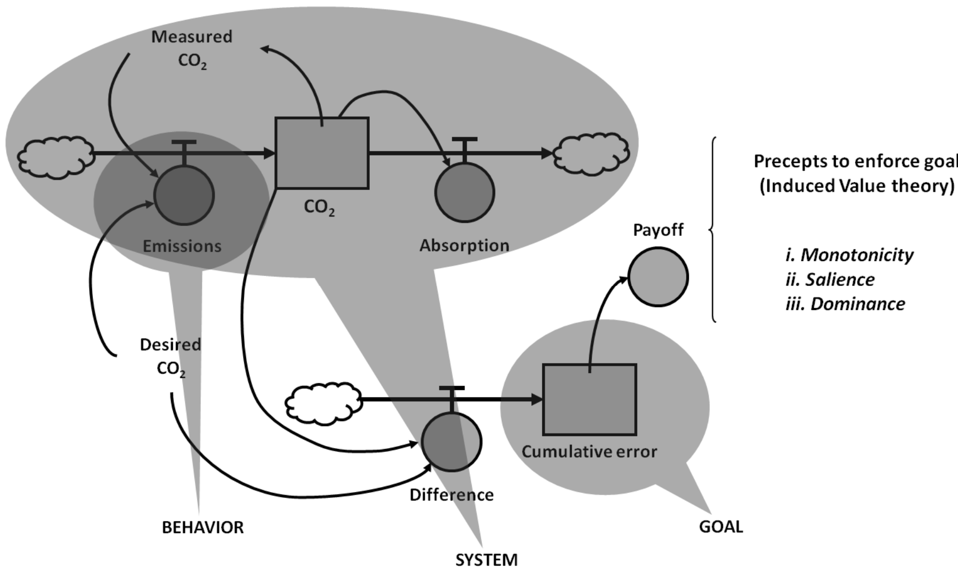

Figure 1. Example of a laboratory on decision-making: control total global emissions of CO_2 to reach a given target for the atmospheric CO_2 concentration. Source: adapted from Moxnes & Saysel (2009).

The basic principles of laboratory experiments are the Induced Value theory (Smith, 1976 and 1982), and the precept of Parallelism (Smith, 1982). The Induced Value theory states that the proper use of a reward medium allows inducing a specific behavior in the agents in such a way that their particular interests do not interfere with the purpose of the experiment. Parallelism, which we explain later in this section, resort to the general principle of induction for dealing with the external validity of the data gathered in the experiments.

Monotonicity, salience and dominance are the sufficient conditions to induce behavior (Smith, 1982). *Monotonicity* means that with a proper reward medium, more is always better (or, alternatively, less is always better). For instance, it can be assumed that every human subject prefers more cash earnings to less, and prefers less hard work to more. *Salience* means that the reward received by subjects depends on their actions and on the actions of the rest of the subjects, and that they understand this. For instance, a reward of

US $1 for every Experimental $1 earned in the market is salient because it depends on subjects' actions. Finally, *dominance* means that changes in subjects' utility come from the reward medium and the rest of influences are irrelevant. For instance, subjects are often concerned about other subjects' reward. The experimental procedures must therefore make irrelevant other subjects' reward by making impossible to know or estimate the rewards earned by them. Regardless of the differences in subjects' characteristics, when we fulfill the monotonicity, salience, and dominance conditions we enforce the goal of the experiment and can make conclusions from observed changes in behavior after changing a control variable. A usual way of satisfying these conditions is to make payments (in local currency) greater than the opportunity cost of subjects (Hey, 1996).

Although the contributions of experimental economics to empirical economic analysis are generally recognized by economists, the criticism about the validity of experimental data persists. Critics of experimental economics argue that experimental results are not representative of real economic phenomena (Loewenstein, 1999; Fatás & Roig, 2004). Smith (1982) addresses this concern by proposing the precept of Parallelism. According to this principle, behavioral regularities will persist in new situation as long as the relevant underlying conditions remain substantially unchanged. Thus, if a particular laboratory environment differs significantly from real world, a new experiment may be conducted to study the effect of such differences on human behavior.

From the point of view of external validity, the simplicity of the experiments is a virtue rather than a defect. The reason for this is that real world is often too complex to approximate closely in the laboratory and futile attempts to do so would decrease the scientific value of the experiment. Experimental simplicity by contrast, allows controlling over the variables and offers the best opportunity to gain insight about the questions that motivated the research (Friedman & Sunder, 1994; Friedman & Cassar, 2004). The method of Experimental Economics is well presented by Friedman & Sunder (1994) and/or Friedman & Cassar (2004), while the main results are summarized in The Handbook of Experimental Economics Results (Plott & Smith, 2008). Now we turn to discuss about laboratory experiments in SD field.

