Illuminating the Logic of Stock Management Failure – How Much Does the (Mis)Understanding of Accumulation Explain?

Jürgen Strohhecker

Frankfurt School of Finance and Management

Sonnemannstraße 9-11, 60314 Frankfurt am Main

Telephone: +49 69 154008-110

E-Mail: j.strohhecker@frankfurt-school.de

ABSTRACT

The stock management (SM) problem is of high relevance for a broad range of decision makers in society, business, and personal affairs. Although in some areas highly sophisticated models and control concepts have been developed, human stock management performance is lamentable. One recent explanation for this failure is offered by a stream of research, which finds evidence for widespread and persistent deficits in understanding how flows accumulate in stocks. This misunderstanding of accumulation (MoA) is proven even among well-educated adults. This research uses laboratory experiments to test the hypothesis that the better people understand accumulation, the higher is their performance in SM tasks. Correlation and univariate regression analysis show that MoA indeed contributes to explaining performance differences in stock management. However, the effect is moderate and vanishes almost completely when intelligence and economic knowledge are included as control variables in a multiple regression model. The value of this paper lies in explicitly testing the relation of MoA and SM, whose existence is widely taken for granted. Future research could explore a broader set of control variables and should increase the number of cases to allow for advanced theory testing using, for example, structural equation modelling.

Key Words: Accumulation, Bathtub Dynamics, Dynamic decision making, Intelligence, Stocks and flows, Stock-flow failure, System Dynamics, Complex Systems

INTRODUCTION – THE CHALLENGE OF DYNAMIC DECISION MAKING

Stock management (SM) is of high relevance for a broad range of decision makers in society, business, and personal affairs. However, human decision making performance in such tasks is generally lamentable. Dynamic decision making research (Brehmer, 1992; Edwards, 1962) has accumulated ample evidence of decision making failures in dynamic complex systems. Such systems consist of stocks and flows and interrelating information links (Forrester, 1961). They are characterized by feedback and delays between cause and effect (Sterman, 1994). Human decision makers perceive these systems often as opaque, incomprehensible and hard to control (Dörner, 1996). And –

on average – they perform rather poorly in managing such systems. These results persist over a wide range of systems. For example, Dörner and colleagues found lamentable results of participants who were asked to act as mayor of the virtual small town "Lohhausen" (Dörner, 1980; Dörner *et al.*, 1994). Reichert and Dörner (1988) report failures when participants are charged with the task of manually controlling the temperature of a refrigerated warehouse. Sterman (1989a) finds average team costs ten times greater than the benchmark using the well-known beer game as an experimental device. In a new product management task a naïve benchmark policy outperformed the subjects in 87 % of the cases (Paich & Sterman, 1993a). Confronted with the challenge to manage a virgin fish stock, 74 % of the participants did overinvest in vessels resulting in a worse-than-optimal achievement of the overall target (Moxnes, 1998). Wittman and Hattrup (2004) report widely varying performance of subjects acting as managers of a tailor's shop, a coal-fired power plant and a high-technology company with a range of substituting products to develop and bring to the markets.

While a well-developed universal theory of dynamic decision making has not yet emerged, the various research efforts over more than two decades have contributed to a better understanding of the "logic of failure," as Dörner (1996) pithily named these phenomena. Although early studies did not find performance in the micro world experiments to correlate with tests of intelligence or personality (Brehmer, 1992; Dörner et al., 1994), more recent research using advanced intelligence structure tests could observe a significant, medium strong relationship between intelligence and performance on complex simulations (Süß, 1996; Wittmann & Hattrup, 2004). Following Cattell's (1963) investment theory and Ackerman's (1996) PPIK theory, these studies decompose intelligence in process and knowledge components.

Misunderstanding of feedback and misunderstanding of delays is another common explanatory pattern for poor human decision making in dynamically complex situations. Sterman (1989a, 1989b) shows that inventory management performance suffers systematically from misperceptions of the feedback structure of the system that has to be managed. Participants used inappropriate anchoring heuristics, misperceive time lags and use open-loop explanations of dynamics. These phenomena have been repeatedly re-observed and corroborated (Barlas & Özevin, 2004; Brehmer, 1992; Cronin & Gonzalez, 2007; Diehl & Sterman, 1995; Kleinmuntz, 1985; Moxnes, 1998; Paich & Sterman, 1993b; Rahmandad *et al.*, 2009).

Recent research has identified an additional potential explanatory factor for poor SM performance – misunderstanding of accumulation (MoA). The seminal study of Booth Sweeney and Sterman (2000) has revealed that a large fraction of highly educated people is unable to infer the behaviour of even the simplest stock-flow-systems consisting of only one stock, one inflow, and one outflow. As no feedback, no time delays, or nonlinearities were incorporated in those simplistic systems, they cannot be characterized as dynamically complex. Nevertheless, the average understanding of these systems' dynamic is lamentable. The subjects showed a rather poor performance in a variety of paper-and-pencil tasks involving such systems, which supports the conclusion that human beings indeed have a poor understanding of accumulation. Subsequent studies by Ossimitz (2002), Sterman and Booth Sweeney (2002, 2007), Cronin and Gonzales (2007) corroborate the conjecture that the misunderstanding of accumulation is a persistent phenomenon, comparable to the deep-rooted problems people have in probabilistic judgements and decision making (Hastie & Dawes, 2001; Kahneman & Tversky, 1972).

This study attempts to contribute to dynamic decision making research by explicitly formulating and testing the hypothesis that understanding accumulation correlates with SM performance. This hypothesis is put to the test by collecting data from two observations in a laboratory experimental setting. The first observation assesses the participants' understanding of accumulation by using a collection of stock and flow tasks that have been developed and employed in previous research. The second observation pursues the purpose of measuring SM performance following the experimental paradigm for investigating dynamic decision making suggested by Brehmer (1992). A dynamic, yet simple computer-based inventory management game with one product, a constant lead time, and costs for ordering, inventory keeping, stock outs, and lost or gained sales is used in a laboratory experiment. SM performance is defined as cumulated total (opportunity) costs, which the participants have to minimize.

The paper continues in Section 2 with a description of the hypothesis to be tested and the research method used. Section 3 describes how the SFT ability is measured and outlines the results. Section 4 provides details on the inventory management task and the assessment of the subjects' performance. Section 5 presents the results of the hypothesis test. The paper concludes with a discussion of limitations and contributions of this research and outlines directions for further research.

HYPOTHESIS AND EXPERIMENTAL DESIGN

Prior work has revealed that people perform rather badly in both rather complex (Croson & Donohue, 2003, 2006; Sterman, 1989a, 1989b) and rather simple SM tasks (Bloomfield *et al.*, 2007). In searching for the simplest dynamic task that people can cope with, Sterman and others developed paper-and-pencil tasks based upon the simplest system possible with one inflow, one stock, and one outflow, with no feedback, time delays, and non-linearity, and found that even well educated subjects still struggle with the understanding of stocks and flows (Booth Sweeney & Sterman, 2000; Cronin *et al.*, 2009; Ossimitz, 2002; Sterman, 2010; Sterman & Booth Sweeney, 2002, 2007). Cronin et al. (2009) find that MoA persists regardless of whether the data are displayed in line graphs, bar graphs, tables, or text; poor performance is robust to changes in the cover story that frames the task and provides a specific context, for example the management of a stock of cash or the amount of water in a bathtub; it is also robust to situations that involve discrete entities or continuously varying quantities; even reducing the task complexity by decreasing the number of data points presented does not increase the understanding of stocks and flows. Cronin et al. (2009) point out that stock and flow thinking capabilities obviously suffer from important and pervasive shortcomings in human reasoning. The authors conclude: A high percentage of people seriously misunderstands "the basic principles of accumulation" (Cronin *et al.*, 2009, p. 128).

