

Does formal system dynamics training improve people's understanding of accumulation?

John D. Sterman
MIT Sloan School of Management
30 Wadsworth Street, E53-351
Cambridge MA 02142 USA
jsterman@mit.edu

Abstract

Prior work shows widespread misunderstanding of stocks and flows, even among highly educated adults. People fail to grasp that any stock rises (falls) when the inflow exceeds (is less than) the outflow. Rather, people often use the *correlation heuristic*, concluding that a system's output is positively correlated with its inputs. Here I report an experiment with MIT graduate students to assess the impact of an introductory system dynamics course on intuitive understanding of accumulation. I use a pre-test-treatment-post-test design; the treatment is the course content. Results show improvement in performance and a reduction in the prevalence of the correlation heuristic. Modest exposure to stocks and flows improves intuitive understanding of accumulation, at least among these highly educated adults. However, there is still evidence of correlational reasoning among a minority of students. I suggest additional experiments to deepen our knowledge of the training required to develop people's understanding of accumulation.

KEYWORDS: accumulation, stocks and flows, correlation heuristic, systems thinking, bathtub dynamics, misperceptions of feedback

Research shows that many people do not understand the distinction between stocks and flows and are unable to infer correctly the behavior of a stock from the behavior of its inflows and outflows (graphical or intuitive integration), or infer the behavior of the net flow from the trajectory of the stock (graphical or intuitive differentiation). Stock-flow problems, even simple ones, are unintuitive and difficult, even for highly educated people with substantial training in Science, Technology, Engineering, and Mathematics (STEM) (Booth Sweeney & Sterman, 2000; Cronin, Gonzalez and Sterman, 2009; Cronin & Gonzalez, 2007; Sterman 2002; Sterman & Booth Sweeney, 2002; Pala and Vennix 2005). In the original study, Booth Sweeney and Sterman (2000) presented highly educated graduate students with a picture of a bathtub and graphs showing the inflow and outflow of water, then asked them to sketch the trajectory of the stock of water in the tub. Although the patterns were simple, fewer than half responded correctly. These results have now been replicated with a variety of other populations (e.g., Pala and Vennix 2005). Importantly, recent work shows that performance remains poor in even simpler tasks and across a wide range of data display and response modes (Cronin, Gonzalez and Sterman 2009). Such stock-flow (SF) failures have important public policy implications, including widespread failure to understand the fundamental relationships between greenhouse gas emissions, atmospheric GHG concentrations, and climate change (Sterman 2008, Sterman and Booth Sweeney 2007).

The prior work clearly establishes widespread misunderstanding of the fundamental principles of accumulation. People fail to grasp that the quantity of any stock, such as the level of water in a tub, rises (falls) when the inflow exceeds (is less than) the outflow. Instead, people often use intuitively appealing heuristics such as assuming that the output of a system is positively correlated with its inputs. That is, people assume that the output (the stock) should “look like” the input (the flow or net flow). Cronin, Gonzalez and Sterman (2009) denote such behavior the *correlation heuristic* and show that such correlational reasoning is common in a wide range of stock-flow tasks. They further show that these stock-flow errors are robust to a wide range of information displays, cover stories and contexts, motivation, and other conditions.

The obvious question is what can be done to improve people's intuitive understanding of accumulation. In particular, are formal courses in system dynamics effective in overcoming people's poor understanding of stocks and flows? The answer may also seem obvious—training students in a specific skill should improve performance. However, several arguments suggest system dynamics training might not help. First, prior studies show that students with extensive STEM training, including calculus, where accumulation is a central concept, do poorly in stock-flow tasks. Second, the literature on transfer of learning across domains (e.g., Holyoak 1987; see references in Genter, Loewenstein and Thompson 2003) generally shows it is difficult for people to transfer insights learned in one domain to another, even when the underlying task structure is isomorphic. Learning to recognize stock-flow structure in a novel situation and apply the principles of accumulation may be subject to the same limitations. Third, as suggested by Cronin et al. (2009), SF failures may be similar to the many errors and biases in probabilistic reasoning documented in the judgment and decision making literature (e.g. Gilovich, Griffin and Kahneman 2002, Kahneman, Slovic and Tversky 1982). Such errors (including insensitivity to sample size, failure to account for regression to the mean, the conjunction fallacy, and many others) are surprisingly resistant to formal training in probability and statistics. Fourth, as suggested by Sterman and Booth Sweeney (2007), correlational reasoning may dominate stock-flow reasoning because the former had high survival value when the human brain evolved, while the latter did not, leading to neural structures in which correlational reasoning is more automatic while stock-flow reasoning requires conscious cognitive effort.

As discussed below, few prior studies examine the impact of formal system dynamics training on SF failure. The results are mixed, with some reporting positive impact of system dynamics training and other showing no impact. Some have small sample sizes, vary multiple factors simultaneously, or suffer from selection bias. Others do not report the nature of the system dynamics training and course content or the educational background and other demographic characteristics of the subjects. None examine the impact of system dynamics training on the use of the correlation heuristic.

Here I report an experiment with a large sample of highly educated graduate students to assess whether a half-term introductory course in system dynamics improves their intuitive understanding of accumulation. The study uses a pre-test–treatment–post-test design. The treatment consisted of the standard course material on stocks and flows, including several class sessions, assigned reading on stocks and flows from Sterman (2000), and an assignment; these are available in the supplement.

Results show improvement in overall performance, and a reduction in the prevalence of the use of the correlation heuristic. Even modest exposure to the concepts of stocks and flows and the principles of accumulation improves the intuitive understanding of these concepts, at least among these highly educated adults. The results should be reassuring to those who teach system dynamics—if there were little improvement, it would call into question the value of current system dynamics syllabi and pedagogical approaches (at least, the one used here).

However, several questions remain. While performance improved, there is still evidence of correlational reasoning among a number of students. Further, the robustness of the improvement is not known. Will these students be able to apply the principles of accumulation in naturalistic contexts they will encounter outside the system dynamics classroom, where there are few or no cues indicating that stock-flow structure and the principles of accumulation are applicable? How durable will student skills be as time passes? I discuss these issues and suggest additional experiments to deepen our knowledge of the education and experiences that can develop people’s intuitive understanding of accumulation.

Misperceptions of Feedback and Stock-Flow Failure

Research in dynamic decision making shows that high levels of *dynamic complexity* lead to systematically biased and suboptimal performance. Dynamically complex systems contain multiple feedback processes, including both positive and negative feedbacks, time delays, nonlinearities, and accumulations (Sterman, 2002). Research further shows that learning in dynamic systems is often slow and weak, even with repeated trials, unlimited time, and

performance incentives (Diehl & Sterman, 1995; Kleinmuntz & Schkade, 1993; Moxnes, 2004, Sterman, 1989a, 1989b). Poor performance in such experimental systems is often attributed to the gap between the complexity of the system and the bounded rationality of human decision-making, specifically, limits on cognitive resources resulting in information overload and computational constraints (Brehmer, 1990, 1995; Gonzalez, 2005; Kleinmuntz, 1985, 1993; Omodei & Wearing, 1995, Jensen and Brehmer, 2003).

Recent work shows that people make persistent mistakes even in the simplest dynamic systems with no feedback processes, time delays, or nonlinearities, including systems consisting of a single stock with one inflow and one outflow, (e.g., Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Sterman & Booth Sweeney, 2007, Cronin, Gonzalez and Sterman 2009, Pala and Vennix 2005). For example, Sterman (2002) describes the “department store” task, which presents participants with a graph showing the number of people entering and leaving a department store each minute over a 30-minute interval (Figure 1). The system consists of a single stock (the number of people in the store) with one inflow (people entering) and one outflow (people leaving). There are no feedbacks, time delays, nonlinearities, or other elements of dynamic complexity. Participants are asked four questions. The first two—“When did the most people enter the store? When did the most people leave the store?”—test whether participants can read the graph and correctly distinguish between inflow and outflow. The next questions—“When were the most people in the store? When were the fewest people in the store?”—test whether participants can infer the behavior of the stock from the flows.

To answer participants could keep a running tally of the number of people in the store minute by minute, $S_t = S_{t-1} + I_t - O_t$. This brute-force method, however, is tedious, error prone, and unnecessary. Rather, if participants understand the principles of accumulation they can answer without any calculation. Like any stock, the number of people in the store rises (falls) when the inflow—the number of people entering each minute—exceeds (is less than) the outflow—number of people leaving each minute. The number entering exceeds the number exiting through $t = 13$ and is less thereafter. Therefore, the most people are in the store when the

two curves cross ($t = 13$). Furthermore, because the number of people in the store rises through $t = 13$ and falls thereafter, the fewest people are in the store either at the beginning or the end of the 30 minutes. To determine which, participants must judge whether the cumulative increase in the store population through $t = 13$ is greater or less than the cumulative decrease from $t = 13$ to 30. Calculation is again unnecessary: participants need only judge whether the area between the rate of entering and the rate of leaving up to $t = 13$ is greater or smaller than the area between the two curves from $t = 13$ on. The area between the curves for $t \geq 13$ on is clearly larger, so the fewest people are in the store at the end of the 30 minutes. As described in Sterman 2002 and Cronin et al. 2009, the task was carefully designed so that area of the region in which outflow exceeds inflow ($t \geq 13$) is twice as large as the area in which inflow exceeds outflow ($t \leq 13$). To test whether people can determine which area is larger, a convenience sample consisting of 12 members of the support staff from the MIT Sloan School of Management were asked which area was greater; all correctly identified the larger area.

Despite the extreme simplicity of the department store task, Cronin, Gonzalez and Sterman (2009) show that performance by a sample of graduate students enrolled in the introductory system dynamics class at the MIT Sloan School of Management was poor. Participants ($N = 173$) were primarily MBA students and graduate students from other MIT departments or from Harvard University. The mean age was 29 and 78% were male. All had taken calculus, and most had strong training in science, technology, engineering, or mathematics (STEM): 71% had a degree in STEM; 28% had a degree in the social sciences, primarily economics. Fully 40% had a prior graduate degree, most in technical fields. Students did the task in class at the beginning of the semester, prior to any exposure to system dynamics concepts, including stocks and flows. As expected for this highly educated population, the vast majority of participants correctly identified when the most people entered and left the store (96% and 95% for Questions 1 and 2, respectively). However, few were able to answer the stock-flow questions correctly (44% and 31% for Questions 3 and 4, respectively). Approximately 17% indicated that it is not possible to determine when the most people were in the store, and 25% said that it is not

possible to determine when the fewest people were in the store. More importantly, 29% incorrectly indicated that the most are in the store when the net inflow is greatest ($t = 8$) and 30% incorrectly conclude that the fewest are in the store when the net outflow is greatest ($t = 17$). These responses, accounting for far more of the erroneous choices than any other, reveal a fundamental confusion about the relationship between stocks and flows. Cronin, Gonzalez and Sterman (2009), using subjects drawn from MIT, Carnegie-Mellon University and George Mason University, show that the poor performance persists when the task is simplified (fewer data points), when the data display is varied (from line graph to bar graph, spreadsheet, or text), when more time is allowed, and when subjects are provided modest incentives and opportunities for learning from outcome feedback.

The results have been extensively replicated (see the review in Pala and Vennix 2005, also Ossimitz 2002, Kapmeier 2004, Kasperidus et al. 2006 and Jensen 2008). In contrast, only a few studies examine whether system dynamics training reduces the incidence of SF failure.

In an unpublished conference paper, Kainz and Ossimitz (2002) used a pre-test, post-test design to explore the impact of a 90-minute stock and flow “crash course”. They report statistically significant improvement in performance on most of the post-test items, though performance generally remained poor.¹ However, the content of the crash course is not available, so it is difficult to know what the students were taught. The post-test items were essentially identical to the pretest items (though properly counterbalanced in presentation order), raising the possibility that the observed improvement reflected memory rather than improved stock-flow reasoning skills. Most important, of 94 subjects who completed the pre-test and crash course only 64 continued the class or “could be induced to participate in the post-test” (p. 5), raising the possibility that the improvement in performance results from selection bias. Students who dropped the course may have believed themselves to be doing poorly, while the better students remained. Similarly, those who “could not be induced to participate in the post-test”

¹ For example, on the simplest bathtub task reported in Booth Sweeney and Sterman (2000), the Kainz and Ossimitz subjects averaged 36% correct in the pre-test and 54% correct in the post-test, both far below the average of 83% correct for the MIT students tested prior to any formal system dynamics training by Booth Sweeney and Sterman.

may have believed that they would do poorly compared to those who took the post-test. In either case the results would be biased toward improvement on the post-test, even if the crash course had no impact. No analysis is presented to rule out either of these potential sources of selection bias. It is therefore not possible to determine whether the crash course actually improved performance in stock-flow reasoning.

In another conference paper, Jensen (2008) compares the performance of students in the “Rabbits and Foxes” task, a simulation game in which participants must equilibrate populations in a predator-prey system. One group of 15 were system dynamics students at the University of Bergen, another group of 22 were engineering students at the Swedish Royal Institute of Technology, and a third group of 10 were first year students at Uppsala University. Success rates were about the same in all three groups, suggesting no impact of either system dynamics or engineering training, though there were differences in the problem-solving strategies applied. However, it is difficult to interpret the results as the sample size is small and there are multiple differences across the groups examined, including differences in their prior training and academic achievement, home institution, gender, native language, and so on.

Lyneis and Lyneis (2003) compared undergraduates at Worcester Polytechnic Institute enrolled in introductory microeconomics to those enrolled in introductory system dynamics. Those taking system dynamics did better on Sterman’s (2002) Department Store task than those taking microeconomics (though performance was still poor). However, the sample size for the system dynamics class was small (14) and the task was given near the end of the course, while the economics students did the task at the beginning of the term. Without a pre-test, post-test comparison and given the multiple differences between the two groups it is difficult to attribute the improvement to the system dynamics training received.