3. LABORATORY EXPERIMENTS IN SYSTEM DYNAMICS

Since the 80s, SD researchers on decision-making have used the methodological framework of laboratory experiments to study DDM. Research in this field emphasizes the link between subjects' behavior and system evolution (Paich & Sterman, 1993). Subjects' decisions alter the state of the system in ways that change the decision environment faced in the future (Edwards, 1962; Brehmer, 1992; Paich & Sterman, 1993).

Laboratory experiments in SD explore DDM utilizing experimental tasks that consist either of microworlds or simulators encompassing an underlying SD model (which can be a model capturing the structure of a supply chain, a corporate environment, some market institutions, etc.) coupled with a user interface, cutting some feedback loops to study subjects' decisions (Gary *et al.*, 2008) or descriptions of scenarios for which subjects have to project their behavior over time or answer some punctual questions about them (Moxnes, 2004).

SD sets experiments up in complex environments exhibiting feedback structures, delays and non-linearities. In this way, SD responds to the criticism of laboratory data not representing real economic phenomena because they are gathered in simple laboratory environments (Loewenstein, 1999; Fatás & Roig, 2004).

In the following sections we examine the application of laboratory experiments in SD. These laboratory studies address issues such as boom-and-bust dynamics, capital investment, corporate management, supply chains among others. To facilitate their comparison, we classify the experimental studies in three broad research lines: estimation of decision rules, theory building and theory testing.

Estimation of Decision Rules

Traditional methods for testing simulation models of human behavior draw on established organizational and economic theory to specify the model, followed by estimation of the parameters and sensitivity test. Specifying the model is relatively straightforward, but discovering and representing the decision rules of the actors is subtle and difficult. These traditional methods are unsatisfying to many economists and simulation modelers as well because such methods are unable to validate the behavioral decision rules since they are based on the "what" of the decisions, not the "why" (Sterman, 1987). Moreover, these methods are heavily based on the assumptions of rational behavior even when these assumptions are contrary to fact (Simon, 1979). Thus, behavioral simulation models must portray decision-making behavior as it is, and not as it might be if decision makers were omniscient optimizers (Sterman, 1987). In this regard, experimental methods offer a complementary approach to traditional methods like econometrics for estimating or bootstrapping decision rules since they use interactive gaming in which people play a role in the system being modeled portraying an institutional context corresponding to that of the model to be tested and are given the same information set, but are free to make decisions any way they wish. In this way, decision rules can be bootstrapped from data on decisions and the information available to subjects at the time they made those decisions (Gary *et al.*, 2008).

In this regard, Sterman (1987) presented the first experiment in SD field devoted to the estimation of a decision rule. In this experiment, Sterman bootstrapped a decision rule for capital investment dynamics in a simple macroeconomic model using experimental data from aggregate capital investment experiments. Similarly, Sterman (1989b) estimated a decision rule for stock management from the *Beer Game*. Both estimation processes identified several cues which account for the poor performance of the subjects, particularly showing that subjects are insensitive to delays and the supply line (orders made, but not received yet). Using experimental data from an experiment on managing a new product, Paich & Sterman (1993) estimated a decision rule for boom-and-bust dynamics, indentifying that subjects fail to utilize important cues like the actual market demand and the growth in demand, which account for the poor performance of the subjects. While these works used data from laboratory experiments, Castillo & Saysel (2005) used data from field experiments on common pool resource management to estimate a behavioral model of the rational choice theory of collective action (Ostrom, 1998) considering the payoff structure used in the experiments and the experimental results, finding that other important behavioral factors to account for subjects' behavior besides those considered by Ostrom's theory are temptation to free ride, awareness and risk perception.

Theory Building

Like economics and/or psychology, SD researchers have used experiments to present behavioral hypotheses, mainly to describe why subjects fail to control dynamically complex systems. In particular, three behavioral hypotheses can be identified: Misperceptions of Feedback, Misperceptions of Bioeconomics and Misperceptions of Basic Dynamics. While these three hypotheses draw on bounded rationality theory and emphasize on the fact that people have poor mental models, each of these focuses in particular issues as we review next.