While the direct investigation of a connection between MoA and SM performance is not in the focus of the research cited above, the conjecture that MoA causes poor SM performance is inherent in the studies' discussion. Already in their seminal study Booth

Sweeney and Sterman (2000) argue that MoA might lie at the root of people's bad performance in dynamically complex environments. In their 2002 article, the authors hypothesize "that much of the complacency about climate change arises from poor systems thinking skills" (Sterman & Booth Sweeney, 2002). And they conclude: "The sooner people understand these dynamics the sooner they will call for leaders who reject do-nothing wait and see policies and turn down the tap—before the tub overflows" (p. 236). Finally, a decisive relation between MoA and SM performance is stated by Cronin et al. (2009, p. 128): "Effective decision making in dynamic settings requires decision makers to understand accumulation." The objective of this research is to contribute to the literature by formulating and testing the following hypothesis:

H1. The better people understand accumulation, the better they perform in managing a dynamic system with stocks and flows.

For testing H1, an experimental research design with two observations and no treatment was deemed appropriate (Trochim & Donnelly, 2007). By the first observation SM performance is measured following the well-established experimental paradigm for investigating dynamic decision making (Brehmer, 1992) using a computer simulated micro world. The second observation's purpose is to come up with a measure for understanding of accumulation (UoA). For this second construct instruments and methods are employed which have already been applied in prior work. Therefore, a specific UoA inventory was compiled using a number of rather simple paper-and-pencil tasks developed by Booth Sweeney and Sterman (2000), Sterman (2002), and Ossimitz (2002). More details on this inventory and information on the outcomes are provided in the following paragraph.

The micro world used in the experiment was a specifically developed inventory management game of much less detail and dynamic complexity than, for example, the beer game used by Sterman (1989a, 1989b) or the Lohhausen simulator used by Dörner et al. (1994). It builds on inventory management games that were developed long ago and have been used as educational instruments in practice and academia for many years (e.g., Renshaw & Heuston, 1957). According to the game's cover story, participants act as production controller in a pump body manufacturing company. They are responsible for one single product and have to decide on the weekly production start rate. Production throughput time is three weeks; in week four, the batch of end products is put into storage, from which they could be delivered to customers according to their needs. As the company produces a range of different pump bodies, batch production is used. The maximum batch size is restricted to 600 units. Smaller batch sizes are possible but do not reduce set-up costs. Of course, more than one batch of a specific pump body can be produced in any week.

The objective is to minimize total costs, which include in addition to set-up cost, inventory holding cost, stock-out costs and lost or gained contribution margins. Stock-out costs account for late deliveries of customer orders that could not be served immediately. These orders are not lost but have to be delivered subsequently. However, if customers experience bad service levels, they place more orders with competing suppliers causing the incoming orders to decrease. An exceptionally good service quality may also lead to increased orders. To account for this, for each lost ordered unit the product's contribution margin is added to total costs, while for each additional unit sold the contribution margin is subtracted from total costs. As a result, total cumulated costs may even become negative, meaning that costs have been offset by additionally gained contribution margins due to extraordinary service quality.

The participants have information on all cost parameters and lead times. They know that customer demand is initially 1000 units per week, not influenced by any random effects and only sensitive to the service level. However, they do not know the exact functional relationship. To allow for dynamic behaviour to unfold over time, the game is started in disequilibrium with empty work in progress (WIP) inventory, 4,000 units of inventory and no order backlog. It is run over a time span of 26 simulated weeks, and therefore, participants have to make 26 decisions. The micro world is developed using Forio Simulate, which provides a flexible modelling language, allows designing modern user interfaces (see Figure 1) and supports web-based gaming and administration. Once a decision is entered and the button "continue simulation" is pressed, the decision outcomes are calculated, and the information displayed on the screen is updated (Figure 1). At the end of the simulation the participants are informed about their overall performance in the game.

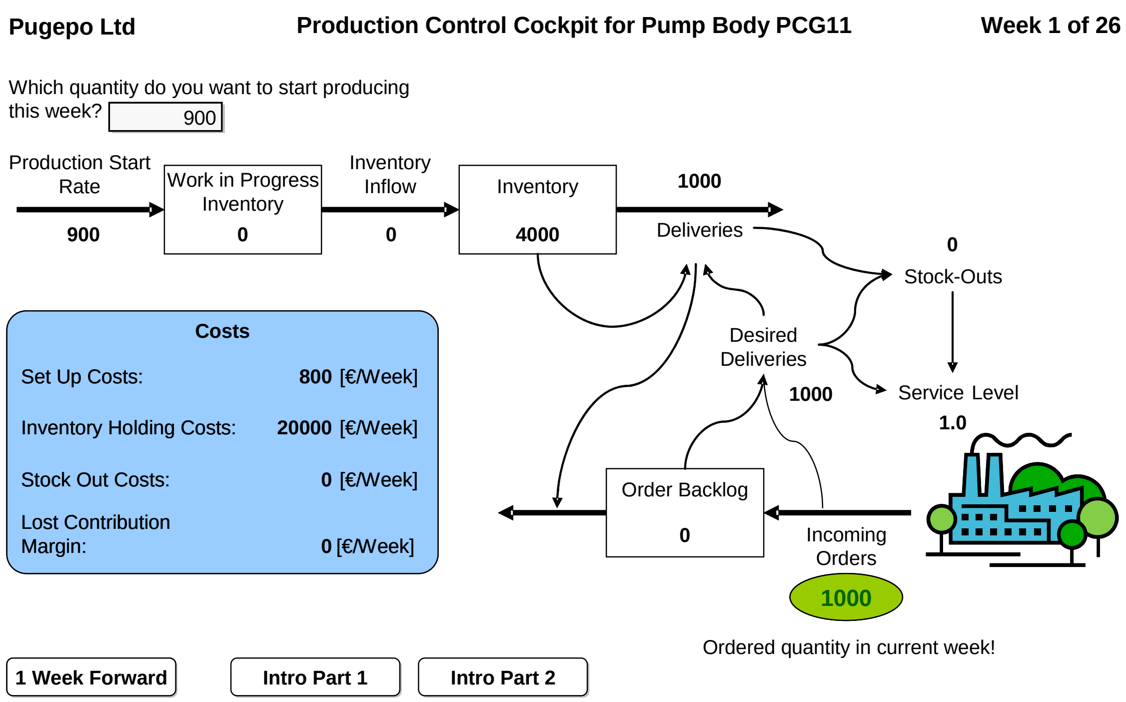


Figure 1: Production and inventory control cockpit

The game is thoroughly introduced: Participants receive both a verbal briefing and on-screen instructions, which precisely describe the setting, the task, and the objective. The participants can go back to this information at any time throughout the game. To minimize the danger of "video-gaming" and exclude learning biases from multiple iterations, only one simulation run is allowed. To allow for thorough reasoning time pressure is kept as low as possible. No explicit deadline for terminating the simulation is set. As the experiment is integrated into a standard 180-minute lecture unit, an implicit end time exists, though.

For testing H1, data were primarily gathered from a laboratory experiment conducted at a German business school. In addition to this, archival data were retrieved from the school's databases allowing controlling for additional influencing factors such as gender, intelligence or knowledge. The experimental sessions were integrated into an elective course on "Operations Management" that is part of the Bachelor of Business Administration program. In September 2009 and September and November 2010, five experimental sessions were performed, involving as participants in total 79 students in their seventh and final semester. The game was played first and took about 20 minutes. The UoA inventory was filled in the second. The participants were allowed to leave afterwards. The average time spent on the inventory was 31 minutes.