The (apparently) only previously published study of the impact of system dynamics training on stock-flow performance is Pala and Vennix (2005). They report three experiments. Using a pre-test, post-test design with a control group they found students enrolled in an introductory system dynamics course showed greater improvement on the department store task

than students enrolled in a research methods course that did not cover stocks and flows. The sample size is large. However, as is often the case in such studies, limitations on permissible manipulations constrained the design. The same task was used in both pre-test and post-test. As Pala and Vennix note, the improvement observed in post-test performance “could be the result of doing the same task twice.” The system dynamics students improved more than those in the research methods course, suggesting a beneficial impact of system dynamics training beyond the impact of students’ memory of the pre-test. However, students were not assigned randomly to the treatment and control conditions, and there were a number of differences between the two groups including age. Hence the larger improvement in post-test performance among the system dynamics students cannot be attributed to the system dynamics training. In the second experiment (the manufacturing task from Booth Sweeney and Sterman 2000), a pre-test, post-test design without control group was used with students taking a full-semester system dynamics class. The post-test was the same as the pre-test, again raising the issue of memory. Even so, overall performance on the post-test was statistically significantly worse. In the third experiment (the climate change task in Sterman and Booth Sweeney 2002), there was no significant improvement in performance in the post-test.

Overall, prior research exploring the impact of formal system dynamics training on SF failure is sparse, and the results mixed. Those studies reporting positive impacts suffer from various design flaws that make it difficult to attribute the positive results to the training. Other experiments show no impact. Additional work is needed to build our understanding of whether and how people’s intuitive understanding of accumulation can be improved.

Method

The study uses a pre-test-treatment-post-test design (Figure 2). Participants were students enrolled in the introductory system dynamics class at the MIT Sloan School of Management in the fall term 2008. The course is divided into two half-semester courses. Both halves are electives, and students may opt to take only the first half or both halves. The

experiment was carried out within the first half-term course, which consists of eleven 80-minute sessions, meeting twice per week. In the fall term 2008 there were two sections of the course, taught in back-to-back time slots. To establish a baseline, students were given the classic department store task (Figure 1) on the first day of the semester as a pre-test. The treatment consisted of the standard course material, which covers principles of system dynamics and tools for dynamic modeling and systems thinking including causal loop diagrams, stock and flow mapping, and computer simulation. Students complete five assignments in the half term. These include: building a simple simulation model of the SARS epidemic (Assignment 1); developing causal diagrams of various business and public policy issues (Assignment 2); stocks and flows, including identification, mapping, graphical integration, and building simple simulation models (Assignment 3); applying their modeling skills to evaluate the business strategy of a firm or of their choice (Assignment 4), and the People Express Management Flight Simulator (Assignment 5). Stocks and flows were introduced in the first class (after the pre-test was administered). Two sessions (sessions 4 and 5) were specifically devoted to stocks and flows; sessions afterwards often used stock and flow diagrams and concepts in developing the examples used in class discussion. Students were assigned reading for each class, including the chapters on stocks and flows from Sterman (2000).²

Pre-test, post-test designs suffer from an intrinsic problem: If the post-test is the same as the pre-test, any performance improvement may arise from memory of the specific pre-test items rather than improvement in the underlying problem solving skills of interest. To avoid performance improvement arising from reuse of the pre-test as the post-test instrument, the post-test consisted of the graphical department store task described in Cronin, Gonzalez and Sterman (Figure 3). The graphical department store task also allows a direct test of the extent to which people rely on the correlation heuristic. The post-test was administered in the 9th class session, two sessions after students completed the assignment on stocks and flows. Administering the

² The syllabus, readings, and the stock-flow assignment are available online at <http://stellar.mit.edu/S/course/15/fa08/15.871ab/>. See also <http://ocw.mit.edu/OcwWeb/Sloan-School-of-Management/15-874Fall2003/CourseHome/>. The stock-flow assignment is assignment 3 (see appendix).

post-test one week after the due date for the stock and flow assignment reduces the chance of priming the students that the post-test must involve stock-flow reasoning. Because the pre- and post-tests are different tasks performance on them does not directly assess the extent of improvement resulting from the treatment. To do so, I compare performance on the post-test to the performance of the students who completed the same graphical department store task on the first day of the same class in the fall term of 2007. As shown below, the demographics of these subjects, whose performance is reported in Cronin et al. (2009), are not statistically significantly different from those of the students who completed the task as the post-test in the Fall of 2008. The difference in performance between these two groups can therefore be interpreted as a measure of the impact of the treatment, that is, of participating in the class and being exposed to the material on stocks and flows.

Administration of the pre- and post-tests followed the protocol described in Cronin et al. (2009) so that the results could be compared. Specifically, students received the pre- and post-tests at the beginning of the first and ninth class periods, respectively. The tasks were administered on paper. Cronin et al. show that question order in the classic department store task (the pre-test) had no impact, so all students received the questions in the same order (shown in Figure 1). Students also provided demographic information such as age, gender, work experience, etc. In the case of the post-test, students were randomly assigned to each of the 8 experimental conditions. For both pre- and post-test, students were given ten minutes to complete the task; as in prior use of these tasks, many students finished far faster. Students were told that participation was voluntary and that the results would not be graded, but would be helpful to the instructor in improving future offerings of the course and in this research.

Subjects

A total of $N_{\text{pre}} = 255$ students completed the pre-test and provided usable demographic information. Of these, $N_{\text{post}} = 173$ completed the post-test. Table 1 summarizes the demographics for the pre- and post-tests and compares the subject pool to the samples reported

in Cronin, Gonzalez and Sterman (2009). Mean age for the pre-test was 28.4 years (range: 19 – 39), with a mean of 4.9 years of work experience (range: 0 to 15), and 71% were male. Seventy percent were 2nd year MBA students, 10% were enrolled in the Leaders for Manufacturing program (LFM, a dual-degree program in which students receive both an MBA and a masters degree in engineering), 15% were MIT graduate students in other programs masters and doctoral students in science, engineering and management, roughly 3% were MIT undergraduates and approximately 3% were graduate students from Harvard and other universities. As in prior semesters of the class the students are highly trained in technical fields: 58% list science, technology, engineering, or mathematics as the field of their highest prior degree; 36% are trained in the social sciences, including economics, business and finance; 3% are trained in architecture; only 3% listed a field in the humanities.

To determine the extent of prior exposure to system dynamics concepts, students were asked whether they had played the Beer Game (Sterman 1989); 86% had done so (the beer game is used as the capstone event in the orientation program for incoming MBA students, hence most of the MBAs had played the game approximately one year prior to enrolling in the class). In addition, 25 students (10%), all 2nd year MBAs, had participated in a half-day workshop on the dynamics of climate change the author conducted in the spring term of 2008. That workshop focused explicitly on stock-flow structure and included several graphical integration exercises with climate change cover stories (Martin 2008, Sterman and Booth Sweeney 2007). Finally, students were asked if they had seen the classic department store (pre-test) task before; only one had, and this subject is excluded from the analysis.

Of those who provided demographics and completed the pre-test, a total of $N_{\text{post}} = 167$ students completed the post-test at the beginning of the ninth class session. The difference in sample size between the pre- and post-test reflects the fact that the course is an elective, so a number of students who attended the first session and completed the pre-test either dropped the course or chose not to attend the day the post-test was administered. An additional 28 students completed the post-test but not the pre-test, indicating that they did not attend the first class in

which the pre-test was administered. Demographic data for these students are not available and they are not included in the analysis.

To assess the extent of selection bias among those who survived the first day and later completed the post-test, first, note that both the pre- and post-test were unannounced so it is not possible for students to skip class to avoid them, ruling out selection bias arising from individual student's self-assessment of their understanding of stocks and flows. Second, the course was oversubscribed, with more people attending the first day (when the pre-test was administered) than ultimately were enrolled. Those who did not continue were thus selected out based on their position on the wait list, not their abilities or backgrounds. To test for differences between the pre- and post-test groups, I compared the demographics of those who completed the pre-test only to those who completed both pre- and post-test. Age was statistically significantly higher in the post-test group ($t = 2.05, p = .042$), but the difference between the group means of 0.4 years is not substantively significant. The only other statistically significant difference is in the proportion of students who had participated in the Beer Game, which increased from 75% among those completing the pre-test only to 91% among those who completed both pre- and post-test (2-sided Wilcoxon test; $p = .0006$). The proportion increased because the course was oversubscribed on the first day; Sloan rules require priority be given to Sloan students, all of whom experience the Beer Game during MBA orientation, compared to students from other programs and universities, most of whom have not played the game. All other factors, including sex, English as a native language, work experience, field of study, highest prior degree, etc. were not statistically significantly different between those who completed the pre-test only and those who completed both pre- and post-test. Further, there were no statistically significant differences on any of the demographic measures between those taking the graphical department store task as the post-test in the Fall of 2008 and those who completed the same task on the first day of class in the Fall of 2007 (at $p < 0.10$). Therefore it is reasonable to compare the performance of these groups to measure the impact of the treatment.

Results: Pre-test

Table 2 presents the results of the pre-test for both the full group that completed the pre-test, $N_{\text{pre}} = 255$, and the subsample of those who later completed the post-test, $N_{\text{post}} = 167$. As in Sterman (2002) and Cronin et al. (2009), responses were considered correct if they were within ± 1 minute of the correct answer. For example, the most people enter the store at $t = 4$; responses of $t = 3, 4$, or 5 were coded as correct. The task is designed so that the key events listed in table 2 are separated by more than two minutes.

Consistent with prior results, these highly educated subjects are able to read the graph and distinguish between those entering and those leaving. For the full sample, N_{pre} , performance on Q1: Most entering was 95% and on Q2: Most leaving was 93%. Most of those responding incorrectly reversed the entering and leaving data, gave the maximum net in- or out-flow instead of the gross flows, or gave the y-axis values rather than the time at which the maximum gross flows occurred.

However, performance on the two stock-flow questions is poor. Only 51% correctly identified when the most people are in the store (Q3), while 14% said it can't be determined and 26% selected $t = 8$, which is the point at which the net inflow to the store reaches its maximum. Only 38% correctly determined when the fewest are in the store (Q4), while 19% said it can't be determined and 22% selected $t = 17$, which is the point at which the net outflow reaches its maximum.

As described above, only 167 of the 255 students who took the pre-test went on to take the post-test. It is important to test for selection bias among those who went on. As previously described, there are no important, statistically significant differences in the demographic attributes of those who took the pre-test only compared to those who did both pre- and post-tests. As a further test of potential selection bias, I compare the fraction correct on each of the four department store questions for those who later took the post-test (the "post" group) to those who only took the pre-test (the "~post" group). There are no statistically significant differences (the Fisher exact test of H_0 : Fraction correct(post) = Fraction correct(~post) yields $p = .76, 1.00, .79,$

.89 for Q1-4, respectively). There do not appear to be any important differences in the responses of those who continued with the class and took the post-test compared to those who did not.

Table 2 also compares the results against the results of the classic department store task reported in Cronin et al. (2009). The data reported by Cronin et al. should be comparable: they were collected in a prior semester of the same course, using the same protocol and under nearly identical circumstances (the first day of class, the same instructor, the same room). The performance of the 2008 students is not statistically different from those of the Cronin et al. group: For the full pre-test sample, the Fisher test yields $p = .82, .68, .17, .15$ for the fraction correct on Q1-4, respectively (results are similar when the N_{post} sample is used).

The pre-test results show that, prior to any exposure to system dynamics concepts, many students have a weak grasp of stock-flow principles. Despite the simplicity of the department store task, fewer than half correctly identify when the most and fewest people are in the store. Nearly a quarter of all respondents mistake the maximum net in- and out-flow rates for the maximum and minimum of the stock of people in the store, a fundamental confusion about the process of accumulation. The question now is whether exposure to these concepts in the course improves their understanding of stock-flow relationships.

Post-test: Method

The post-test uses the graphical department store task shown in Figure 3. Cronin, Gonzalez and Sterman (2009) use this task to determine the extent to which people erroneously rely on the “correlation heuristic” in assessing the behavior of stock-flow systems.

Suggested by prior work (Booth Sweeney and Sterman, 2000), the correlation heuristic is a form of pattern matching in which people assume that the output of a system (e.g., the number of people in the store) should “look like” the input (the flow or net flow of people into the store). Booth Sweeney and Sterman (2000) found extensive use of the correlation heuristic among erroneous responses to simple tasks including inferring the level of water in a tub or the cash balance of a firm from graphs of the inflows and outflows. These results have been replicated

with diverse student populations (e.g., Atkins et al., 2002; Ossimitz, 2002; Pala and Vennix, 2005). The graphical department store task was designed to test the extent to which people rely on the correlation heuristic, and to identify which cues—inflow, outflow, or net flow—people select as the basis for estimating the behavior of the stock.

As shown in Figure 3, each of the eight different conditions for the graphical department store task consists of a graph showing the flow of people entering and leaving a store over 30 minutes. Participants were directed to draw the number of people in the store throughout the 30 minutes on a blank graph placed directly beneath the flow graph. The eight flow patterns ranged from constant flows to more complex shapes. Note that no numerical scales are provided for the flow data or for the blank graph for the subjects' response. The graph for the stock includes a point indicating the initial number of people in the store. To avoid biasing participant responses, that point is placed at the midpoint of the vertical axis. In all cases, it is possible to answer correctly without knowledge of calculus and without carrying out any calculations.

Subject responses were coded correct or incorrect, and correlations between the pattern drawn for the stock and any of the flows, if present, were noted. A response was judged qualitatively correct if it was consistent with basic stock-flow principles: (i) the stock is rising, constant, or falling when the net inflow is positive, zero, or negative, respectively; and (ii) the rate of change (slope) of the stock is increasing (decreasing) when the net flow is increasing (decreasing). Participants were not penalized for drawing patterns that were not quantitatively correct or that did not show the number in the store beginning at the initial point provided on the graph. Erroneous responses were coded to determine whether the correlation between the stock and inflow or net flow was +1 (perfect pattern matching), or -1. A correlation of -1 indicates perfect pattern matching, but with the pattern inverted; such inversion might occur when the net flow is positive but falling (e.g., condition 5); in such a case the participant realizes that the stock is rising, but still erroneously concludes the stock follows the shape of the net flow.