Sterman (1989a and 1989b) presented the hypothesis of Misperceptions of Feedback to explain why subjects perform poorly in environments characterized by dynamic complexity. According to this hypothesis, subjects fail to assess correctly the nature and significance of the causal structure of a system, particularly the linkages between their decisions and the environment (Sterman, 1989a, p. 324). In Sterman (1989a) aggregate capital investment experiment, subjects play the role of manager for the entire capital-producing sector of a simulated economy making capital investment decisions to satisfy demand. Subjects' decisions led to costly oscillations in orders. In Sterman (1989b) experiment of the *Beer Game*, each brewery consists of four sectors: retailer, wholesaler, distributor, and factory, where one subject manages each sector ordering cases of beer in the face of uncertain demand. Again, subjects' decisions led to costly oscillations in orders. In both experiments, Sterman explained the results by recurring to bounded rationality theory (Simon, 1995 and 1979) and heuristics or simple decision rules (Tversky & Kahneman, 1974), which are an extension of bounded rationality (Kleinmuntz, 1993). Sterman found that subjects' behavior can be explained by an anchoring and adjustment heuristic (Tversky & Kahneman, 1974) that misperceives the delay in acquiring orders and ignores the supply line making more orders than necessary, supporting this way the Misperceptions of Feedback hypothesis.

Inspired on the Misperceptions of Feedback hypothesis, Moxnes (1998a and 1998b) established his hypothesis of Misperceptions of Bioeconomics to explain why persons mismanage bioeconomic resources. Moxnes ruled out the commons problem, which is known to cause overexploitation (Gordon, 1954; Hardin, 1968), by the design of the experiments. In the fishery experiment (Moxnes, 1998a), the commons problem was ruled out by implementing private fish stocks (fjords). Subjects had to manage a fjord by investing in a fishing fleet to extract the fish resource. Subjects overinvested in the fleet reducing the resource to levels below the optimal. In the reindeer and lichen experiment

(Moxnes, 1998b), the commons problem was ruled out by setting quotas for reindeer. Subjects had to manage a reindeer herd in order to avoid an overgrazing of lichen and the die-off of reindeers. As in the previous experiment, subjects overexploited the resources. Summarizing the key insights from the previous experiments, Moxnes (2000) stands out that subjects' behavior is explained by heuristics intentionally rational for static, flow resources, but not for dynamic, stock resources, showing that the misperceptions of feedback that people have about bioeconomic resources are a challenge for management of this type of resources, beyond the commons problem.

The laboratory experiments surveyed thus far in theory building are characterized by considerable complexity about model structure. However, there are experiments that reduce the complexity to minimum levels to assess the understanding of basic systems thinking concepts. Sweeney & Sterman (2000) experiment was the first experiment in this regard. Subjects were faced with three tasks with basic stock and flow structures describing a problem for which they had to project its behavior over time. In spite of the simplicity of the tasks, the average performance was about 55% according to the measure criterions used by the authors. Later, Ossimitz (2002) and Kainz & Ossimitz (2002) carried out a series of experiments to study if the poor performance observed in Sweeney & Sterman experiment was due either to failures in discerning between stocks and flows or failures in reading and interpreting graphs. The performance was even worse than in Sweeney & Sterman experiment and these studies demonstrated that failures to discern between stocks and flows do not depend upon whether the tasks require the subjects to read or to draw graphs or not. These three experiments gave rise to what is called the Misperceptions of Basic Dynamics hypothesis, suggesting that people intuitively use an attractive but erroneous heuristic that matches the shape of the output of the system to the shape of the input, behavior known as the correlational or pattern matching heuristic.

Theory Testing

The behavioral hypotheses previously mentioned have been tested by running experiments modifying some experimental conditions, increasing this way the robustness of the results presented initially. Additionally, some works have tested the bounded rationality theory in dynamically complex markets. Next we review some of the works devoted to these purposes.