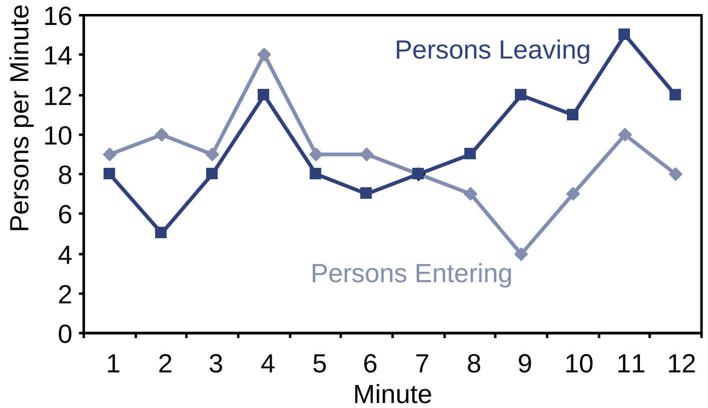
While for some disciplines and decision tasks empirical evidence for the appropriateness (Depositario *et al.*, 2009; Elliott *et al.*, 2007) or inappropriateness (Fleming, 1969; Peterson, 2001; Vinson & Lundstrom, 1978) of students as experimental subjects exist, I am not aware of any investigation on this issue for the stock management task central to our research. Thus, it is assumed that the participants are similar to real world decision makers in terms of general personal characteristics (like intelligence or personality) and basic education received; obviously, student participants lack the level of experience that managers have accumulated. However, using students as participants has the advantage that the results can be compared to previous studies that have mostly also relied upon students.

The experimental design follows suggestions from experimental economics (e.g. Friedman *et al.*, 2004; Guala, 2005) as well as experimental psychology (e.g. Kantowitz *et al.*, 2009). Applying Smith's (1976, 1982) induced value theory participants are not only motivated by an informative individual analysis of the test results but also incentivized by a monetary reward. The financial incentive was linked to both the participant's performance in the inventory management game and the UoA inventory. For the game, the lower the cumulated total costs the more money was paid out up to the maximum of 9.50 €. For the test, the cash-out was calculated according to the percentage of correct answers with a maximum of 10 €. On average, 9.68 € were achieved by the participants for an exercise of about 50 minutes.

ASSESSMENT OF UNDERSTANDING OF ACCUMULATION

For assessing UoA ability, five relatively simple paper-and-pencil tasks are compiled that have already been used in prior studies in an identical or very similar form. Each task is designed to measure the participants' understanding of stocks and flows and their ability to infer their behaviour over time. The type of the tasks ranged from sketching behaviour over time patterns, reading and interpretation of line graphs to multiple choice questions.

The first task is taken from Kainz and Ossimitz (2002) and referred to as a rainwater tank (RWT) task. The second task is adapted from the department store task developed by Sterman (2002) and illustrated in Figure 2. The third task intends to test whether the participants are aware of the difference between the net flow "budget deficit" and the stock "national debt." It is adapted from Ossimitz (2002) and referred to as a budget deficit (BD) problem. No graphical presentation of information is given or required in this task. Instead, it consists of five multiple choice questions, which have to be answered by checking one of four possible answers. Task number four and five are taken from Booth Sweeney & Sterman (2000). The fourth task is the so-called manufacturing case (MC). The fifth and last task in the UoA inventory is the bath tub (BT) task.



During which minute were the most people in the bank branch?
During which minute were the fewest people in the bank branch?

Figure 2: Illustration of the bank branch UoA task

The five tasks of the UoA inventory include, all in all, 14 subtasks. Subtasks are assessed on a right (1) or wrong (0) basis. Based on these results a percentage of correct answers is calculated for each one of the five main tasks. Finally, a UoA score is determined as the average of these five percentage values. Appendix 1 displays bivariate Pearson correlations for the five main tasks. The consistently moderate correlations between the participants' results on the tasks indicate that indeed a variety of aspects regarding the construct UoA is covered. Correlations for all 14 subtasks are provided in Appendix 2. These too confirm the broadness of the UoA measure. Nevertheless, based on this 14-item scale, reliability is acceptable (Cronbach's α = .760).

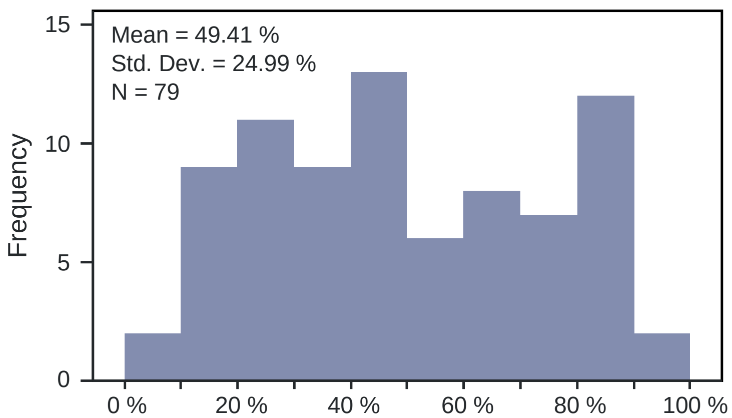


Figure 3: Histogram of the participant's UoA score

Figure 3 provides a histogram for the results of the UoA test filled in by 79 participants involved in the study. With a mean value of .4940, a standard deviation of .2499, a minimum performance of .0667 and a maximum of .9333, the participants' understanding of accumulation has to be rated as bad. Considering the rather low level of difficulty in the tasks, the result is devastating. However, this study adds to the pool of bad results found by previous work (Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Cronin et al., 2009; Kainz & Ossimitz, 2002; Ossimitz, 2002; Sterman, 2002; Sterman & Booth Sweeney, 2002, 2007). Once more, it demonstrates a profound and notable shortcoming in human reasoning: the inability of even smart and well-educated people to understand the dynamic relationships between stocks and flows, that is, the process how flows into and out of a stock accumulate over time. Cronin et al. (2009) demonstrate that poor SFT performance persists regardless of the cover story, the display format of the data, and the quantity of information provided. They reveal that learning is slow when tasks can be done repeatedly and outcome feedback is provided. Moreover, they show that modest incentives do not improve performance.

RESULTS FROM THE INVENTORY MANAGEMENT GAME

Following Sterman (1989a, 1989b), Süß (1996), and many others, this study uses total accumulated (opportunity) costs in a dynamic simulation game as the measure for SM performance. Inherent in the game's design is that increasing the production order rate also increases inventory costs and set-up costs (step-fixed) yet decreases stock-out costs and lost contribution margin. More frequent production orders increase set-up costs but decrease inventory costs. Consequently, total accumulated costs can be seen as a balanced measure for decision quality in the inventory game. From a theoretical financial perspective, the net present value of costs would be the most advisable performance measure to choose. However, as participants' subjective time preferences often do not follow the economic rationality paradigm (al-Nowaihi & Dhami, 2006; Loewenstein & Prelec, 1992), the situation for the participants is simplified and undiscounted cumulative costs as a performance measure are used.

Although the complexity of the simulator used is relatively moderate, Figure 4 illustrates that the system still can be characterized as dynamically complex. It consists of four stocks. Two negative feedback loops control deliveries and incoming orders. A first order perception smoothing structure delays the effect of delivery quality on incoming customer orders. Production and delivery lead times introduce further delays and require the participants to look ahead. Before providing an overview of the participants' performance in the inventory game, a benchmark strategy and benchmark results are introduced and a few exemplary dynamics created by participants are shown.