The eight flow patterns divide into three groups. Group I consists of conditions 1 and 2 and should be the easiest: participants need only realize that the net flow is constant, determine

whether it is positive or negative, and draw a straight line with positive or negative slope. Group II consists of conditions 3, 4 and 5. These all have constant outflow and linear inflow: participants must determine whether the net flow is positive or negative, note whether the net flow is increasing or decreasing, and then draw a curve that is rising or falling at an increasing or decreasing rate. Group III comprises conditions 6, 7, and 8 and should present the greatest difficulty: These have constant outflows but nonlinear patterns for the inflow: participants must determine whether the net flow is positive or negative, then determine whether the net flow is increasing or decreasing in each part of the thirty minute interval, and sketch a path that shows the stock rising or falling with qualitatively correct changes in slope.

Table 3 presents the results for the post-test and compares them to the results in Cronin et al. (2009). The Cronin et al. results provide a useful comparison because they were collected on the first day of the same class in the prior year (2007). As shown in table 1, the demographics of the two groups are essentially identical. Further, the task was administered to the two groups under the same protocol, by the same instructor, in the same room, at nearly the same time of day. The main difference between the post-test group and the subjects in 2007 is that the 2007 subjects received the task on the first day of class while the 2008 subjects received it in the 9th class session, after studying stocks and flows.

Results improve significantly in the post-test compared to the results from 2007 (Figure 4, Table 3). Overall, 25% of the participants responded incorrectly, nearly half the rate of 46% in 2007, a highly statistically significant reduction ($p < 6 \times 10^{-6}$ by the Fisher test). Performance improved in all conditions except condition 2, where the difference (25% vs 22.2% incorrect) is not statistically significant ($p = 1$). Performance improved in all three groups, though the improvement in Group I is not statistically significant, perhaps because of the generally high performance and comparatively small sample size compared to the other groups. In Group I, the simplest tasks with constant flows, the fraction incorrect fell from 21% to 15% (but is not statistically significant; $p = .62$). Performance in Group II, three tasks with linearly changing net flows, improved significantly, from 46% incorrect to 15% incorrect ($p = 2.3 \times 10^{-5}$). In Group

III, the most difficult tasks, with nonlinear net flows, performance improved significantly, from 64% incorrect to 41% incorrect ($p = .006$).

Turning to the prevalence of pattern matching (the correlation heuristic), Figure 4 and Table 3 also show the fraction of those responding incorrectly whose responses exhibit perfect correlation with the input cues (the stock trajectory drawn by the subject is perfectly correlated, +1 or -1, with the inflow or net flow). The fraction of those responding incorrectly whose answers exhibit correlation fell overall, from 71% in Fall 2007 to 51% in Fall 2008, significant at $p = .024$. The proportion of erroneous responses exhibiting correlation fell in all three groups, though the drop is significant only for Group II. The incidence of erroneous use of the correlation heuristic relative to all subjects fell substantially and significantly, from 32.6% in Fall 2007 to 12.7%, $p = 1.8 \times 10^{-6}$.

Impact of Demographics on Performance

People's understanding of accumulation and the extent to which they learn from the course material may be affected by demographic characteristics including prior degrees, prior field of study, and so on. Booth Sweeney and Sterman (2000) found some evidence for such effects on a variety of graphical integration tasks, including some evidence of a field effect (those with more technical training did somewhat better) and weak evidence of a gender effect (males performed better than females). Kainz and Ossimitz (2002) also report a gender effect. The supplement (Appendix 2) reports both nonparametric tests and multivariate logistic regression models to explore the impact of subject demographics on performance.

Overall, there are no consistent and statistically significant effects of subject demographics on performance for both the pre- and post-test, with the exception of a gender effect. Males outperform females, even after controlling for other demographic attributes including prior field of study, current degree program, field of study, age and work experience.

An important issue is the extent to which performance on the post-test is predicted by performance on the pre-test. If correct responses on the pre-test are highly predictive of post-test success, then it may be that students did not benefit from the course material on stocks and flows

but rather that those who understood accumulation prior to the course simply did well on both pre- and post-tests, while those who did poorly on the pre-test also did poorly on the post-test. The logistic regressions suggest this is not the case (Supplement Table S-3). First, as expected, performance on pre-test questions Q1 and Q2, which assess whether subjects can interpret the graph, is not predictive of post-test success. More important, performance on pre-test questions Q3 and Q4, which assess whether subjects understand stocks and flows, is not predictive of post-test success. Prior understanding of stocks and flows does not explain the improvement on the post-test.

Discussion and conclusions

Research shows that people, including highly educated adults with substantial training in STEM, or quantitative social sciences, have poor understanding of stocks and flows and the principles of accumulation. These difficulties are not due to limits on working memory or mental computation capability, or to any easily correctable task features such as unfamiliar data presentation format, task context or cover story, insufficient time or lack of motivation (Gonzalez and Cronin 2007, Cronin, Gonzalez and Sterman 2009). Rather, people's difficulties with accumulation appear to be a robust cognitive deficit analogous to the difficulties people have in probabilistic reasoning. The challenge is how to overcome this difficulty.

Here I use a pre-test, post-test design to assess the extent to which a half-semester system dynamics course improves people's ability to apply the principles of accumulation. The subjects were students enrolled in the first half-semester introductory system dynamics class at the MIT Sloan School of Management. The pre-test, the classic department store task, showed the poor performance typical of this and other populations reported in prior work. The treatment consisted of the standard course material, including only eight class sessions, only two of which were completely devoted to the concepts of stocks and flows. Students were also assigned to read the chapters on stocks and flows in Sterman (2000), and completed four assignments before the post-test was administered, only one of which was focused on stocks and flows. That

assignment (see the appendix) covers stock and flow identification, mapping stock and flow networks in various situations, one example of graphical integration, and construction of simple simulation models illustrating first-order linear positive and negative feedback.

The results of the post-test show that performance on the graphical department store task improved substantially and statistically significantly compared to a demographically similar set of subjects who did the task at the beginning of the term the previous year. The overall fraction correct improved significantly, with the error rate falling by nearly half. Among those who responded incorrectly, the fraction using the correlation heuristic (matching the pattern of the stock to the pattern of the inflow or net flow) dropped substantially and significantly as well. Even the relatively brief exposure to stock-flow concepts provided by several class sessions, readings, and a single stock-flow assignment appear to improve people's abilities to recognize stock-flow structure and correctly apply the principles of accumulation.

Nearly all subject attributes had little or no impact on performance, for either the pre-test or post-test. Unsurprisingly, age, work experience, whether English was the student's native language, and experience playing the Beer Game had no impact. In contrast, one might expect that prior educational background might have a strong impact on people's ability to recognize stock and flow structure and apply the principles of accumulation. Surprisingly, however, the degree program in which the students were enrolled, which included both MBA students and graduate students in engineering and science, had no effect. Forty percent of the subjects earned a bachelor of science or engineering degree in their undergraduate training, and more than a quarter possessed a prior graduate degree, yet this factor was also not significantly related to performance on either the pre-test or post-test. Even more surprising, the subjects' field of study had no significant impact on performance. Nearly 60% of the subjects were trained in STEM (Science, Technology, Engineering, or Mathematics), another roughly 36% were trained in the social sciences, primarily economics and business, with the remainder trained in humanities or architecture. Yet field of study was not statistically significantly related to performance.

Among demographic factors, only gender was statistically significant, with males on

average performing better than females even after controlling for other demographic attributes such as field and prior education. A robust literature seeks to explain gender (and other differences) in mathematics achievement and STEM participation (see, e.g., Gallagher and Kaufman 2005). One possibility is that the effect is an artifact of variations in student ability correlated with gender but not captured by the demographic data collected. Tests for such unmeasured covariation might examine differences in college grade point average or scores on standardized tests used in admissions decisions such as the SAT and GMAT. Another possibility is stereotype threat, in which subtle cues in the task or task environment trigger negative stereotypes of female mathematics ability, lowering performance for women by enhancing anxiety (e.g., Spencer, Steele and Quinn 1999). Future work should explore these issues by manipulating cues that may trigger stereotypes.

Success on the stock-flow questions in the pre-test is associated with success in the post-test, as expected: those who understand stocks and flows prior to taking the course should do well on both pre- and post-test. However, pre-test stock-flow performance is only marginally significant as a predictor of performance on the post-test. The weak association of pre-test and post-test performance provides evidence that the course material improved the subjects' understanding of and ability to apply the principles of accumulation. The large improvement in performance compared to those who did the graphical department store task prior to taking the course is encouraging news for those who teach system dynamics.

However, several issues remain. Roughly 25% of the subjects still did the post-test incorrectly, and of these, half showed evidence of correlational reasoning. While performance improved by nearly half, and the incidence of correlational reasoning fell significantly, a disturbingly large minority of subjects still did not exhibit strong understanding of stock-flow concepts. As the instructor in the course, I hypothesize that the number of classes, problems to work, and assignments involving stock-flow concepts is simply not enough to provide sufficient practice for these concepts to become more broadly and deeply understood and internalized by the students. Given these results, the even shorter exposure to stock-flow concepts provided in

short academic and commercial training workshops is unlikely to be effective in overcoming the correlation heuristic and helping people learn the principles of accumulation. Those teaching system dynamics in other formats, and with other groups, should carry out evaluative research to assess the impact of their curriculum and pedagogy on student learning.

A second issue relates to the unusual characteristics of the subject population in this study. Graduate students at MIT are highly selected for top academic performance and capability; they have far more training in STEM and other quantitative disciplines (economics, business) than the average person. Still their understanding of stocks and flows prior to exposure to the course is extremely poor. These results are similar to prior studies with this population (Booth Sweeney and Sterman 2000). It may be that there is simply insufficient variance in prior education, field of study and other subject demographics to detect any effects. Alternatively, prior training and education may actually have only a weak effect. Developing and transferring knowledge from one domain or example to others is difficult (Holyoak 1987), so prior training in STEM disciplines may not be effective in helping people reason correctly about accumulations. As discussed by Sterman and Booth Sweeney (2002, 2007), the prevalence of the correlation heuristic in people's responses to stock-flow problems may reflect an evolutionary process. The ability to detect correlations among cues in the environment is highly adaptive, while the ability to relate stocks and flows offered no reproductive advantage. When modern humans evolved there was no need to graph or tabulate flows and use that data to make inferences about the stocks they affected. It was far easier for people to monitor the level of important stocks and take corrective actions when stock levels departed from their desired values—that is, to use standard negative feedback to control key stocks (when the fire burns low, put more wood on; when food stocks drop, gather more). Thus the ability to detect correlations and make inferences using them may have evolved as an automatic ability while relating stocks and flows requires training and significant cognitive effort. To gain insight into these issues, future studies should also explore people's reasoning processes through, e.g., verbal protocols.

Further, as discussed in Cronin, Gonzalez and Sterman (2009), much early mathematics

education emphasizes correlational reasoning, perhaps reinforcing the tendency to use the correlation heuristic. Further research into the failure of the educational system to teach principles of accumulation and other systems thinking skills, and effective methods to teach these concepts in the K-12 grades, is sorely needed (Booth Sweeney and Sterman 2007). There are some promising experiments underway, and the stock of educational materials for pre-college settings is growing (see the Creative Learning Exchange, <http://www.clexchange.org>, for examples). However, although the prior literature has replicated the basic findings of the original “bathtub dynamics” study with diverse populations including K-12 students, these replications do not allow inferences to be drawn about the effectiveness of different training methods or abilities of different subject populations because they were not done under controlled conditions. Future work should focus on reducing sources of unexplained variation across conditions by using standard protocols and tasks and by collecting and analyzing subject demographics.

It is also unclear how robust and durable the improved understanding of accumulation exhibited by the students tested here will be. Will their understanding of stocks and flows become internalized and readily recalled outside of the classroom and in later life? Will these students be able to recognize stock-flow structures and apply the principles of accumulation in everyday, naturalistic settings in which there are no special cues or prompts to trigger the relevance of stocks and flows? To address these questions, researchers should undertake long-term prospective longitudinal studies to assess the ability of subjects to recognize stock-flow situations and apply the principles of accumulation correctly in naturalistic settings. Such studies will be difficult. Besides the obvious challenge of subject recruitment for periods of years, such studies must avoid priming the participants to think about stocks and flows. Nevertheless, long-term follow up study is an essential next step towards the development of effective curriculum and pedagogy to develop people’s intuitive systems thinking abilities.

References

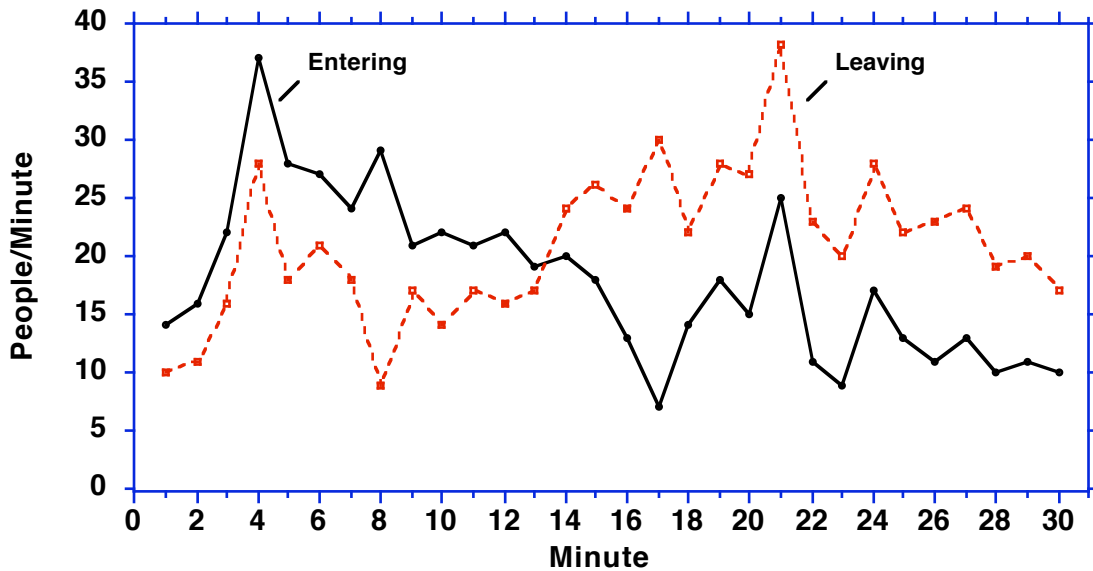
- Atkins, P., Wood, R., & Rutgers, P. (2002). The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes*, **88**(587-604).
- Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, **16**(4), 249-286.
- Booth Sweeney, L. and J. Sterman (2007). Thinking about systems: Students' and their teachers' conceptions of natural and social systems. *System Dynamics Review* **23**(2-3): 285-312.
- Brehmer, B. (1990). Strategies in real-time, dynamic decision making. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 262-279). Chicago: University of Chicago Press.
- Brehmer, B. (1995). Feedback delays in complex dynamic decision tasks. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cronin, M., and Gonzalez, C. (2007). Understanding the building blocks of system dynamics. *System Dynamics Review*, **23**(1), 1-17.
- Cronin, M., C. Gonzalez and J. D. Sterman (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.
- Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, **62**(2), 198-215.
- Gallagher, A. and Kaufman, J. (2005). *Gender Differences in Mathematics: An Integrative Psychological Approach*. Cambridge, UK: Cambridge University Press.
- Genter, D., Lowenstein, J., and Thompson, L. (2003) Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, **95**(2), 393-408.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Gonzalez, C. (2005a). Decision support for real-time dynamic decision making tasks. *Organizational Behavior & Human Decision Processes*, **96**, 142-154.