The Misperceptions of Feedback hypothesis (Sterman, 1989a and 1989b) was expanded in order to be tested in follow-up experiments. The new experiments address different assumptions to provide more data that help to test Misperceptions of Feedback. Most of these experiments vary different characteristics that help either to increase or decrease the dynamic complexity of the environment faced by subjects and provide improved decision-making aids. In this regard, Bakken (1993) experiment on capital-intensive industries varies conditions that enhance the familiarity of the industry and the frequency of market instabilities. Paich & Sterman (1993), Langley *et al.* (1998) and Gary & Wood (2008) experiments on a new product launch boom-and-bust dynamics varies the strength of key feedback processes for product lifetime, the strategies of the simulated competitor, and the number of decision variables and the presence of competition respectively. Sengupta & Abdel-Hamid (1993) experiment on software development projects uses different information displays. Diehl & Sterman (1995) and Atkins *et al.* (2002) experiments on a stock management problem varies simultaneously the strength of feedback and the length of the delays involved in the task; additionally, Atkins *et al.* also uses different information displays. Barlas & Özevin (2004) experiment on a stock management problem varies the demand pattern, the type of delay and the decision interval. Young *et al.* (1997) and Howie *et al.* (2000) experiments on capital investment study subjects' behavior when the system falls into uncontrollable feedback loops and when subjects are faced with different information displays respectively. Domínguez *et al.* (1998) and Größler *et al.* (2000) experiments on corporate management use different information displays; and Wu & Katok (2006) experiment of the *Beer Game* implements a simpler demand distribution than the original game. The experimental results of these studies are consistent with Sterman's original results: poor performance with respect to the experiments' goals due to the use of heuristics that systematically misperceive the causal structure of the system, lending support to Misperceptions of Feedback. With only a few counterintuitive effects (*e.g.*, Bakken treatment of frequency of market instabilities and Barlas & Özevin treatment of demand pattern), these results show that complexity matters since when the dynamic complexity of the environment was increased (decreased), results worsened (improved) with respect to either the optimal or other benchmarks (Bakken, 1993; Paich & Sterman, 1993; Diehl & Sterman, 1995; Young *et al.*, 1997; Langley *et al.*, 1998; Atkins *et al.*, 2002; Barlas & Özevin, 2004; Wu & Katok, 2006; Gary & Wood, 2008). Moreover, most of the experiments that manipulated the information displays show that decision-making aids may reduce the negative effects of Misperceptions of Feedback (Domínguez *et al.*, 1998; Größler *et al.* 2000; Howie *et al.*, 2000; Atkins *et al.*, 2002).

In order to establish if the Misperceptions of Bioeconomics were due to the use of quite complex simulators, Moxnes (2004) designed a new experiment to study management of reindeers and lichen with simplified underlying SD models: one with only one stock and other with two stocks. Moxnes found that the basic tendency towards Misperceptions of Bioeconomics remains when the experiment is simplified, showing that inappropriate mental models are responsible for the poor understanding of the dynamics of bio economics no matter the complexity. Moreover, Moxnes found that subjects' behavior can be explained by an anchoring and adjustment heuristic, lending support to bounded rationality theory.

As the two previous hypotheses, Misperceptions of Basic Dynamics have also been put to the test. In this regard, Sterman & Sweeney (2002) assessed understanding of climate change by asking subjects to identify the likely response of temperature to various scenarios for CO_2 emissions or concentrations. Similarly, Sterman & Sweeney (2007) and Sterman (2008) presented subjects with a scenario for the evolution of atmospheric CO_2 and asked them to describe the emissions trajectory required to realize it. In the same line, Moxnes & Saysel (2009) asked subjects to manage a simulator to control total global emissions of CO_2 to reach a given target for the stock of CO_2 in the atmosphere and in subsequent treatments they presented subjects with aids to develop proper mental models to help them to understand the system structure. Although the tasks only required an understanding of stocks and flows and basic facts about climate change, the majority of subjects did not achieve the experiments' goals. Sterman and Sweeney's experiments found that many people used the correlational heuristic, while Moxnes & Saysel found that the tendency to use this heuristic disappears when subjects understand the structure of the system. Besides these studies assessing the understanding of climate change, Misperceptions of Basic Dynamics have also been tested in other contexts. For instance, Jensen & Brehmer (2003) asked subjects to establish equilibrium in a simple predator and prey system by managing the population of foxes in an experiment with a simple underlying SD model that only included to stocks. Although about 50% of the subjects accomplished the task, a great part of that 50% did it through and error and essay strategy, showing a poor understanding of the system. In a recent study, Cronin *et al.* (2009) designed an experiment to identify if poor understanding of accumulation was due to the inability to interpret graphs, lack of contextual knowledge, motivation, or cognitive capacity, which are common reasons for poor performance in DDM studies and stock and flow contexts. Subjects had to answer simple questions regarding inflows and outflows and the behavior of a stock from the behavior of the flows based on a graph showing the number of people entering and leaving a department store. Results showed that even in the most favorable conditions, the failure to discern between stocks and flows persists. Moreover, Cronin *et al.* found that many people, including highly educated individuals with strong technical training, use the correlational heuristic.