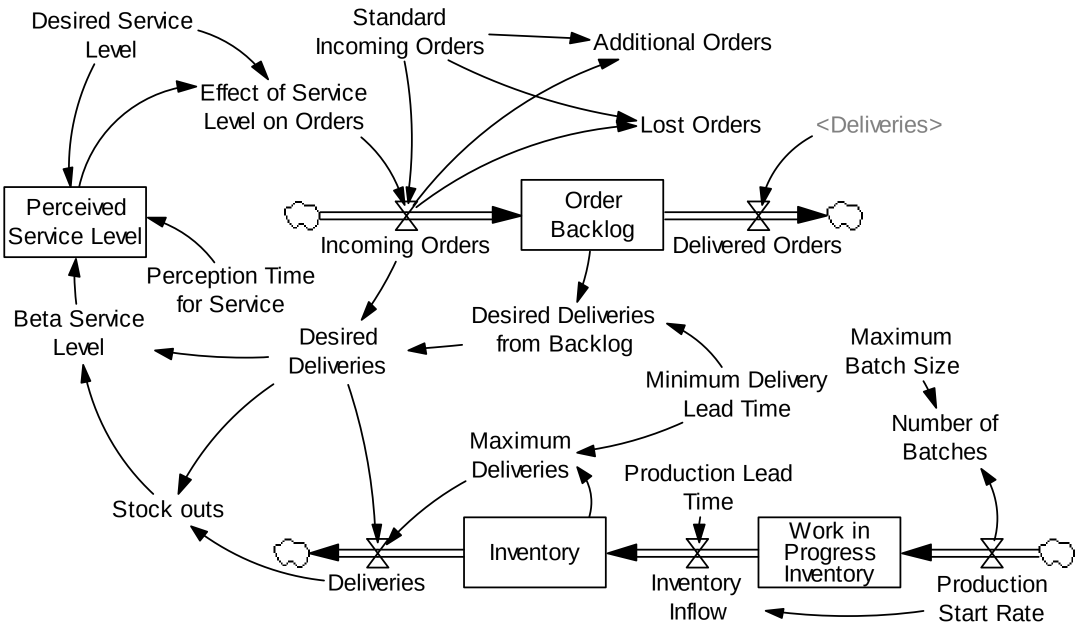


Figure 4: Core stock and flow structure of the inventory management game

A "fair" benchmark decision rule should be designed so that it could, at least in principle, be followed by bounded rational, human decision makers – at least if the principles of stocks, flows and accumulation are well understood. In particular, it should not violate the Baker Criterion (Sterman, 2000, p. 517) meaning that it should not be based on hindsight. An appropriate decision-making rule is suggested by Sterman (1989a, 1989b; 2000) and illustrated in Figure 5. It is based on incoming orders, which are converted to production orders by additionally accounting for adjustments to the inventory and work in progress inventory levels. These adjustments are derived from gradually closing the gaps between two stocks, for instance, inventory and desired inventory. It is obvious from the stock and flow diagram provided in Figure 5, that an understanding of accumulation is required to design and follow the decision rule suggested by Sterman for stock management tasks.

Using policy parameter optimization (Coyle, 1985) for the Pugepo micro world case leads to a safety stock coverage of 0.156 weeks, an inventory adjustment time of 9.71 weeks and a WIP adjustment time of 4.30 weeks. Based on these parameter values, the decision rule translates into a production start rate shown in Figure 7 and results in a stock management performance measure of total cumulated opportunity costs equal to -8,540 € by week 26.

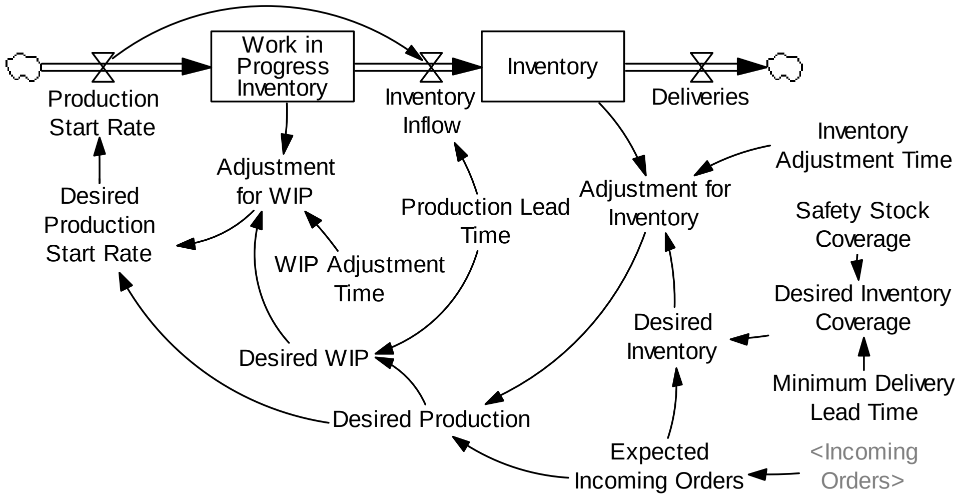


Figure 5: Decision making rule for the benchmark policy

Behaviour-over-time charts for the benchmark scenario are provided by Figure 6, Figure 7, and Figure 8 respectively. It has to be highlighted again that the simulation is deliberately started in disequilibrium. Desired WIP is therefore clearly higher than actual WIP resulting in production orders above incoming orders for weeks one to three. Consequently the WIP inventory is quickly filled up to the desired level, while at the same time the inventory level of initially 4000 units is reduced to the level of slightly more than one week's demand. Service level is kept at its maximum. The sharp decline in service quality in week four cannot be avoided. With an initial inventory of 4000 units, a four-week lead time and a demand that rises above the initial level of 1000 units per week due to the 100 % service level provided in weeks one to three, stock outs in week four are inevitable. So is the drop in demand in week five, as customers react to bad service quality. However, with a sufficient amount of inventory at hand, service level can be restored to the maximum from week five on, and demand slowly recovers. From week 22 onwards it has reached its equilibrium level of 1105 units per week. As the simulation ends in week 26 and no costs beyond that time are included in the performance measure, it is optimal to cut back the production start rate to zero for the last four weeks.