- Holyoak, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 332-340.
- Jensen, E. (2008) Does system dynamics or control theory help you to strike a balance? Proceedings of the 2008 International System Dynamics Conference. Available at: www.systemdynamics.org/conferences/2008/proceed/papers/JENSE203.pdf.
- Jensen, E., & Brehmer, B. (2003). Understanding and control of a simple dynamic system. *System Dynamics Review*, 19(2), 119-137.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kainz, D. and Ossimitz, G. Can Students Learn Stock-Flow-Thinking? An Empirical Investigation. Proceedings of the 2002 International System Dynamics Conference. Available at: www.systemdynamics.org/conferences/2002/proceed/papers/Kainz1.pdf.
- Kapmeier, F. (2004) Findings from four years of bathtub dynamics at higher management education institutions in Stuttgart. Proceedings of the 2004 International System Dynamics Conference. Available at: www.systemdynamics.org/conferences/2004/SDS_2004/PAPERS/197KAPME.pdf.
- Kasperidus, H-D, Langfelder, H. and Biber, P. (2006), Comparing Systems Thinking Inventory Task Performance in German Classrooms at High School and University Level. Proceedings of the 2006 International System Dynamics Conference. Available at: www.systemdynamics.org/conferences/2006/proceed/papers/KASPE299.pdf.
- Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, 31(6), 680-702.
- Kleinmuntz, D. N. (1993). Information processing and misperceptions of the implications of feedback in dynamic decision making. *System Dynamics Review*, 9(3), 223-237.
- Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science*, 4(4), 221-227.
- Lyneis J, Lyneis D. (2003). Bathtub dynamics at WPI. Presented at the 21st International Conference of the System Dynamics Society, New York. System Dynamics Society: Albany, NY.

- Moxnes, E. (2004). Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review*, **20**(2), 139-162.
- Omodei, M., & Wearing, A. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, **27**, 303-316.
- Ossimitz, G. (2002). Stock-flow-thinking and reading stock-flow-related graphs: An empirical investigation in dynamic thinking abilities. Paper presented at the International System Dynamics Conference.
- Pala, O., & Vennix, J. A. M. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review*, **21**(2), 147-172.
- Spencer, S., Steele, C. and Quinn, D. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, **35**(1), 4-28.
- Sterman, J. D. (1989a). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes*, **43**(3), 301-335.
- Sterman, J. D. (1989b). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, **35**(3), 321-339.
- Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Irwin/McGraw-Hill.
- Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, **18**, 501-531.
- Sterman, J. D. (2008). Risk Communication on Climate: Mental Models and Mass Balance. *Science* **322**: 532-533 (24 October).
- Sterman, J. D., & Booth Sweeney, L. (2002). Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review*, **18**(2), 207-240.
- Sterman, J. D., & 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.

Figure 1. “Classic” Department Store Task (Serman, 2002; Cronin, Gonzalez, and Serman, 2009).

The graph below shows the number of people *entering* and *leaving* a department store over a 30-minute period.



Please answer the following questions.

Check the box if the answer cannot be determined from the information provided.

1. During which minute did the most people enter the store?

Minute _____

Can't be determined

2. During which minute did the most people leave the store?

Minute _____

Can't be determined

3. During which minute were the most people in the store?

Minute _____

Can't be determined

4. During which minute were the fewest people in the store?

Minute _____

Can't be determined

Figure 2. Experimental Design.

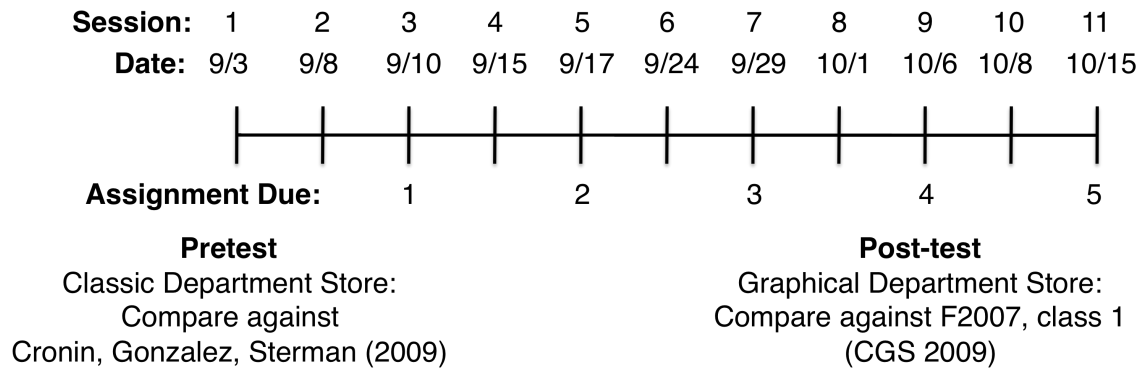
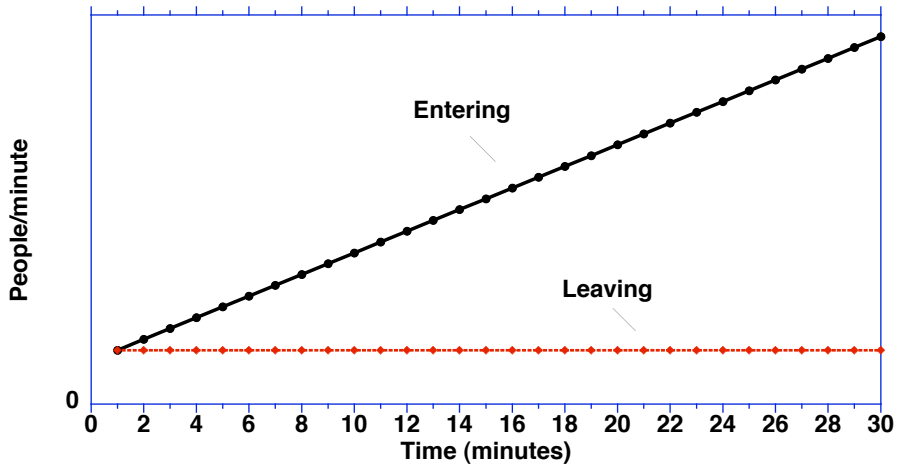


Figure 3. The “graphical” department store task, used to identify the prevalence of the correlation heuristic (Cronin, Gonzalez and Stermann, 2009). Subjects were randomly assigned to receive one of eight different patterns for the number of people entering and leaving the store (see Fig. 3 parts 2 and 3, below).

The graph below shows the number of people **entering** and **leaving** a department store over a 30 minute period.



In the space below, graph the number of people in the store over the 30 minute interval. You do not need to specify numerical values. The dot at time zero shows the initial number of people in the store.

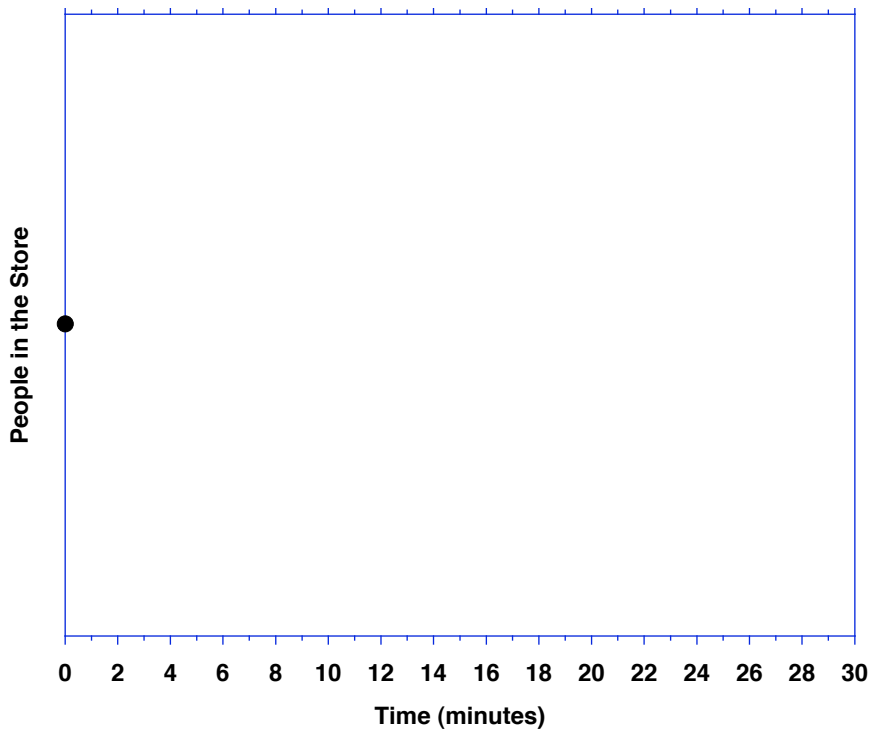
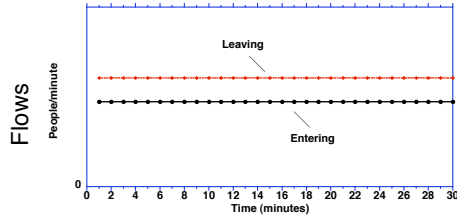


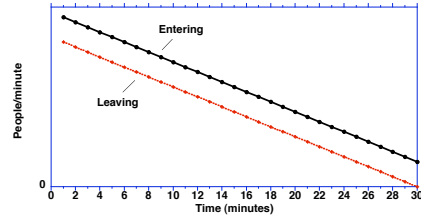
Figure 3 (continued). Correct and typical incorrect responses for the graphical department store task (Cronin, Gonzalez and Sterman 2009).

1. Constant Flows; $I < O$



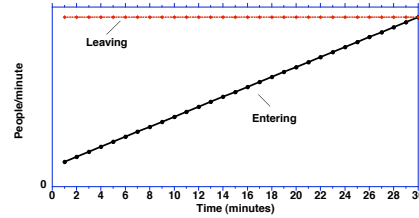
Net flow is constant and < 0 .
Stock falls linearly.

2. Linear decline in both I and O,
Constant Net Flow, $I > O$



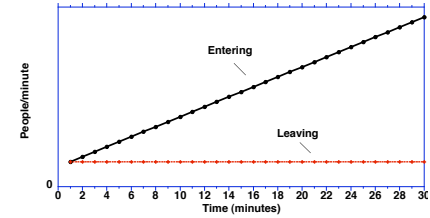
Net flow is constant and > 0 .
Stock rises linearly.

3. Constant Outflow, Linear increase in Inflow; $I \leq O$



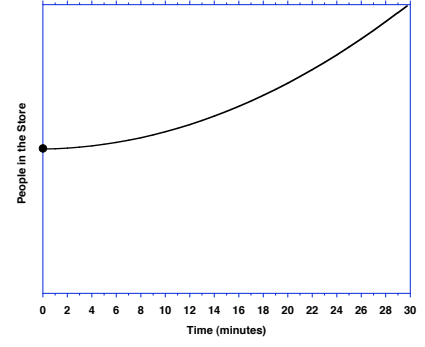
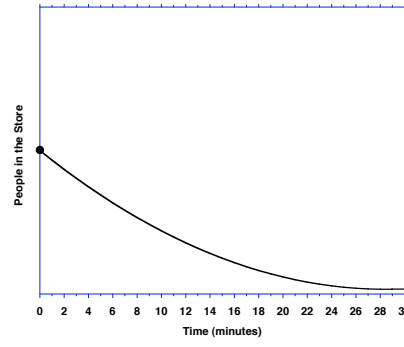
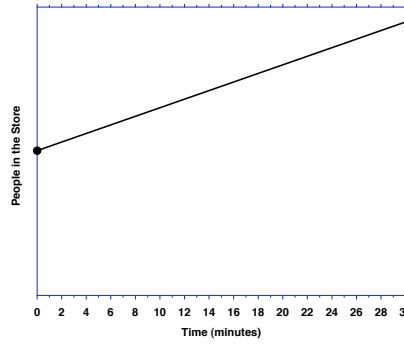
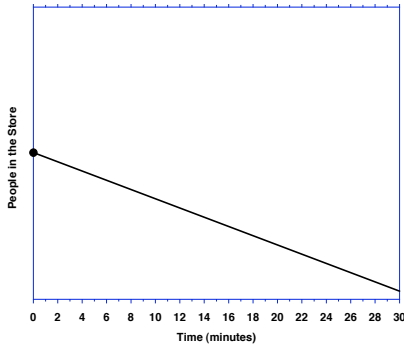
Net flow < 0 , rises linearly to 0 by
time 30. Stock falls at decreasing
rate, is constant at $t=30$

4. Constant Outflow, Linear increase in inflow; $I \geq O$



Net flow > 0 , rises linearly
throughout. Stock rises at
increasing rate from initial
equilibrium.