Different from most experiments in economics where markets are relatively simple (*e.g.*, Plott & Smith, 2008), SD experiments offer the possibility to test bounded rationality theory in dynamically complex markets while studying more specific economic issues such as equilibrium and/or commodity cycles. In this regard, Kampmann (1992) carried out experiment to test if different price institutions could reduce the negative impacts of bounded rationality. To do this, Kampmann designed experimental markets which involved three price institutions in addition to two market complexity conditions. The author found that complexity worsens subjects' performance and although performance is improved in the presence of market institutions, it remains significantly below the optimal calculated from rational expectations (Muth, 1961). Studying commodity cycles, Arango (2006) increased the complexity of the experimental markets by varying the length of delays of investment decisions and capacity lifetimes. Arango found no evidence of rational expectations; instead, he observed cyclical tendencies in prices as market complexity increased. In a follow-up experiment, Arango increased even more the length of delays and capacity lifetimes. This time, subjects' decisions led to a well-defined oscillatory behavior in prices showing that the rationality of decisions could play an important role in commodity cycles, beyond traditional economic theory which attributes such fluctuations to exogenous causes (*e.g.*, Deaton & Laroque, 1992; Deaton, 1999). In all these experiments, the authors found that subjects' behavior can be interpreted in terms of an anchoring and adjustment heuristic, lending support to bounded rationality theory. Figure 2 show some results obtained by Arango.

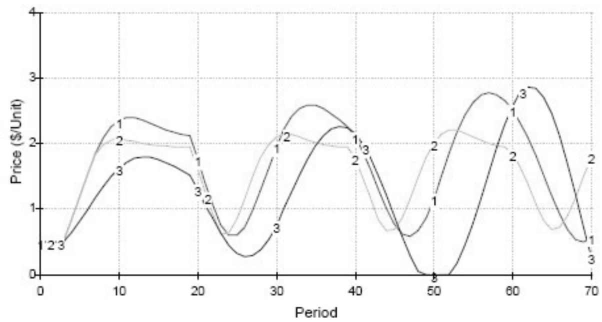

Figure 2. Simulated prices of the market with an anchoring and adjustment heuristic with different parameters: from literature (lines 1 and 2) and parameters obtained from the average estimates of the experimental results (line 3). Source: Arango (2006).

4. DISCUSSION

Results

The bootstrapping process from experimental data is a complementary approach to traditional methods for testing simulation models (Sterman, 1987, 1989a and 1989b; Arango, 2006). Moreover, the bootstrapping process can be useful on strategy research since there are opportunities to test the decision rules identify in experiments through bootstrapping decision rules using field data to see whether the rules explain variations in firm decisions and whether these variations are a source of performance heterogeneity among firms (Gary *et al* 2008). In this sense, the decision rules identified in experiments on multi-stage supply chains (Sterman, 1989b) and on boom-and-bust dynamics (Paich & Sterman, 1993) could be tested using appropriate data from the field.

In general, the literature of theory building and theory testing of laboratory experiments in SD shows that subjects' rationality deteriorates in the presence of dynamic complexity. Understanding the dynamic complexity of tasks poses cognitive difficulties. When facing dynamic complexity, subjects make their decisions based on heuristics or simple decision rules that work as mental short cuts to reduce the complexity of the tasks. Frequently, such heuristics lead to systematic deviation from optimal decisions. The use of heuristics is also observed in relatively simple tasks, showing that systematical misperceptions between the subjects' decisions and the environment are present in a great variety of tasks.

Bounded rationality theory in the form of heuristics involving dynamics explains better the decisions both in highly complex environments and relatively simple environments. (*e.g.*, Sterman, 1989a and 1989b; Kampmann, 1992; Bakken, 1993; Paich & Sterman, 1993; Diehl & Sterman, 1995; Moxnes, 2000 and 2004; Sterman & Sweeney, 2002 and 2007; Barlas & Özevin, 2004; Arango, 2006; Sterman, 2008; Cronin *et al.*, 2009). This it is not necessarily surprising since people's rationality is bounded or limited within certain contexts (Conlisk, 1996; Größler *et al.*, 2004), in particular within complex situations, where behavior may be governed by different laws than those used in simple systems (Plott, 1982; Gigerenzer, 2004).

The experimental results discussed above do not imply that people make irrational decisions because, in general, people try to be rational (March, 1994) or they look for satisfactory solutions (Simon, 1955). What the results show instead, is the cognitive difficulties of making complex decisions or of making decisions in complex environments.