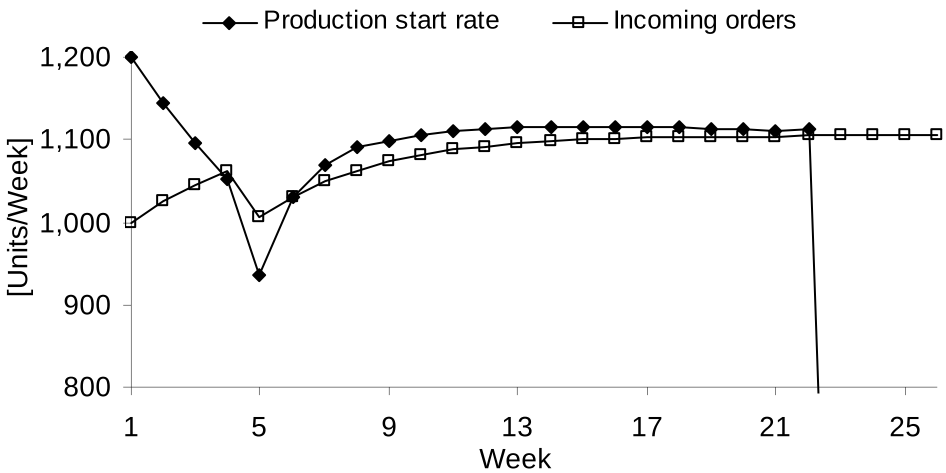


Figure 6: Benchmark production start rate and incoming orders

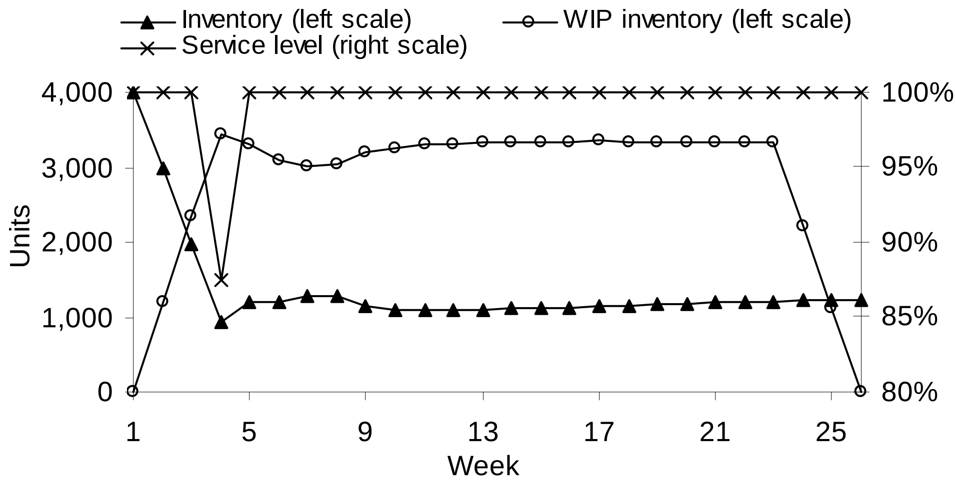


Figure 7: Benchmark inventory, WIP inventory and service level behaviour

The cost curves for the benchmark scenario are shown in Figure 8. Stock-out costs can almost completely be avoided; inventory holding costs decrease rapidly as the initially high inventory level is reduced. From week five onwards, they fluctuate slightly and remain nearly constant after week 12. As from week 1 to 22 two batches are produced, set-up costs amount to 800 € per week for this time period. Contribution margin balance is always negative. It is calculated as follows: lost contribution margin (due to bad service quality) minus additional contribution margin (due to good service quality). As a result of good service delivery the demand is always greater than the initial 1000 units per week, resulting in additional contribution margin being consistently larger than lost contribution margin. Total costs accumulate in the benchmark scenario to -8450 € indicating that the additional contribution margin more than compensates inventory holding, stock out and set-up costs.

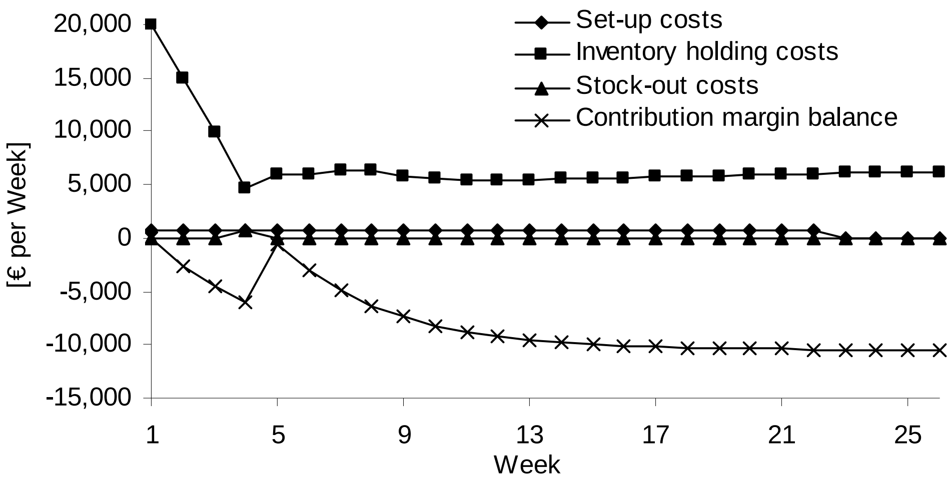


Figure 8: Benchmark cost curves

The benchmark decision rule formulated above compares favorably with the five basic principles for modelling human decision making postulated by Sterman (Sterman, 2000, p. 516-520). Therefore, it could be followed by the participants. But only very few come close. As an example, Figure 9 shows the sequences of decisions over the 26 weeks for three participants. By having a steady hand and avoiding hectic up and down movements in the production start rate, the very good participants shown achieve total cumulated costs of 5,050 €, which is not far from the benchmark. That a frenzy of activity does lead to much worse results makes the second (the average) participant obvious. A result of 208,970 € of cumulated costs is the consequence of a high variability in production orders even towards the end of the game. The third example of a very poorly performing participant shows that she, or he has not grasped the huge damage that consistently producing too few units does to the customers' satisfaction and their ordering. This decision making shows that the demand erosion feedback loop, which is fuelled by bad delivery quality, is completely misunderstood.

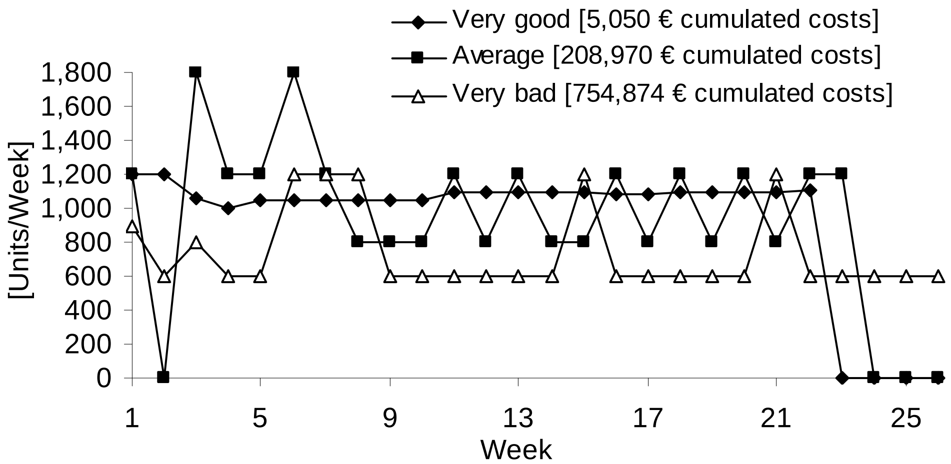


Figure 9: A comparison of decision sequences on the production start rate

Based on 72 completed and usable simulation runs, Figure 10 shows the – clearly non-normal – distribution of the participants' SM performance (SMP). Descriptive statistics are provided in Appendix 3. When comparing these results to the benchmark, the conclusion is at hand that the logic of failure of human decision makers is once more impressively demonstrated. This inference is supported by comparing the participants' results to the outcome of one of the simplest decision rules that one can come up with: Just placing the incoming orders as production orders – clearly a "no-brainer" policy – would result in accumulated total costs of 146,905 €. Only 39 out of 72 participants are able to outperform the no-brainer policy. 33 participants do even worse – with the poorest result being more than five times higher than the no-brainer outcome. Another extremely simple and obviously inadequate decision rule would be to keep production orders constant at the initial value of 1,000 units per week from week 1 to week 22. Even this "Do-nothing" decision rule would outperform 27 participants who manage to accumulate more than 180,110 € in total costs.

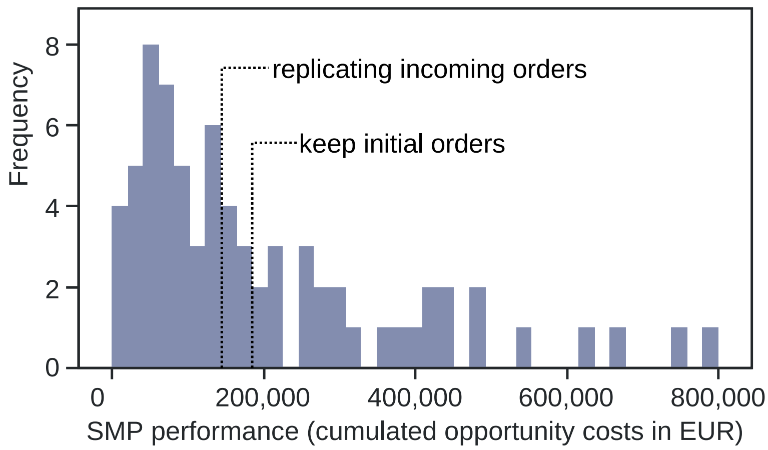


Figure 10: SM performance distribution

Although the computer based micro world used in this study is significantly less complex than previously used simulators, the difficulties of human decision makers in dynamically complex environments can be reproduced. Most of the explanations discussed in the literature could probably be applied too: misperceptions of time lags (Sterman, 1989a, 1989b), misperceptions of feedback and open-loop thinking (Brehmer, 1992; Moxnes, 1998; Sterman, 1989b), anchoring (Dörner, 1996; Tversky & Kahneman, 1974), defensive routines and ballistic behaviour (Dörner, 1996; Sterman, 2000), etc. However, if and how (mis)understanding of accumulation and (mis)performance in stock management are related is for the remainder of this study of special interest and will be investigated in the following paragraph.