Correct Response



Response Showing Correlation

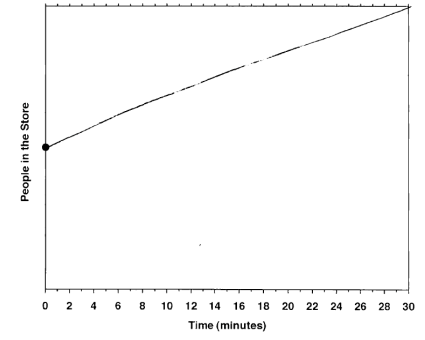
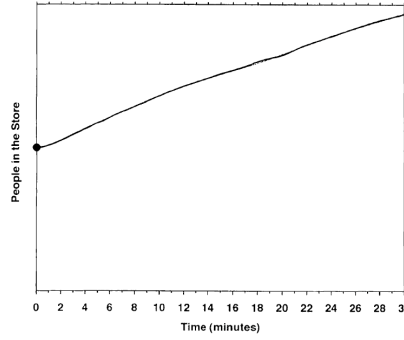
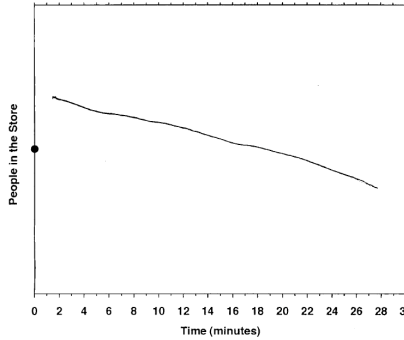
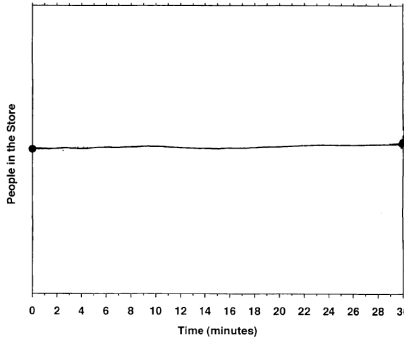


Figure 3 (continued). Correct and typical incorrect responses for the graphical department store task (Cronin, Gonzalez and Sterman 2009).

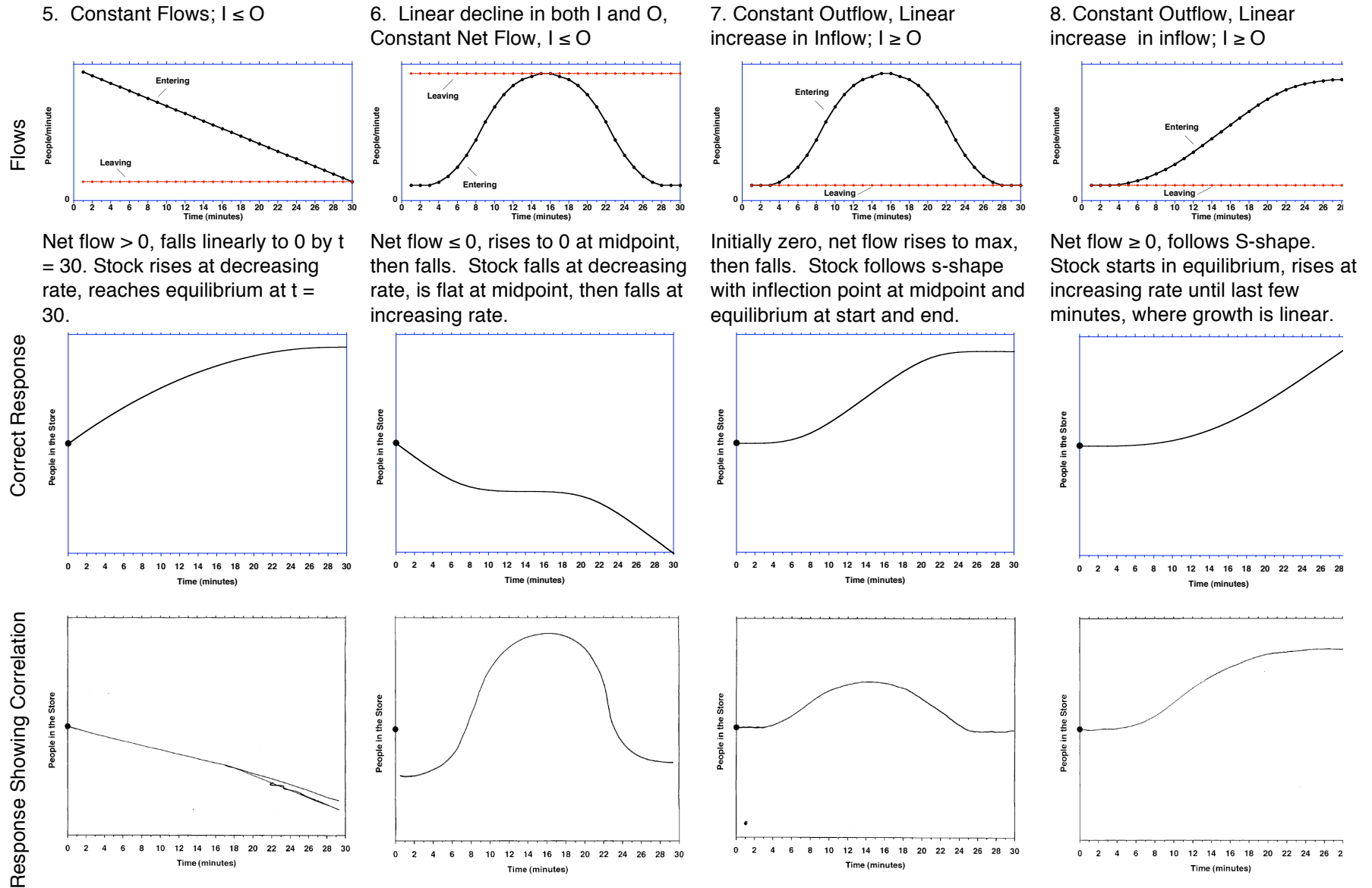


Table 1. Subject demographics for the pre-test, post-test, and comparison groups reported in Cronin, Gonzalez and Sterman (2009).

	1. Pre-test, Fall 2008 (N = 255)	2. Classic Department Store (Cronin et al.) (N = 173)	3. Post-test, Fall 2008 (N = 167)	4. Graphical Department Store, Fall 2007 (Cronin et al.) (N = 282)
Age	28.4 $\sigma = 3.3$ range 19-39	29.2 $\sigma = 4.7$ range 20-46	28.8 $\sigma = 3.1$ range 22-39	28.2 $\sigma = 3.4$ range 20-44
Years of Work Experience	4.9 $\sigma = 2.7$ range 0-15	NA	5.1 $\sigma = 2.7$ range 0-15	4.8 $\sigma = 2.9$ range 0-22
Gender = Male	.706	.789	.701	.706
Native language = English	.533	.500	.557	.630
Program				
1 st year MBA	0	.243	0	0
2 nd year MBA	.702	.139	.701	.716
LFM	.098	.046	.120	.067
Other MIT grad student	.145	.295	.156	.111
Other University	.028	.127	.006	.064
Undergraduate	.028	.012	.018	.039
Highest Prior Degree				
High School	.028	.012	.018	.039
BA	.306	.162	.283	.292
BS	.400	.422	.416	.377
Masters	.239	.353	.253	.249
Ph.D.	.012	.052	.018	.018
Other (JD, MD, etc.)	.016	0	.012	.025
Field of highest degree¹				
STEM	.582	.674	.600	.577
Social Science	.357	.267	.338	.385
Humanities	.033	.041	.031	.034
Architecture	.029	.017	.031	.004
Beer Game?	.855	.503	.910	.862
Climate Change Workshop?	.098	0	.096	0

Table 2. Pre-test results. The “Response” column lists the key events in the classic department store task (Figure 1). Each column shows the proportion of subjects who selected the response indicated in each row of column 1 for each of the four questions in the task. N_{pre} = results for full sample of 255 subjects who completed the pre-test. N_{post} = pre-test results for the 167 who also completed the post-test in class 9. CGS = results for 173 subjects who completed the pre-test on the first day of class in a prior semester, as reported by Cronin, Gonzalez and Sterman (2009). Bold indicates the correct answer.

Response:	Q1: Most Entering			Q2: Most Leaving			Q3: Most in Store			Q4: Fewest in Store		
	N_{pre}	N_{post}	CGS	N_{pre}	N_{post}	CGS	N_{pre}	N_{post}	CGS	N_{pre}	N_{post}	CGS
Max Entering t=4	.953	.958	.960	.012	.018	0	.006	.006	.035	.004	.006	.006
Max Leaving t=21	.004	.006	.012	.933	.934	.948	0	0	.006	.012	.012	.017
Max in Store t=13	0	0	0	0	0	0	.510	.503	.439	.035	.036	.023
Fewest in store t=30	0	0	0	0	0	0	0	0	.006	.380	.377	.312
Max Net Inflow t=8	.012	.006	.023	.008	.006	0	.263	.305	.289	0	0	0
Max Net Outflow t=17	0	0	0	.012	.006	.035	.035	.030	.035	.224	.240	.295
Initial in Store t=1	0	0	0	0	0	0	0	0	0	.090	.078	.069
Can't be Determined	.012	.012	0	.012	.012	0	.141	.138	.168	.192	.204	.249
Other	.016	.018	.006	.020	.024	.012	.027	.012	.012	.035	.036	.012
No Answer	.004	0	0	.004	0	.006	.016	.006	.012	.027	.012	.017

Figure 4. Graphical Department Store Task: Post-test results (Fall 2008) compared to results from first day of class in Fall 2007 (Cronin, Gonzalez and Serman, 2009). Left: percent responding incorrectly. Right: % of those responding incorrectly exhibiting pattern matching (using the correlation heuristic).

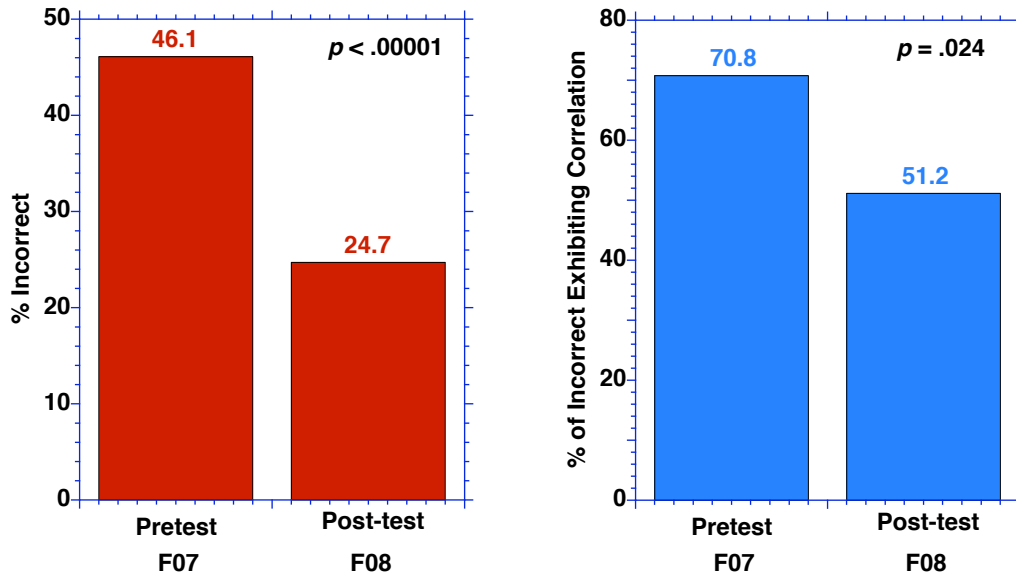


Table 3. Graphical Department Store Task: Post-test results (Fall 2008) compared to results from first day of class in Fall 2007 (Cronin, Gonzalez and Sterman, 2009). “Condition” refers to the pattern of inflow and outflow received (Figure 3).

Condition	Incorrect		% Incorrect exhibiting correlation		N	
	Fall 2008	Fall 2007	Fall 2008	Fall 2007	Fall 2008	Fall 2007
1 ¹	4.8%	16.7%	0.0%	33.3%	21	36
2 ²	25.0%	22.2%	40.0%	55.6%	20	37
3	22.7%	41.7%	40.0%	68.8%	22	37
4	4.8%	55.6%	100.0%	88.9%	21	34
5	15.8%	44.4%	0.0%	80.0%	19	35
6	47.8%	69.4%	36.4%	56.0%	23	36
7	15.0%	47.2%	100.0%	57.1%	20	33
8	60.0%	80.6%	75.0%	88.9%	20	34
All	24.7%	46.1%	51.2%	70.8%	166	282
	p = 5.7 x 10 ⁻⁶		p = 0.024			
Group I	14.6% 20.5%	33.3% 46.7%	41	73		
	p = .62		p = .66			
Group II	14.5% 46.2%	33.3% 79.6%	62	106		
	p = 2.3 x 10 ⁻⁵		p = .0096			
Group III	41.3% 64.1%	61.5% 69.7%	63	103		
	p = .0060		p = .47			

1. Inflow, Outflow and Net Flow are all constant. A subject’s response was coded as showing correlation if the response was also constant (a horizontal line).
2. Inflow and Outflow are correlated; Net Flow is constant. A subject’s response was coded as showing correlation if the response was either correlated to the inflow (a straight line) or was constant, matching the pattern of the net flow (a horizontal line).

Supplement to:

Does formal system dynamics training improve people's understanding of accumulation?

Appendix 1

Appendix 1 contains the syllabus for the course and the third assignment, which covers stocks and flows, including identifying and distinguishing stocks and flows, mapping the stock and flow structure of systems, graphical integration, and formulating and simulating simple models. Full information including all assignments and other materials is available on the course website, <http://stellar.mit.edu/S/course/15/fa08/15.871ab/>.