Experimental results in the SD field show that people undervalue the importance of delays, misperceive the workings of stock and flow relationships, and are insensitive to nonlinearities that may alter the strengths of different feedback loops as the system evolves (Moxnes, 2000). People have poor mental models and have limited cognitive capabilities to infer the behavior of the systems as complexity increases (Kampmann, 1992; Paich & Sterman, 1993; Diehl & Sterman, 1995; Young *et al.*, 1997; Langley *et al.*, 1998; Atkins *et al.*, 2002; Barlas & Özevin, 2004; Moxnes, 2004; Arango, 2006; Gary & Wood, 2008). This is also true for simple system structures (*e.g.*, Sterman & Sweeney, 2002 and 2007; Sterman, 2008; Cronin *et al.*, 2009). However, literature also shows that information displays and other artifacts designed to aid decision makers can be useful to reduce the effects of bounded rationality (Domínguez *et al.*, 1998; Größler *et al.* 2000; Howie *et al.*, 2000; Atkins *et al.*, 2002; Kainz & Ossimitz, 2002; Moxnes & Saysel, 2009), showing an interesting line of research.

Method

Unlike most economic experiments, whose structure is static and which are reset each period (*e.g.*, Plott & Smith, 2008), SD experiments provide a decision environment more complex and hence closer to reality. Therefore, laboratory experiments in SD enforce the precept of Parallelism because their experimental design usually has an underlying SD model which mimics real world decisions by incorporating feedback structures, delays and non-linearities (Gary *et al.*, 2008). In this sense, SD experiments have a great potential to contribute to SD research on economic issues by designing dynamically complex economic environments and testing relevant economic theories in such environments.

While providing a realistic decision environment is the main contribution of SD to experimental research, much of the body research on DDM from the SD field lacks of the formal protocols of experimental methods, particularly those related to the Induced Value theory. For instance, in some experiments on basic dynamics testing (*e.g.*, Sterman & Sweeney, 2002 and 2007) participants received no monetary rewards. This lack of formality frequently decreases the value of the experimental results since it is widely believed among economists that performance-based rewards are necessary to establish external validity. In an extensive review, Camerer & Hogarth (1999) found that the effects of incentives are mixed. However, in spite of this heterogeneity they found that incentives may reduce the variance of responses. In this sense, incentives are a way of producing higher-quality data (Smith & Walker, 1993; Camerer & Hogarth, 1999). Furthermore, since the formal methods in economic experimentation have already been established and accepted by the mainstream economic community (it is essentially impossible to report experimental research in economics journals if subjects have not been financially motivated), adoption of such protocols[1] by the SD community would enable to better position experimentation in SD as a valid tool for research in dynamic and complex environments.

Education

Microworlds, also known as Management Flight Simulators, could be used for laboratory experiments; however, because microworlds are methodological different from the laboratory experiments, we do not include them in this review. Microworlds are designed to enhance people understanding of dynamically complex systems (*e.g.*, van Ackere *et al.*, 1997; Dyner *et al.*, 2009), while laboratory experiments are designed to study about how people make decisions. Thus, while subjects use Microworlds for learning, laboratory experiments are used to learn about subjects' mental models on decision-making tasks. Nevertheless, Microworlds can actually use the information coming from laboratory experiments in order to design them more comprehensive and improve the learning through their use.

In general, the literature reports that more research is needed to understand better the problem of decision-making in dynamic and complex environments. Many of the surveyed works complement and extend previous research by creating alternative experimental settings which are used to investigate the limitations of the original studies. For example, Moxnes & Saysel (2009) extend Sterman & Sweeny (2002 and 2007) and Sterman (2008) experiments to study how pedagogic aids to develop mental models affect subjects' behavior. By creating environments that recreate reality, SD and experimentation contribute to theory building and knowledge creation in the field of decision-making under dynamic complexity.

REFERENCES

Arango, S (2006). *Essays on Commodity Cycles Based on Expanded Cobweb Experiments of Electricity Markets*. PhD Thesis, University of Bergen, Social Science Faculty. Bergen, Norway.

Atkins, PWB, Wood, RE & Rutgers, PJ (2002). The Effects of Feedback Format on Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* **88**: 587-604.

Bakken, BE (1993). *Learning and Transfer of Understanding in Dynamic Decision Environments*. PhD Thesis, Massachusetts Institute of Technology, Sloan School of Management. Cambridge (MA), USA.

[1] Following Hey's (1996) seven practical requirements for laboratory experiments is a good practice for developing laboratory experiments according to the basic principles of experimental economics.