TESTING THE HYPOTHESIS

Hypothesis H1 that was stated above positively relates the understanding of accumulate to the performance in managing a stock and flow system. It is operationalised as follows:

H1o: The higher a decision maker's UoA score, the lower is his or her SM performance score (that measures total cumulated opportunity costs).

First, the hypothesis is tested using univariate regression analysis and non-parametric correlation analysis. The standardized regression coefficient β is negative as postulated and indicates a moderate relationship (-.312, p = .008). Consequently, the hypothesis cannot be rejected. However, as Figure 11 illustrates, only a small proportion of the variance can be explained, and R^2 is low (.098). As neither SMP nor UoA is normally distributed, the nonparametric Spearman Rank Order Correlation Coefficient is also calculated. With ρ = -.314 (p=.007) again a moderate and significant relationship is found giving no reason to reject H1.

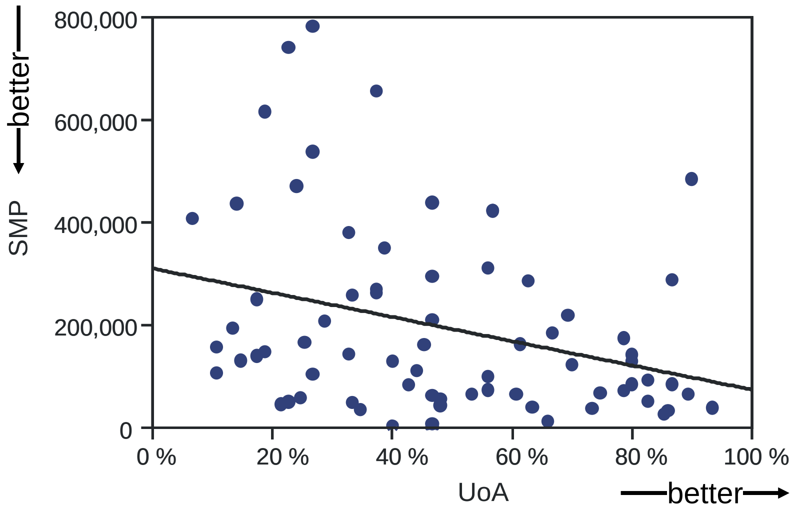


Figure 11: Scatter plot showing the relationship between UoA and SMP

Second, a multiple regression model is formulated and tested to control for further effects investigated by previous dynamic decision making research. Based upon Ackerman's (1996), Süß' (1996) and Wittmann & Hattrup's (2004) findings, general cognitive ability (G), knowledge and gender are also included in the regression model. As the participants are students, for these three predictors archival data could be retrieved from the school's databases. Applicants have to pass through an entrance assessment centre, which includes completion of the BIS test short form (Jäger *et al.*, 1997). For 63 participants the BIS general intelligence score was available. Admittedly, these data are two to three years older than the data gathered in the laboratory. However, psychological research shows that general cognitive ability is rather stable over long time periods (Larsen *et al.*, 2008; Lyons *et al.*, 2009), which seems to justify the use of these data. Following Ackerman's (1996) theory, knowledge is operationalised as economic knowledge (EKS), which is measured averaging the participants' percentage achievements in typical economic and business administration courses.[1]

[1] The following course grades are included: Investment and finance, management accounting, decision analysis, organisational behaviour, microeconomics, bookkeeping, national accounts.

Model		Unstandardized Coefficients		Std Coefficients			Collinearity Statistics	
Dep.	Predictors	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
SMP	(Constant)	612,762	160,982		3.806	.000		
	UoA	-77,327	124,753	-.099	-.620	.538	.554	1.806
	G	-36,540	18,123	-.263	-2.016	.048	.833	1.201
	EKS	-334,856	219,254	-.202	-1.527	.132	.807	1.239
	GENDER	1,366	59,020	.003	.023	.982	.739	1.354

Table 1: Multiple regression results (N = 63, R^2 = .121)

The results for the extended regression model are compiled in Table 1. As it is obvious that gender is not significant in predicting SMP, this factor is excluded and the regression is updated (Table 2). Still, the result changes considerably: Compared to the univariate model, UoA shows a greatly reduced and non-significant impact on SMP in the multivariate regression model. Obviously, due to missing values for G and EKS, the number of cases is reduced. However, a repeated univariate regression based on the same 63 cases does result in an only slightly reduced beta coefficient (β = -.279, p = 0.027, R^2 = 0.078). Therefore, the main effect has to be attributed to including G and EKS as control variables.

Model		Unstandardized Coefficients		Std Coefficients			Collinearity Statistics	
Dep.	Predictors	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
SMP	(Constant)	613,969	151,006		4.066	.000		
	UoA	-75,856	106,441	-.097	-.713	.479	.748	1.337
	G	-36,604	17,760	-.264	-2.061	.044	.852	1.173
	EKS	-335,770	213,838	-.203	-1.570	.122	.834	1.199

Table 2: Multiple regression results (N = 63, R^2 = .178)

The collinearity statistics included in Table 2 does not show variance inflation factors (VIF) near or greater than 10, which should evoke concern (Myers, 1990). However, according to Bowerman and O'Connell (1990), an average VIF of greater than one indicates that multicollinearity may be biasing the regression model. Although the correlation matrix provided in Appendix 4 does not show problematic correlations of above .80 between the predictor variables, UoA significantly correlates with both EKS and G. Considering these results, it cannot be excluded that multicollinearity indeed poses a problem and the interpretation and generalisation of the results has to be done with care.

In summing up the results from statistically testing the hypothesis H1, two main findings can be stated: First, univariate analysis finds a significant moderate relationship between UoA and SM performance; H1 can clearly not be rejected. Second, multivariate analysis produces somewhat blurred results. The relationship between UoA and SMP is diminished in its strength and turned insignificant when intelligence and economic knowledge are included as control variables. Consequently, H1 had to be rejected. However, a certain degree of multicollinearity and the violation of the multi-normal distribution assumption cast a shadow on the persistence of this conclusion.

DISCUSSION, LIMITATIONS AND CONCLUSIONS

In our modern, vastly and globally connected world, coping with dynamic decision making challenges our daily routine. Increasing the understanding of human key performance drivers in such situations is theoretically and practically important. Initiated and influenced by prior research on stock and flow thinking capabilities, this study focuses on the issue whether poor understanding of accumulation indeed predicts poor performance in managing a dynamic stock and flow system. The hypothesis that this relationship does exist can be derived from discussions by, for example, Cronin et al. (2009), Sterman (2002), or Booth Sweeney & Sterman (2000). Those studies, however, primarily attempt to measure UoA performance and try to provide explanations for the poor results – for instance, the widespread use of a simple correlation heuristic (Cronin et al., 2009). They do not investigate the consequences of poor UoA on the stock management abilities. This research seeks to fill this research gap by conducting a laboratory experiment and statistically testing the hypothetical relationship.

Experimental results reported here corroborate human deficits in reasoning and dynamic decision making: first, the human inability to correctly understand the relation between flows and stocks over time, and second, our incapability to just come close to the performance of heuristic decision rules in a dynamically complex system. Regarding the study's main objective, that is the investigation and testing of the relationship between (mis)understanding of accumulation and (mis)management of dynamic stock and flow systems, first evidence could be assembled, yet results have to be regarded as preliminary. While correlation and univariate regression analysis find a significant relation of moderate strength, more sophisticated regressions models with intelligence and knowledge included as control variables cast doubt on the importance of UoA as a predictor for SMP. Due to some missing data for the control variables, N decreases from 72 to 63. Therefore, the most important limitation of this study may be seen in an insufficient sample size. To address this issue, additional experiments are being scheduled to increase the number of cases (and hopefully at the same time decrease the problems with violation of statistical assumptions).