A prior version of the course and assignments is available on:

<http://ocw.mit.edu/OcwWeb/Sloan-School-of-Management/15-874Fall2003/CourseHome/>.



Massachusetts Institute of Technology
Sloan School of Management

15.871 Introduction to System Dynamics
15.872 System Dynamics II

Fall 2008

GENERAL INFORMATION

- Background:** 15.871 (Introduction to System Dynamics) is a 6 unit course meeting in H1. 15.872 (System Dynamics II) is a 6 unit course meeting in H2. Together they constitute the introductory sequence in system dynamics. You can take 871 alone or both 871 and 872. Successful completion of both 871 and 872 is a prerequisite for advanced courses in system dynamics, work as an RA or TA in the field, as well as careers using system dynamics.
- 871 & 872**
- Schedule:** Section A: Monday and Wednesday, 8:30 – 10:00 in E51-345.
Section B: Monday and Wednesday, 10:00 – 11:30 in E51-345.
- Instructor:** John Sterman, E53-351, 617.253.1951 (v), 617.258.7579 (f), jsterman@mit.edu
- Office hours:** My door is always open to students, or make an appointment by email.
- TAs:** REDACTED
- TA Sessions:** The TAs will lead a weekly review session in which they will answer questions about assignments in progress and discuss solutions to past assignments. There are two recitations: Friday, 10:00 – 11:30 and Friday, 14:30 – 16:00, both in E51-325. You may attend either one. The first session will be Friday, Sept. 5.
- Grading** Assignments: 85%
- Emphasis:** Class participation: 15%
- Each assignment is graded on a 10-point scale. Two points will be forfeited for assignments handed in late. Assignments handed in more than 1 class late will receive no credit. *This policy will be strictly enforced.*
- Web Site:** We will be using Stellar <<http://stellar.mit.edu/S/course/15/fa08/15.871ab>> to post course materials online. Non-MIT students can access Stellar after being added by the course administrator. The site contains the syllabus, assignments, simulation models, reading list, helpful hints, software access, and other useful information. We will use it to send emails with information such as hints for assignments, schedule changes for TA sessions, etc. You can also use the site to find partners for group assignments, or to pose questions to the class as a whole.
- Handouts:** Available on the class Stellar site. Any extra hard copies will be available outside the instructors' offices.

Objectives and Scope

Why do so many business strategies fail? Why do so many others fail to produce lasting results? Why do many businesses suffer from periodic crises, fluctuating sales, earnings, and morale? Why do some firms grow while others stagnate? And how can a firm identify and design high-leverage policies, policies that are not thwarted by unanticipated side effects?

Accelerating economic, technological, social, and environmental change challenge managers to learn at increasing rates. And we must increasingly learn how to design and manage complex systems with multiple feedback effects, long time delays, and nonlinear responses to our decisions. Yet learning in such environments is difficult precisely because we never confront many of the consequences of our most important decisions. Effective learning in such environments requires methods to develop systems thinking, to represent and assess such *dynamic complexity* – and tools managers can use to accelerate learning throughout an organization.

15.871 and 872 introduce you to system dynamics modeling for the analysis of business policy and strategy. You will learn to visualize a business organization in terms of the structures and policies that create dynamics and regulate performance. System dynamics allows us to create ‘microworlds,’ management flight simulators where space and time can be compressed, slowed, and stopped so we can experience the long-term side effects of decisions, systematically explore new strategies, and develop our understanding of complex systems. We use simulation models, case studies, and management flight simulators to develop principles of policy design for successful management of complex strategies. Case studies of successful strategy design and implementation using system dynamics will be stressed. We consider the use of systems thinking to promote effective organizational learning.

The principal purpose of modeling is to improve our understanding of the ways in which an organization's performance is related to its internal structure and operating policies as well as those of customers, competitors, suppliers and other stakeholders. During the course you will use several simulation models to explore such strategic issues as fluctuating sales, production and earnings; market growth and stagnation; the diffusion of new technologies; the use and reliability of forecasts; the rationality of business decision making; and applications in health care, energy policy, environmental sustainability, and other topics.

Students will learn to recognize and deal with situations where policy interventions are likely to be delayed, diluted, or defeated by unanticipated reactions and side effects. You will have a chance to use state of the art software for computer simulation and gaming. Assignments give hands-on experience in developing and testing computer simulation models in diverse settings.

No prior computer modeling experience is needed.

Those on the wait list, those who did not register through the Sloan bidding system, and listeners are welcome only if space permits (in that order).

Texts and Software

Required Text:

1. Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World* (Text and CD-ROM). Irwin/McGraw Hill. ISBN 0-07-238915X. (Available at the MIT Coop.)
2. Occasional articles and case studies (to be made available via Stellar).

The syllabus notes the days for which these readings should be prepared (NOTE: *before* the class in which we discuss them). Additional readings will be handed out on an occasional basis. The syllabus also indicates which sections of the text you should be sure to read to learn the material you will need to do the assignments, and which sections you can skim (NOTE: 'skim' ≠ 'skip').

In addition, we will be using modeling software. Several excellent packages for system dynamics simulation are available commercially, including **iThink**, from High Performance Systems, **Powersim**, from Powersim Corporation, and **Vensim**, from Ventana Systems. All are highly recommended. You may wish to learn more about these packages, as all are used in the business world, and expertise in them is increasingly sought by potential employers. For further information, see the following resources:

iThink: See the Isee Systems web site at <www.iseesystems.com>.

Powersim: See the Powersim web site at <www.powersim.com>.

Vensim: See the Ventana Systems web site at <www.vensim.com>.

In this course, we will be using the **Vensim** Personal Learning Edition (**VensimPLE**) by Ventana Systems. VensimPLE is free for academic use. VensimPLE is available for Windows only. However, Mac users with Intel-based Macs can easily run Vensim using a PC emulator such as Parallels, VMWare, or Darwin. VensimPLE comes with on-line user's guide and help, and also a folder of demo models. Download VensimPLE from <www.vensim.com/venple.html>.

NOTE: The disc that comes with the *Business Dynamics* textbook includes a version of VensimPLE. However, the version available online is newer and has enhanced functionality. Be sure to download the current version from the Vensim website above. All the Vensim models on the text CD work with the new version.

15.871/15.872 SCHEDULE

Date		Class	Topic	Reading Due	Assn Out	Assn Due
9/3	W	1	Introduction: Purpose, tools and concepts of system dynamics	Read <i>Business Dynamics</i> [BD], Ch. 1	#1	
9/8	M	2	System Dynamics Tools Part 1: Problem definition and model purpose; intro to causal mapping	Read BD, Ch. 3, Ch. 4		
9/10	W	3	System Dynamics Tools Part 2: Building theory with causal loop diagrams	Read BD, Ch. 5 (Skim sections 5.4, 5.6)	#2	#1
9/15	M	4	System Dynamics Tools Part 3: Mapping the stock and flow structure of systems	Read BD, Ch. 6 (Skim sections 6.2.7, 6.2.8, 6.2.9, 6.3.4, 6.3.6)		
9/17	W	5	System Dynamics Tools Part 4: Dynamics of stocks and flows	Read BD, Ch. 7	#3	#2
9/22	M		<i>NO CLASS: MIT HOLIDAY</i>			
9/24	W	6	Growth Strategies Part 1: Modeling innovation diffusion and the growth of new products	Read BD, Ch. 8; Ch. 9.1 (Skim 9.1.2, 9.1.3); 9.2, 9.3 (Skim sections 9.3.5 - end)		
9/29	M	7	Growth Strategies Part 2: Network externalities, complementarities, and path dependence	Read BD Ch. 10 (Skim section 10.2)	#4	#3
10/1	W	8	Growth Strategies Part 3: Modeling the evolution of new medical technologies	<i>Please Prepare:</i> Homer 1996/1984, “The Evolution of a Radical New Technology: The Implantable Cardiac Pacemaker”		

Date		Class	Topic	Reading Due	Assn Out	Assn Due
10/6	M	9	Interactions of Operations, Strategy, and Human Resource Policy: People Express	<i>Please Prepare:</i> People Express (A)	#5	#4
10/8	W	10	<i>Guest Lecture:</i> System Dynamics at General Motors (Dr. Mark Paich)	TBA		
10/13	M		<i>NO CLASS: Columbus Day Holiday</i>			
10/15	W	11	Managing Hyper Growth: Lessons from People Express. END OF 15.871	TBA		#5
10/20-10/24			<i>Sloan Innovation Period:</i> No Classes			
10/27	M		15.872 begins: see next page			

NOTE ON ACADEMIC STANDARDS

We expect the highest standards of academic honesty and behavior from all participants in class. The course Stellar site, <<http://stellar.mit.edu/S/course/15/fa08/15.871ab>>, contains an important document describing academic standards at MIT Sloan. The document discusses standards for citing the work of others (proper referencing to avoid plagiarism), and standards for individual and group work. Please be sure to read this document. If you have any questions about standards and expectations regarding individual and team assignments, please ask us after you have read the standards and before doing the assignments.

15.872 SCHEDULE

Date		Class	Topic	Reading Due	Assn Out	Assn Due
10/27	M	1	System Dynamics in Action: Re-engineering the supply chain in a high-velocity industry	Read BD, Ch. 11 (Skim sections 11.6, 11.7).	#1	
10/29	W	2	Managing Instability Part 1: Formulating and testing robust models of business processes	Read BD, Sections 13.1, 13.2.1-13.2.9, 13.3 and 13.4		
11/3	M	3	Managing Instability Part 2: The Beer Game (Bullwhip) Effect	Read BD, Sections 17.1, 17.2 and 17.3	#2	#1
11/5	W	4	Managing Instability Part 3: Forecasting and Feedback: how (not) to forecast	Read BD, Ch. 16		
11/10	M		<i>NO CLASS: MIT HOLIDAY</i>			
11/12	W	5	Cutting corners and working overtime: Service quality management	Read BD, Sections 14.1-14.4	#3	#2
11/17	M	6	Managing Instability Part 4: Business cycles, real estate crises and speculative bubbles	Read BD, Sections 17.4 and 17.5		
11/19	W	7	<i>Guest Lecture:</i> Jay W. Forrester	Read Forrester, <i>From the Ranch to System Dynamics: An Autobiography</i>		
11/24	M	8	System Dynamics in Action: Applications of System Dynamics to Environmental and Public Policy Issues	Read Meadows, "The Global Citizen" (selections)		

Date		Class	Topic	Reading Due	Assn Out	Assn Due
11/26	W	9	Process Improvement and the dynamics of organizational change	TBA	#4	#3
12/1	M	10	Overcoming the service quality death spiral	TBA		
12/3	W	11	Late, expensive, and wrong: The dynamics of project management	Read BD, Sections 2.3 and 6.3.4		
12/8	M	12	Project management (cont.): Firefighting in new product development	TBA		
12/10	W	13	System Dynamics in Action: The implementation challenge Conclusion: How to keep learning. Follow-up resources. Career opportunities. Course evaluations	Read BD, Ch. 22		#4

NOTE ON ACADEMIC STANDARDS

We expect the highest standards of academic honesty and behavior from all participants in class. The course Stellar site, <<http://stellar.mit.edu/S/course/15/fa08/15.871ab>>, contains an important document describing academic standards at MITSloan. The document discusses standards for citing the work of others (proper referencing to avoid plagiarism), and standards for individual and group work. Please be sure to read this document. If you have any questions about standards and expectations regarding individual and team assignments, please ask us after you have read the standards and before doing the assignments.

System Dynamics Group
Sloan School of Management
Massachusetts Institute of Technology

15.871 Introduction to System Dynamics

Fall 2008

Professor John Sterman

Assignment 3

Mapping the Stock and Flow Structure of Systems

Assigned: Wednesday 17 September 2008; Due: Monday 29 September 2007

Please do this assignment in a group totaling three people.

This assignment will give you practice with the structure and dynamics of stocks and flows. Stocks and flows are the building blocks from which every more complex system is composed. The ability to identify, map, and understand the dynamics of the networks of stocks and flows in a system is essential to understanding the processes of interest in any modeling effort.

☛ **To do this assignment effectively be sure to read *Business Dynamics*, ch. 6 and 7.**

A. Identifying Stock and Flow Variables

The distinction between stocks and flows is crucial for understanding the source of dynamics. In physical systems it is usually obvious which variables are stocks and which flows. In human and social systems, often characterized by intangible, “soft” variables, identification is more difficult.

☐ A1. For each of the following variables, state whether it is a stock or a flow, and give units of measure for each.

Name	Type	Units
<i>Example:</i> Inventory of beer	Stock	Cases
<i>Example:</i> Beer order rate	Flow	Cases/week

	Name	Type	Units
a.	Company Revenue		
b.	Customer service calls on hold at your firm's call center		
c.	GDP (Gross Domestic Product)		
d.	US trade deficit		
e.	Products under development		
f.	Employee Experience		
g.	Corporate accounts receivable		
h.	Book value of inventory		
i.	Promotion of Senior Associates to Partner at a consulting firm		
j.	Incidence of attacks on corporate web sites		
k.	Greenhouse gas emissions of the US		
l.	Euro/dollar exchange rate		
m.	Employee morale		
n.	Interest Rate on 30-year US Treasury Bond		
o.	Your firm's cost of goods sold (COGS)		

B. Mapping Stock and Flow Networks

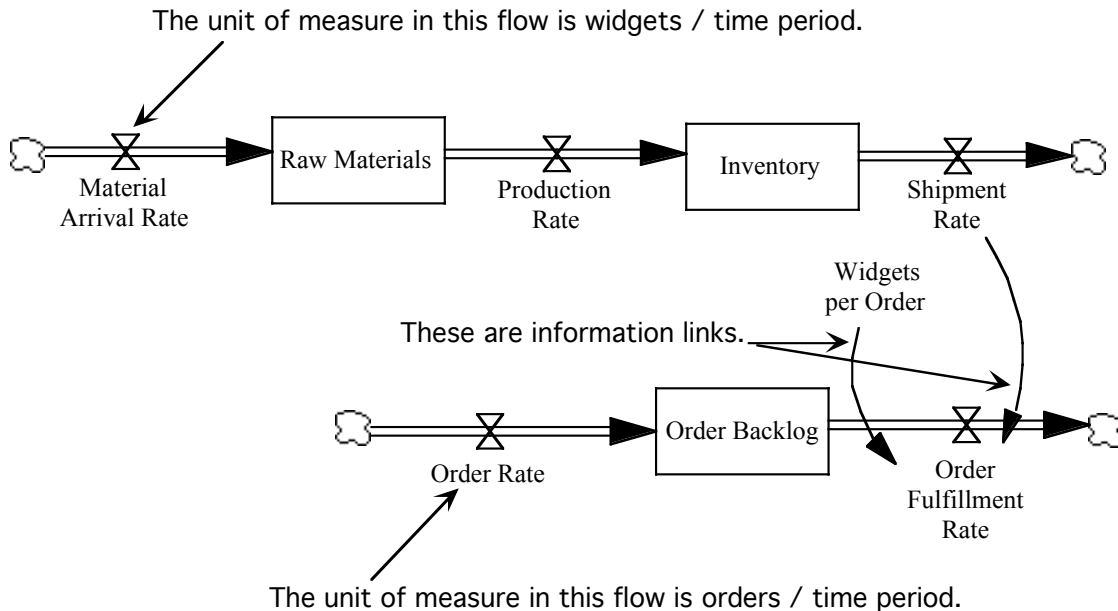
Systems are composed of interconnected networks of stocks and flows. Modelers must be able to represent the stock and flow networks of people, material, goods, money, energy, etc. from which systems are built.