Barlas, Y & Özevin, MG (2004). Analysis of Stock Management Gaming Experiments and Alternative Ordering Formulations. *Systems Research and Behavioral Science* **21**: 439-470.

Brehmer, B (1992). Dynamic Decision Making: Human Control of Complex Systems. *Acta Psychologica* **81**: 211-241.

Camerer, CF & Hogarth, RM (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty* **19**: 7-42.

Castillo, D & Saysel, AK (2005). Simulation of Common Pool Resource Field Experiments: A Behavioral Model of Collective Action. *Ecological Economics* **55**: 420-436.

Conlisk, J (1996). Why Bounded Rationality? *Journal of Economic Literature* **34**: 669-700.

Cronin, MA, Gonzalez, C & Sterman, JD (2009). Why Don't Well-educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens. *Organizational Behavior and Human Decision Processes* **108**: 116-130.

Deaton, A. (1999) Commodity Prices and Growth in Africa. *The Journal of Economic Perspectives* **13**: 23-40.

Deaton, A & Laroque, G (1992). On the Behaviour of Commodity Prices. *The Review of Economic Studies* **59**: 1-23.

Diehl, E & Sterman, JD (1995). Effects of Feedback Complexity on Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* **62**: 198-215.

Domínguez, JA, Ruiz, JC, Domingo, MA & González, MM (1998). Our Ten Years of Work on Transparent Box Business Simulation. *Proceedings of the 16th International Conference of the System Dynamics Society*. Québec City, Canada.

Dyner, I, Larsen, ER & Franco, CJ (2009). Games for Electricity Traders: Understanding Risk in a Deregulated Industry. *Energy Policy* **37**: 465-471.

Edwards, W (1962). Dynamic Decision Theory and Probabilistic Information Processing. *Human Factors* **4**: 59-73.

Fatás, E & Roig, JM (2004). Una Introducción a la Metodología Experimental en Economía. *Cuadernos de Economía* **27**: 7-36.

Friedman, D & Cassar, A (2004). *Economics Lab: An Intensive Course in Experimental Economics*. Routledge: London.

Friedman, D & Sunder, S (1994). *Experimental Methods: A Primer for Economists*. Cambridge University Press: Cambridge.

Gary, MS & Wood, RE (2008). *Mental Models, Decision Rules, Strategies, and Performance Heterogeneity*. MIT Sloan School of Management Research Paper No. 4736-09. Sloan School of Management, MIT.

Gary, MS, Kunc, M, Morecroft, JDW & Rockart, SF (2008). System Dynamics and Strategy. *System Dynamics Review* **24**: 407-429.

Gigerenzer, G (2004). Fast and Frugal Heuristics: The Tools of Bounded Rationality. In D Koehler & N Harvey (eds.), *Blackwell Handbook of Judgment and Decision Making*. Blackwell: Oxford.

Gordon, HS (1954). The Economic Theory of a Common-Property Resource: The Fishery. *The Journal of Political Economy* **62**: 124-142.

Größler, A, Maier, FH & Milling, PM (2000). Enhancing Learning Capabilities by Providing Transparency in Business Simulators. *Simulation & Gaming* **31**: 257-278.

Größler, A, Milling, PM y Winch, G (2004). Perspectives on Rationality in System Dynamics –A Workshop Report and Open Research Questions. *System Dynamics Review* **20**: 75-87.

Hardin, G (1968). The Tragedy of the Commons. *Science* **162**: 1243-1248.

Hey, JD (1996). *Experimentos en Economía*. Fondo de Cultura Económica: México, D.F.

Howie E, Sy, S, Ford, L & Vicente, KJ (2000). Human-Computer Interface Design Can Reduce Misperceptions of Feedback. *System Dynamics Review* **16**: 151-171.

Jensen, E & Brehmer, B (2003). Understanding and Control of a Simple Dynamic System. *System Dynamics Review* **19**: 119-137.

Kainz, D & Ossimitz, G (2002). Can Students Learn Stock-Flow-Thinking? An Empirical Investigation. *Proceedings of the 20th International Conference of the System Dynamics Society*. Palermo, Italy.

Kampmann, CPE (1992). *Feedback Complexity and Market Adjustment: An Experimental Approach*. PhD Thesis, Massachusetts Institute of Technology, Sloan School of Management. Cambridge (MA), USA.

Kleinmuntz, DN (1993). Information Processing and Misperceptions of the Implications of Feedback in Dynamic Decision Making. *System Dynamics Review* **9**: 223-237.