Additionally, a larger number of cases would allow for applying more advanced research instruments – for example, structural equation modelling (SEM) techniques that require $N \geq 100$. While multicollinearity still can obscure SEM based theory testing, for moderate levels errors tend to be small (Grewal *et al.*, 2004). Further advantages of using SEM include the ability to construct latent variables, cover direct and indirect relationships and distinguish between the structural model and the measurement model (e.g. Montfort *et al.*, 2004).

This study does not necessitate a fundamental alteration of the research design. The inventory used for measuring the UoA performance seems to be appropriate. Earlier results could be confirmed. Based on, for example, classical test theory or item response theory, future work could investigate the reliability and validity of the UoA inventory used in this and previous research. Such attempts would contribute to the discussion on the adequacy of the paper-and-pencil tasks proposed in the literature (and also used in this study) for truly measuring understanding of stocks and flows. Although the inventory management micro world used here was newly developed, no obvious shortcomings could be detected while conducting the study and analysing its outcomes. The cost-minimising objective was sufficiently simple for the participants to grasp. The task – deciding on a production start rate for 26 weeks in a row – was well understood, at least as one could conclude from the rather few questions participants asked after the introduction. Obviously, with just this one micro world used in the experiment the existence of a framing effect (Frisch, 1993; Maule & Villejoubert, 2007) cannot be excluded. Therefore, future research could aim for replicating (and extending) the results using micro worlds of similar complexity but with different context.

This research's novelty lies in the focus on (mis)understanding of accumulation as a factor in (mis)management of stocks. It can be seen as a first attempt to determine the contribution of UoA to predict decision making performance in stock management tasks. Further research should help to build and test a more elaborate theory. Psychological constructs as personality and interests could be integrated as well as dynamic decision making traits: misperception of feedback, delays, uncertainty, etc. Since such theories are more and more complex, testing becomes increasingly demanding. However, application of more sophisticated theories to various domains is particularly promising. Based on those theories, improved methods could be developed to, firstly, support decision making in dynamic stock and flow systems and, secondly, to select the best people for the most demanding stock and flow management tasks. Last but not least, attempts to educate all of us to reach higher levels of performance in stock and flow problems could be fostered.

APPENDIX

		PRWT	PBB	PBD	PMC	PBT
PRWT	Pearson Correlation	1	.262	.028	.129	.259
	Sig. (2-tailed)		.020	.808	.259	.021
PBB	Pearson Correlation	.262	1	.230	.215	.283
	Sig. (2-tailed)	.020		.041	.057	.011
PBD	Pearson Correlation	.028	.230	1	.092	.302
	Sig. (2-tailed)	.808	.041		.418	.007
PMC	Pearson Correlation	.129	.215	.092	1	.364
	Sig. (2-tailed)	.259	.057	.418		.001
PBT	Pearson Correlation	.259	.283	.302	.364	1
	Sig. (2-tailed)	.021	.011	.007	.001	

Appendix 1: Bivariate correlations for main task performance (N = 79)

		PRWT1	PRWT2	PRWT3	PBB3	PBB4	PBD1	PBD2	PBD3	PBD4	PBD5	PMC1	PMC2	PMC3	PBT
PRWT1	Pearson Correlation	1	.656	.161	.039	.258	-.051	-.064	.210	-.127	-.004	.047	.063	-.008	.214
	Sig. (2-tailed)		.000	.155	.732	.022	.655	.576	.063	.266	.973	.680	.581	.942	.059
PRWT2	Pearson Correlation	.656	1	.207	.123	.466	.106	.015	.230	-.198	.069	.063	.116	.108	.271
	Sig. (2-tailed)	.000		.068	.278	.000	.352	.895	.041	.080	.545	.580	.310	.346	.016
PRWT3	Pearson Correlation	.161	.207	1	.026	.070	-.085	.048	.212	.010	-.049	.099	.218	.085	.063
	Sig. (2-tailed)	.155	.068		.823	.537	.458	.676	.060	.927	.668	.385	.053	.458	.579
PBB3	Pearson Correlation	.039	.123	.026	1	.633	.137	.153	.235	.142	.101	.046	.222	.118	.202
	Sig. (2-tailed)	.732	.278	.823		.000	.230	.179	.037	.210	.375	.688	.049	.301	.074
PBB4	Pearson Correlation	.258	.466	.070	.633	1	.169	.157	.208	.076	.157	.062	.200	.289	.310
	Sig. (2-tailed)	.022	.000	.537	.000		.137	.168	.066	.505	.168	.585	.078	.010	.005
PBD_1	Pearson Correlation	-.051	.106	-.085	.137	.169	1	.674	.397	.380	.417	-.027	.050	.166	.247
	Sig. (2-tailed)	.655	.352	.458	.230	.137		.000	.000	.001	.000	.816	.662	.143	.028
PBD_2	Pearson Correlation	-.064	.015	.048	.153	.157	.674	1	.403	.417	.323	.047	.113	.045	.290
	Sig. (2-tailed)	.576	.895	.676	.179	.168	.000		.000	.000	.004	.681	.323	.695	.009
PBD_3	Pearson Correlation	.210	.230	.212	.235	.208	.397	.403	1	.334	.403	-.116	-.008	.107	.259
	Sig. (2-tailed)	.063	.041	.060	.037	.066	.000	.000		.003	.000	.308	.942	.348	.021
PBD_4	Pearson Correlation	-.127	-.198	.010	.142	.076	.380	.417	.334	1	.307	-.045	.010	.001	.059
	Sig. (2-tailed)	.266	.080	.927	.210	.505	.001	.000	.003		.006	.696	.930	.990	.606
PBD_5	Pearson Correlation	-.004	.069	-.049	.101	.157	.417	.323	.403	.307	1	.102	.059	.250	.239
	Sig. (2-tailed)	.973	.545	.668	.375	.168	.000	.004	.000	.006		.373	.604	.026	.034
PMC1	Pearson Correlation	.047	.063	.099	.046	.062	-.027	.047	-.116	-.045	.102	1	.752	.350	.297
	Sig. (2-tailed)	.680	.580	.385	.688	.585	.816	.681	.308	.696	.373		.000	.002	.008
PMC2	Pearson Correlation	.063	.116	.218	.222	.200	.050	.113	-.008	.010	.059	.752	1	.371	.378
	Sig. (2-tailed)	.581	.310	.053	.049	.078	.662	.323	.942	.930	.604	.000		.001	.001
PMC3	Pearson Correlation	-.008	.108	.085	.118	.289	.166	.045	.107	.001	.250	.350	.371	1	.212
	Sig. (2-tailed)	.942	.346	.458	.301	.010	.143	.695	.348	.990	.026	.002	.001		.060
PBT	Pearson Correlation	.214	.271	.063	.202	.310	.247	.290	.259	.059	.239	.297	.378	.212	1
	Sig. (2-tailed)	.059	.016	.579	.074	.005	.028	.009	.021	.606	.034	.008	.001	.060	

Appendix 2: Bivariate correlations for the 14 UoA scale items

	N	Minimum	Maximum	Mean	Std. Dev.	Skewness		Kurtosis	
						Statistic	Std. Error	Statistic	Std. Error
SMP	72	1,360	783,460	197,359.10	182,412.11	1.420	.283	1.568	.559

Appendix 3: Descriptive statistics for SM performance

		EKS	G	UoA	SMP
EKS	Pearson Correlation	1	.211	.402	-.298
	Sig. (2-tailed)		.096	.001	.018
G	Pearson Correlation	.211	1	.379	-.343
	Sig. (2-tailed)	.096		.002	.006
UoA	Pearson Correlation	.402	.379	1	-.279
	Sig. (2-tailed)	.001	.002		.027
SMP	Pearson Correlation	-.298	-.343	-.279	1
	Sig. (2-tailed)	.018	.006	.027	