For each of the following cases, construct a stock and flow diagram that properly maps the stock and flow networks described.

- Not all the variables are connected by physical flows; they may be linked by information flows, as in the example below.
- You may need to add additional stocks or flows beyond those specified to complete your diagram (but keep it simple). Be sure to consider the boundary of your stock and flow map. That is, what are the sources and sinks for the stock and flow networks? Are you tracking sources and sinks far enough upstream and downstream? This process of deciding how far to extend the stock and flow network is called “challenging the clouds” because you question whether the clouds are in fact unlimited sources or sinks.
- Consider the units of measure for your variables and make sure they are consistent within each stock and flow chain.

Example: A manufacturing firm maintains an inventory of finished goods from which it ships to customers. Customer orders are filled after a delay caused by order processing, credit checks, etc. Map the stock and flow structure, drawing on the following variables: Inventory, Raw Materials, Production, Order Backlog, Order Rate.

Solution:



Comment: There are two linked stock and flow networks here: first, the physical flow of materials as they are fabricated into products and shipped to customers; second, the flow of orders. The two networks are linked because there is a direct relationship between physical shipments and order fulfillment (assuming no accounting glitches or inventory shrinkage!)—every time a product is physically shipped, the order is removed from the backlog and denoted as filled. The link between the Shipment Rate and Order Fulfillment Rate is an information link, not a material flow. Note that considering the units of measure helps identify the linkages between the two stock and flow chains. The units of all flows in the materials chain are widgets/time period, and the units of the materials and inventory stocks are widgets. The units of the order flows are orders/time period. The order fulfillment rate is then given by the number of widgets shipped per period divided by the number of widgets per order, to yield orders/time period for the order fulfillment rate. Note also that only the information links directly connecting the stock and flow networks are captured. Other information links that must exist are not represented. For example, the shipment rate must depend on the finished goods inventory (no inventory—no shipments). The purpose of this exercise, however, is to map the stocks and flows, so these feedbacks can be omitted for now. Later you will integrate stock and flow maps with causal-loop diagrams to close the feedback loops in a system. Note that the shipment rate, material arrival rate, and order fulfillment rate were not included in the group of variables listed in the description but must be introduced to complete the stock-and-flow network. Note also that the solution omits some structure that might be added if the purpose of the model required it—for example, inventory shrinkage and order cancellation flows, and the installed base of product (the stock filled up by shipments). The model could be disaggregated further, e.g., splitting the order backlog into two stocks, “orders awaiting credit approval,” and “orders approved.” The choice of detail is always governed by the purpose of the model.

- B1. A computer manufacturer maintains a large call center operation to handle customer inquiries. Customers with questions or problems call a toll free number for help. In this firm, incoming calls are answered by a voice recognition system that routes calls, based on the customer's choice, either to an automated system or to a live customer service agent (CSA). Callers choosing to work their way through the automated help process can, at any time, press "0" to speak to an agent, or, of course, hang up. Callers electing to speak to a CSA may be placed on hold until an agent becomes available. If the call is answered before the customer gets frustrated and hangs up, the CSA may be able to resolve the issue. Often, however, the CSA is unable to solve the problem and forwards the call to a supervisor or specialized department such as technical support. The issue may or may not be resolved by these specialists. Map the stock and flow structure of calls as they flow through the system.
- ☛ In reality, customer inquiries arrive by phone, by email, and by live chat from the firm's website. You don't need to consider these channels separately. Likewise, do not attempt to separate inbound calls into different categories such as billing problems or tech support questions. Assume there is a single flow of calls coming in to the system. These calls are then divided into those electing the automated system and those electing to speak to an agent.
- B2. The ability of the firm above to answer calls quickly depends on the size and skill of their CSA staff. Map the stock and flow structure for the number of CSAs. In mapping the stocks of CSAs, distinguish between "generalists" and "specialists". Generalists are the front line agents who initially field calls; specialists are the tech support and other more highly trained people who handle the more complex inquiries generalists are unable to resolve. Call center work is stressful and turnover among both types of CSAs is high. Further, new hires are inexperienced and less productive; these are known in the firm as "rookies." Many rookies quit before they become experienced. The firm does not hire into specialist CSA positions from outside; rather, they promote some of the experienced generalists into the better-paid specialist positions.
- ☛ Such firms maintain many call centers around the world (Dell, for example, has roughly 27,000 CSAs located in dozens of call centers around the world). However, you should aggregate all such centers into a single category.
- B3. Map the stock and flow structure for the adoption and diffusion of new products. To provide a concrete context, consider the adoption of DVD players in the United States. Initially, before DVDs and DVD players were developed, everyone in the US was unaware that such an innovation existed. After DVD players were introduced to the market, people moved through various stages. Some gradually became aware of the product. Some may then enter the market (actively seeking information about different models, prices and features). Some of these people decide to buy a unit, thus becoming an adopter of the innovation. Many adopters are happy with their purchase; they may even replace their first units when they are lost, wear out, or become obsolete. Other people may decide they don't get enough benefits from the product and don't replace their initial units, or abandon the DVD if a better product is introduced to the market (e.g., Blu-Ray). Such individuals become former adopters.

- Map two distinct stock and flow chains. The first tracks the flows of people as they move from being unaware through awareness, adoption, and, perhaps, abandonment. The second should track the flows of DVD player purchases and discards. The installed base of a product, while related to the number of adopters, can have different dynamics.
- Show, using information links, how the two stock and flow chains are connected. Specifically, show how purchases and discards are related to the stocks and flows of people as they move from being unaware to adoption.
- Challenge the clouds. What happens to the old units people discard?

C. Dynamics of Accumulation

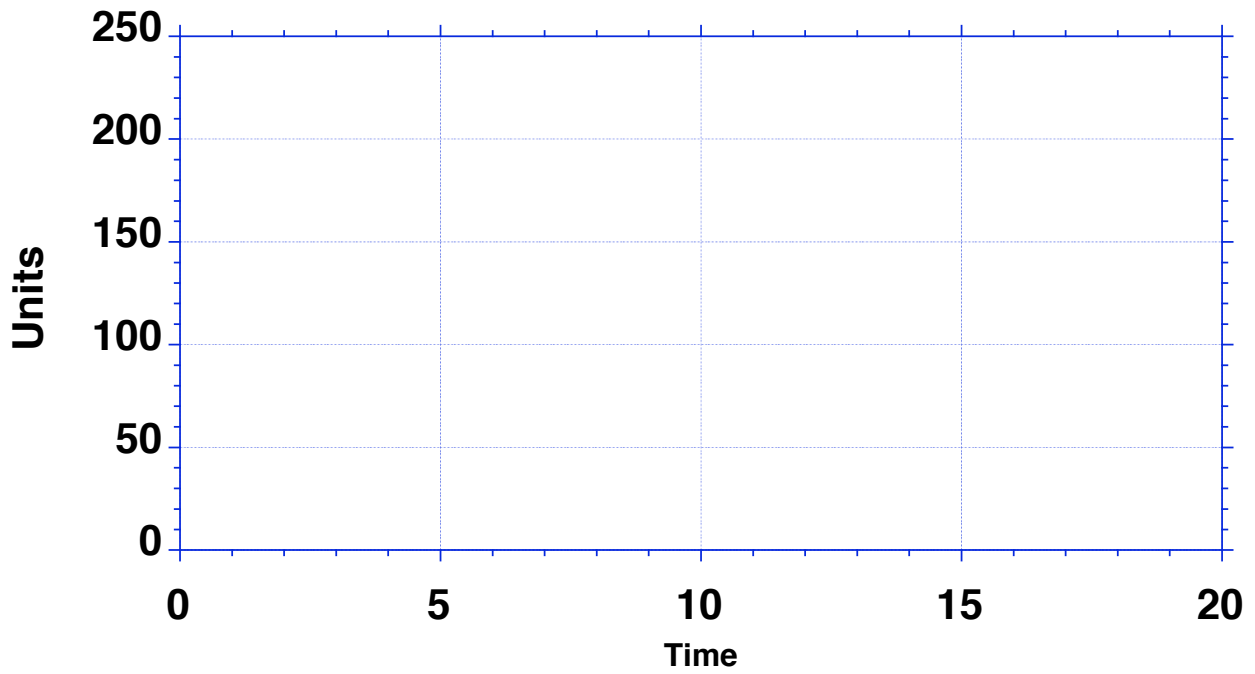
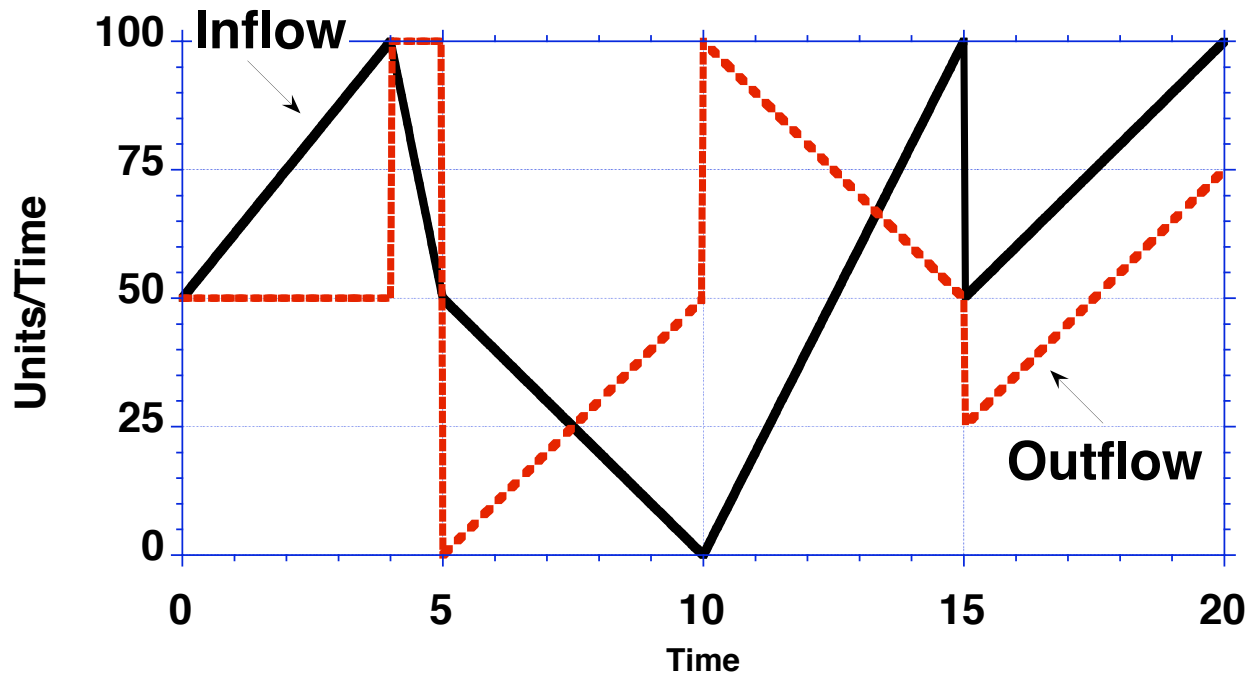
Stocks are accumulations. The difference between the inflows and outflows of a stock accumulates, altering the level of the stock variable. The process of accumulation gives stocks inertia and memory and creates delays. Since realistic models are far too complex to solve with formal analysis, it is important to understand the relationship between flows and the behavior of stocks intuitively.

- **The goal is to develop your intuition about stocks and flows. Be sure to read Chapter 7 first.**

□ C1. Consider the following system:



The top graph on the next page shows the behavior of the inflow and outflow for the stock. On the graph provided below, draw the trajectory of the stock given the inflow and outflow rates shown. Indicate the numerical values for any maxima or minima, and for the maximum or minimum values of the slope for the stock. Assume the initial quantity in the stock is 100 units.



D. Linking stock and flow structure with feedback

Now we will simulate a simple stock flow system with feedback. Build and simulate a simple model of the US national debt and budget deficit.

- ☛ Follow the instructions below precisely. Do not add structure beyond that specified.
- ☛ Begin the simulation of the model in 1988 so that there is some replication of history. In Vensim, Select **Settings...** under the **Model** menu. Then set the Initial Time = 1988, Final Time = 2088, and Time Step = 0.0625 years. Check the box to save the results every Time Step. Finally, set the unit of measure for time to Years.
- ☛ To keep your model simple:
 - Your model should have a single stock, the National Debt. The debt accumulates the Net Federal Deficit. The only flow altering the debt is the net deficit (do not represent the issuance and maturity of the debt). In 1988 the national debt was approximately \$2.5 trillion (2.5E12).
 - The net federal deficit is the difference between Government Expenditure and Government Revenue.
 - Government Revenue is exogenous and constant. In 1988, revenue was approximately \$900 billion/year (900E9).
 - Government Expenditure consists of Interest paid on the debt and Expenditures on Programs (all non-interest expenditures).
 - Expenditures on Programs are exogenous and constant. In 1988 expenditures on programs were about \$900 billion/year, about the same as Revenue.
 - Interest payments are the product of the debt and the interest rate.
 - The interest rate is exogenous and constant. In 1988 the average interest rate on the debt was approximately 7%/year (.07/year).
- ☛ As always, document your model and make sure every equation is dimensionally consistent.