Langley, PA, Paich, M & Sterman, JD (1998). Explaining Capacity Overshoot and Price War: Misperceptions of Feedback in Competitive Growth Markets. *Proceedings of the 16th International Conference of the System Dynamics Society*. Québec City, Canada.

Loewenstein, G (1999). Experimental Economics from the Vantage-Point of Behavioural Economics. *The Economic Journal* **109**: Features, F25-F34.

March, JG (1994). *A Primer on Decision Making: How Decisions Happen*. Free Press: New York.

Moxnes, E (1998a). Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics. *Management Science* **44**: 1234-1248.

__________ (1998b). Overexploitation of Renewable Resources: The Role of Misperceptions. *Journal of Economic Behavior and Organization* **37**: 107-127.

__________ (2000). Not Only the Tragedy of the Commons: Misperceptions of Feedback and Policies for Sustainable Development. *System Dynamics Review* **16**: 325-348.

__________ (2004). Misperceptions of Basic Dynamics: The Case of Renewable Resource Management. *System Dynamics Review* **20**: 139-162.

Moxnes, E & Saysel, AK (2009). Misperceptions of Global Climate Change: Information Policies. *Climate Change* **93**: 15-37.

Muth, JF (1961). Rational Expectations and the Theory of Price Movements. *Econometrica* **29**: 315-335.

Ossimitz, G (2002). Stock-Flow-Thinking and Reading Stock-Flow-Related Graphs: An Empirical Investigation in Dynamic Thinking Abilities. *Proceedings of the 20th International Conference of the System Dynamics Society*. Palermo, Italy.

Ostrom, E (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action. *American Political Science Review* **92**: 1-22.

Paich, M & Sterman, JD (1993). Boom, Bust, and Failures to Learn in Experimental Markets. *Management Science* **39**: 1439-1458.

Plott, CR (1982). Industrial Organization Theory and Experimental Economics. *Journal of Economic Literature* **20**: 1485-1527.
Plott, CR & Smith, VL (2008). *The Handbook of Experimental Economics Results*. North Holland: Amsterdam.
Roth, AE (1983). Toward a Theory of Bargaining: An Experimental Study in Economics. *Science* **220**: 687-691.
Sengupta, K & Abdel-Hamid, TK (1993). Alternative Conceptions of Feedback in Dynamic Decision Environments: An Experimental Investigation. *Management Science* **39**: 411-428.
Simon, HA (1995). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* **69**: 99-118.
__________ (1979). Rational Decision Making in Business Organizations. *The American Economic Review* **69**: 493-513.
Smith, VL (1976). Experimental Economics: Induced Value Theory. *The American Economic Review* **66**: 274-279.
__________ (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review* **72**: 923-955.
__________ (1994). Economics in the Laboratory. *The Journal of Economic Perspectives* **8**: 113-131.
Smith, VL & Walker, JM (1993). Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry* **31**: 245-261.
Sterman, JD (1987). Testing Behavioral Simulation Models by Direct Experiment. *Management Science* **33**: 1572-1592.
__________ (1989a). Misperceptions of Feedback in Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* **43**: 301-335.
__________ (1989b). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science* **35**: 321-339.
__________ (2008). Risk Communication on Climate: Mental Models and Mass Balance. *Science* **322**: 532-533.
Sterman, JD & Sweeney, LB (2002). Cloudy Skies: Assessing Public Understanding of Global Warming. *System Dynamics Review* **18**: 207-240.
__________ (2007). Understanding Public Complacency about Climate Change: Adults' Mental Models of Climate Change Violate Conservation of Matter. *Climatic Change* **80**: 213-238.
Sweeney, LB & Sterman, JD (2000). Bathtub Dynamics: Initial Results of a Systems Thinking Inventory. *System Dynamics Review* **16**: 249-286.
Tversky, A & Kahneman, D (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* **185**: 1124-1131.
van Ackere, A, Warren, K & Larsen, ER (1997). Maintaining Service Quality under Pressure from Investors: A Systems Dynamics Model as a Hands-On Learning Tool. *European Management Journal* **15**: 128-137.
Wu, DY & Katok, E (2006). Learning, Communication, and the Bullwhip Effect. *Journal of Operations Management* **24**: 839-850.

Young, SH, Chen, CP, Wang, S-W & Chen, CH (1997). The Landmine Structure and the Degrees of Freedom of Decision-Making. *Proceedings of the 15th International Conference of the System Dynamics Society*. Istanbul, Turkey.