Appendix 4: Bivariate correlations (N = 63)

REFERENCES

Ackerman PL. 1996. A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence* **22**(2): 227-257

al-Nowaihi A, Dhami S. 2006. A note on the Loewenstein-Prelec theory of intertemporal choice. *Mathematical Social Sciences* **52**(1): 99-108

Barlas Y, Özevin MG. 2004. Analysis of stock management gaming experiments and alternative ordering formulations. *Systems Research and Behavioral Science* **21**(4): 439-470

Bloomfield RJ, Gino F, Kulp S. 2007. Behavioral causes of the bullwhip effect in a single echelon. SSRN

Booth Sweeney L, Sterman JD. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review* **16**(4): 249-294

Bowerman BL, O'Connell RT. 1990. *Linear statistical models : an applied approach* (2nd ed.). PWS-Kent Pub. Co.: Boston

Brehmer B. 1992. Dynamic decision making: human control of complex systems. *Acta Psychologica* **81**: 211-241

Cattell RB. 1963. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* **54**(1): 1-22

Coyle RG. 1985. The use of optimization methods for policy design in a system dynamics model. *System Dynamics Review* **1**(1): 81-91

Cronin M, Gonzalez C. 2007. Understanding the building blocks of system dynamics. *System Dynamics Review* **23**(1): 1-17

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**(1): 116-130

Croson R, Donohue K. 2003. Impact of POS data sharing on supply chain management: An experimental study. *Production and Operations Management* **12**(1): 1-11

Croson R, Donohue K. 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science* **52**(3): 323-336

Depositario DPT, Nayga RM, Wu X, Laude TP. 2009. Should students be used as subjects in experimental auctions? *Economics Letters* **102**(2): 122-124

Diehl E, Sterman JD. 1995. Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes* **62**(2): 198-215

Dörner D. 1980. On the difficulties people have in dealing with complexity. *Simulations and Games* **11**(1): 87-106

Dörner D. 1996. *The logic of failure. Strategic thinking for complex situations*. Metropolitan Books: New York

Dörner D, Kreuzig HW, Reither F. 1994. *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität* (Unveränd. Nachdr. der Ausg. von 1983 / mit einem Beitrag von Dietrich Dörner Zwölf Jahre danach: Lohhausen im Rückblick. ed.). Huber: Bern et al.

Edwards W. 1962. Dynamic decision theory and probabilistic information processing. *Human Factors* **4**(2): 59-73

Elliott WB, Hodge FD, Kennedy JJ, Pronk M. 2007. Are M.B.A. Students a Good Proxy for Nonprofessional Investors? *Accounting Review* **82**(1): 139-168

Fleming JE. 1969. Managers As Subjects In Business Decision Research. *Academy of Management Journal* **12**(1): 59-66

Forrester JW. 1961. *Industrial Dynamics*. Productivity Press: Cambridge MA

Friedman D, Cassar A, Selten R. 2004. *Economics lab: An intensive course in experimental economics*. Routledge: London

Frisch D. 1993. Reasons for Framing Effects. *Organizational Behavior & Human Decision Processes* **54**(3): 399

Grewal R, Cote JA, Baumgartner H. 2004. Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science* **23**(4): 519-529

Guala F. 2005. *The methodology of experimental economics*. Cambridge University Press: Cambridge

Hastie R, Dawes RM. 2001. *Rational choice in an uncertain world. The psychology of judgement and decision making*. Sage Publications: Thousand Oaks, Calif. ; London

Jäger AO, Süß H-M, Beauducel A. 1997. *Berliner Intelligenzstruktur-Test. BIS-Test. Form 4.* . Hogrefe: Göttingen et al.

Kahneman D, Tversky A. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* **3**(3): 430-454

Kainz D, Ossimitz G. 2002. Can students learn stock-flow-thinking? An empirical investigation. In SD Society (Ed.), *20th International Conference of the System Dynamics Society*: Palermo

Kantowitz BH, Roediger HL, Elmes DG. 2009. *Experimental psychology* (International student ed., 9th ed. ed.). Wadsworth Cengage Learning: Belmont, Calif.

Kleinmuntz DN. 1985. Cognitive heuristics and feedback in a dynamic decision environment. *Management Science* **31**(6): 680-702

Larsen L, Hartmann P, Nyborg H. 2008. The stability of general intelligence from early adulthood to middle-age. *Intelligence* **36**(1): 29-34

Loewenstein G, Prelec D. 1992. Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics* **107**(2): 573

Lyons MJ, York TP, Franz CE, Grant MD, Eaves LJ, Jacobson KC, Schaie KW, Panizzon MS, Boake C, Xian H, Toomey R, Eisen SA, Kremen WS. 2009. Genes Determine Stability and the Environment Determines Change in Cognitive Ability During 35 Years of Adulthood. *Psychological Science (Wiley-Blackwell)* **20**(9): 1146-1152

Maule J, Villejoubert G. 2007. What lies beneath: Reframing framing effects. *Thinking & Reasoning* **13**(1): 25-44

Montfort Kv, Oud J, Satorra A. 2004. *Recent developments on structural equation models : theory and applications*. Kluwer Academic: Dordrecht ; London

Moxnes E. 1998. Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics. *Management Science* **44**(9): 1234-1248

Myers RH. 1990. *Classical and modern regression with applications* (2nd ed.). PWS-KENT: Boston

Ossimitz G. 2002. Stock-flow-thinking and reading stock-flow-related graphs: An emperical investigation in dynamic thinking abilities. In SD Society (Ed.), *20th International Conference of the System Dynamics Society*: Palermo, Italy

Paich M, Sterman JD. 1993a. Boom, bust, and bailures to learn in experimental markets. *Management Science* **39**(12): 1439-1458

Paich M, Sterman JD. 1993b. Boom, bust, and failures to learn in experimental markets. *Management Science* **39**(12): 1439-1458

Peterson RA. 2001. On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research* **28**(3): 450-461

Rahmandad H, Repenning N, Sterman J. 2009. Effects of feedback delay on learning. *System Dynamics Review* **25**(4): 309-338

Reichert U, Dörner D. 1988. Heurismen beim Umgang mit einem "einfachen" dynamischen System. *Sprache und Kognition* **7**(1): 12-24

Renshaw JR, Heuston A. 1957. The game monopologs, *RAND Research Memorandum*:

Smith VL. 1976. Experimental Economics: Induced Value Theory. *American Economic Review* **66**(2): 274

Smith VL. 1982. Microeconomic System as an Experimental Science. *American Economic Review* **72**(5): 923

Sterman JD. 1989a. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* **43**(3): 301-335

Sterman JD. 1989b. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science* **35**(3): 321-339

Sterman JD. 1994. Learning in and about complex systems. *System Dynamics Review* **10**(2-3): 291–330

Sterman JD. 2000. *Business dynamics: Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill: Boston

Sterman JD. 2002. All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501-531

Sterman JD. 2010. Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review* **26**(4): 316-334

Sterman JD, Booth Sweeney L. 2002. Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review* **18**(2): 207-240

Sterman JD, Booth Sweeney L. 2007. Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change* **80**(3-4): 213-238

Süß H-M. 1996. *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Hogrefe: Göttingen et al.

Trochim WMK, Donnelly JP. 2007. *The research methods knowledge base* (3 ed.). Atomic Dog Publishing

Tversky A, Kahneman D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157): 1124-1131

Vinson DE, Lundstrom WJ. 1978. The Use of Students as Experimental Subjects in Marketing Research. *Journal of the Academy of Marketing Science* **6**(2): 114

Wittmann WW, Hattrup K. 2004. The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science* **21**(4): 393-409