Answer the following questions.

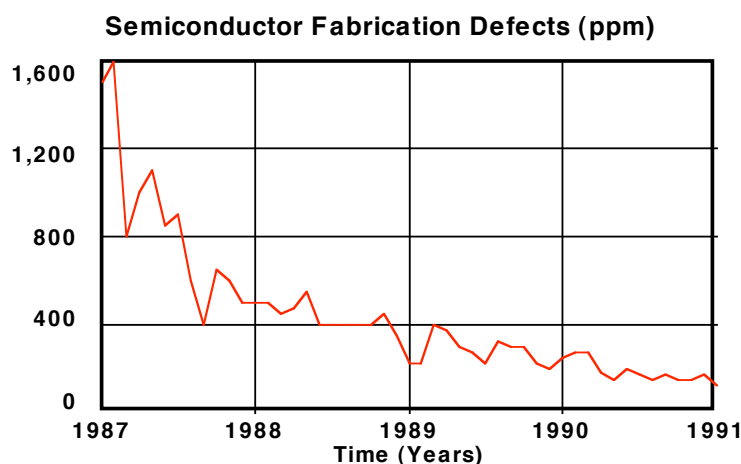
- a. What kind of feedback loop is created in your model?
- b. What is the initial deficit (given the base case parameters)?
- c. How long does it take for the deficit to double?
- d. What is the relationship between the doubling time and the interest rate? (To discover a relationship, you may want to simulate with extreme interest rates—say, between 1% per year and 15% per year).
- e. Hand in your model (diagram and equation listing) and answers to the above questions. You need not hand in plots, but you should describe briefly how you arrived at your answers.

E. Modeling Goal-Seeking Processes

All goal-seeking processes consist of negative feedback loops. In a negative loop, the system state is compared to a goal, and the gap or discrepancy is assessed. Corrective actions respond to the sign and magnitude of the gap, bringing the state of the system in line with the goal.

For example, consider programs designed to improve the quality of a process in a company. The process could be in manufacturing, administration, product development—any activity within the organization. Improvement activity is iterative. Members of an improvement team identify sources of defects in a process, often ranking benefits of correcting them using a Pareto chart. They then design ways to eliminate the source of the defect, and try experiments until a solution is found. They then move on to the next most critical source of defects. Quality professionals refer to this iterative cycle as the “Plan—Do—Check—Act” or “PDCA” cycle (also known as the Deming cycle, for the late quality guru W. Edwards Deming). In the PDCA process, the improvement team: (1) plans an experiment to test an improvement idea, (2) does the experiment, (3) checks to see if it works, then (4) acts—either planning a new experiment if the first one failed or implementing the solution and then planning new experiments to eliminate other sources of defects. The team continues to cycle around the PDCA loop, successively addressing and correcting root causes of defects in the process. This learning loop is not unique to TQM: All learning and improvement programs, including 6-s, follow an iterative process similar to the PDCA cycle.

The figure below shows data on defects from the wafer fabrication process of a mid-size semiconductor firm (from Figure 4-5 in *Business Dynamics*). The firm began its TQM program in 1987, when defects were running at a rate of roughly 1500 parts per million (ppm). After the implementation of TQM, the defect rate fell dramatically, until by 1991 defects seem to reach a new equilibrium close to 150 ppm—a spectacular factor-of-ten improvement. Note that the decline is rapid at first, then slows as the number of defects falls.



- E1. Create a model of the improvement process described above and compare its behavior to the data for the semiconductor firm. Once you have formulated your model, make sure the units of each equation are consistent. Hand in the diagram for your model and a documented model listing.

- * Follow the instructions below precisely. Do not add structure beyond that specified.
- The state of the system is the defect rate, measured in ppm. The defect rate in 1987 was 1500 ppm.
 - * The defect rate is not a rate of flow, but a stock characterizing the state of the system—in this case, the ratio of the number of defective dies to the number produced.
- The defect rate decreases when the improvement team identifies and eliminates a root cause of defects. Denote this outflow the “Defect Elimination Rate.”
- The rate of defect elimination depends on the number of defects that can be eliminated by application of the improvement process and the average time required to eliminate defects.
- The number of defects that can be eliminated is the difference between the current defect rate and the theoretical minimum defect rate. The theoretical minimum rate of defect generation varies with the process you are modeling and how you define “defect.” For many processes, the theoretical minimum is zero (for example, the theoretical minimum rate of late deliveries is zero). For other processes, the theoretical minimum is greater than zero (for example, even under the best imaginable circumstances, the time required to build a house or the cycle time for semiconductor fabrication will be greater than zero). In this case, assume the theoretical minimum defect level is zero.
- The average time required to eliminate defects for this process in this company is estimated to be about 0.75 years (9 months). The average improvement time is a function of how much improvement can be achieved on average on each iteration of the PDCA cycle, and by the PDCA cycle time. The more improvement achieved each cycle, and the more cycles carried out each year, the shorter the average time required to eliminate defects will be. These parameters are determined by the complexity of the process and the time required to design and carry out experiments. In a semiconductor fab, the processes are moderately complex and the time required to run experiments is determined by the time needed to run a wafer through the fabrication process. Data collected by the firm prior to the start of the TQM program suggested the 9 month time was reasonable.
- Equipment wear, changes in equipment, turnover of employees, and changes in the product mix can introduce new sources of defects. The defect introduction rate is estimated to be constant at 250 ppm per year.

E2. Run your model with the base case parameters, and *hand in the plot*.

a. *Briefly* describe the model’s behavior.

b. How well does your simulation match the historical data? Are the differences likely to be important if your goal is to understand the dynamics of process improvement and to design effective improvement programs?

c. Does the stock of defects reach equilibrium after 9 months (the average defect elimination time)? Referring to the structures in your model, explain why or why not.

- E3. Experiment with different values for the average defect elimination time. What role does the defect elimination time play in influencing the behavior of other variables?
- E4. The stock reaches equilibrium when its inflows equal its outflows. Set up that equation and solve for the equilibrium defect rate in terms of the other parameters.
 - a. What determines the equilibrium (final) level of defects? Why?
 - b. Does the equilibrium defect rate depend on the average time required to eliminate defects? Why/Why not?
- E5. Explore the sensitivity of your model's results to the choice of the time step or "dt" (for "delta time").
 - * Before doing this question, read Appendix A in *Business Dynamics*.
 - a. Change the time step for your model from 0.125 years to 0.0625 years. Do you see a substantial difference in the behavior?
 - b. What happens when dt equals 0.5 years? Why does it behave as it does?
 - c. What happens when dt equals 1 year? Why does the simulation behave this way?

Appendix 2. Impact of demographics on performance

Appendix 2 shows the impact of subject demographics on performance in both the pre-test and post-test. Table S-1 shows the significance levels of tests of each individual demographic variable on the fraction correct for each of the four questions in the pre-test. Considering the first two questions, which assess whether subjects can interpret the graph, none of the demographic variables have a statistically significant impact on the fraction correct, with the exception of English as a native language for Q1 only ($p = .044$). There is, however, no plausible reason for native language to matter for the question of when the most people left the store but not for when the most people entered the store. On the two stock-flow questions (Q3 and Q4, most and fewest in the store, respectively), age, work experience, English as a native language, prior experience with the beer game, and participation in the half-day climate change workshop have no significant effects on performance. However, there is a highly significant gender effect, with males outperforming females ($p < .0001$). The degree program in which the student is enrolled has a marginally significant effect for both questions. The highest prior degree has at best a marginal effect on Q4 only, and the prior field of study (STEM, social science, humanities, or architecture) has a significant effect on Q3 only. Table S-1 also shows the Spearman rank correlations among responses on each of the pre-test questions. As one would expect, correct responses on Q1 and Q2 are highly correlated ($r = .68, p < .0001$): if one cannot determine when the most people enter the store, one is also unlikely to know when the most are leaving. Also as expected, correct responses on the two stock-flow questions (Q3 and Q4) are highly correlated ($r = .67, p < .0001$): if one cannot determine when the most people are in the store, one is also unlikely to know when the fewest are in the store. Performance on the graphical interpretation questions tends to improve performance on the stock-flow questions, but much more weakly, and the impact is statistically significant only for the correlation of Q1 and Q3: the ability to read the graph is necessary but far from sufficient to understand the stock-flow structure of the task.

Table S-1. Impact of subject demographics on pre-test performance. Entries are the significance levels (p -values) for sex, English, beer game, climate change workshop from 2-sided Wilcoxon test; for program, highest prior degree and field from Kruskal-Wallis test; for Age and Work Experience from the χ^2 test of the likelihood ratio derived from univariate logistic regression. Bold values show $p < .05$.

	Q1 Most Entering	Q2 Most Leaving	Q3 Most in Store	Q4 Fewest in Store
Age	.575	.512	.541	.790
Work Exp.	.858	.094	.824	.874
Sex	.341	1.00	< .0001	< .0001
English	.044	.299	.738	.850
Program	.420	.686	.050	.063
Highest Prior Degree	.984	.534	.220	.099
Field	.715	.825	.025	.161
Beer Game	.292	.740	.170	.448
Climate Change Workshop	.414	.574	.915	.277
Spearman Correlations				
Q1		.683 <i>p < .0001</i>	.153 <i>p = .015</i>	.098 <i>p = .119</i>
Q2			.210 <i>p = .0008</i>	.112 <i>p = .074</i>
Q3				.671 <i>p < .0001</i>

Table S-1 shows the results of univariate tests; a more appropriate test would account for the relationships among the different demographic variables. Table S-2 reports multivariate logistic models with the fraction correct on Q1-4 as the dependent variable. Results show even smaller impact of demographics than revealed by the univariate tests. For Q1, none of the demographic variables are statistically significant at $p < .05$. For Q2, age and work experience are significant at $p < .05$. For Q3 and Q4, only sex had a significant effect ($p < .001$ for both). Results were similar when the logistic regression excluded those variables the univariate analysis suggested had no impact (i.e., age, work experience, native language, beer game experience and participation in the climate change workshop).

Table S-2. Statistical significance of subject demographics on pre-test performance, multivariate logistic regression. Entries are p -values for each effect. Values of $p < .05$ in **bold** print.

	Q1 Most Entering	Q2 Most Leaving	Q3 Most in Store	Q4 Fewest in Store
Age	.110	.026	.472	.677
Work Exp.	.153	.012	.920	.798
Sex	.485	.511	.0005	.0007
English	.059	.165	.536	.463
Program	.683	.241	.250	.389
Highest Prior Degree	.997	.300	.885	.752
Field	.924	.943	.288	.689
Beer Game	.829	.275	.183	.205
Climate Change Workshop	.501	.683	.613	.114
N	240	240	240	240

Turning to performance on the post-test, Table S-3 shows the impact of subject demographics on post-test performance. Univariate tests show statistically significant effects of sex, field of study, and of course which of the eight patterns of people entering and leaving the subject received. However, as in the case of the pre-test, multivariate logistic regression shows even weaker effects of the demographics. Table S-3 shows the significance levels for the

demographic and other variables in a set of logistic regressions incorporating combinations of the demographic variables, the pattern of inflow and outflow received in the post-test, and performance on each of the four pre-test questions. As in the pre-test, there is a strong effect gender, but the effects of age, work experience, native language, beer game experience and participation in the climate change workshop in which stock and flow concepts were discussed are not statistically significant in predicting success on the post-test. The effects of the degree program in which the student is enrolled, highest prior degree, and field of study (STEM, social science, humanities, or architecture) also were not significant.

With so many correlated regressors, the validity of the logistic regression is questionable, so successive models eliminated variables that appeared to offer no explanatory power. The results remain similar. Which of the eight patterns the subject received is always highly significant, along with the impact of gender, with males outperforming females.

While the univariate Wilcoxon tests show that correctly responding on the two stock-flow questions in the pre-test does predict post-test success, the effects are not robust in the logistic regressions. Responding correctly on pre-test Q3 (when are the most in the store?) does improve the odds of success in the post-test, but the effect is marginally significant. Performance on pre-test Q4 (when are the fewest in the store?) is not significant. As a final test, Table S-3 also reports tests the extent to which those getting both pre-test Q1 and Q2 correct, and/or Q3 and Q4 correct, predicts post-test performance. As expected, the impact of getting both graph interpretation questions correct is not significant. Getting both Q3 and Q4 correct, indicating those with the best grasp of stock-flow concepts, does predict post-test performance slightly, but the effect is not statistically significant.

Table S-3. Determinants of performance on graphical department store task. Univariate *p*-values from logistic regression for Age and Work Experience, from nonparametric tests (Wilcoxon or Kruskal-Wallis) for all other variables.

	Post-test, % incorrect						
	Univariate effects (<i>p</i>)	Logistic Regression (<i>p</i>)					
Age	.775	.338					
Work Exp.	.762	.801					
Sex	.003	.017	.027	.051	.032	.017	.017
English	.534	.803					
Program	.201	.866	.916				
Highest Prior Degree	.131	.835	.664				
Field	.039	.368	.431				
Beer Game	.853	.443					
Climate Change Workshop	.563	.868					
Condition (1-8)	< .0001	.004	.003	.001	.002	.002	.002
Pre-test Q1 correct	.256	.223	.401	.270			
Pre-test Q2 correct	.355	.349	.512	.371			
Pre-test Q3 correct	.002	.076	.083	.054	.057		
Pre-test Q4 correct	.002	.836	.786	.852	.854		
Pre-test Q1 and Q2 correct						.512	
Pre-test Q3 and Q4 correct						.110	.081