Boom, bust, and failures to learn in experimental markets

Mark Paich, 719-389-6445
Colorado College

and

John D. Sterman, 617-253-1951
MIT Sloan School of Management

Working Paper 3441-92-BPS July, 1992

Boom, bust, and failures to learn in experimental markets

Mark Paich* and John D. Sterman§

Abstract

Boom and bust is a pervasive dynamic for new products. Word of mouth, marketing, and learning curve effects can fuel rapid growth, often leading to overcapacity, price war, and bankruptcy. Previous experiments suggest such dysfunctional behavior can be caused by systematic 'misperceptions of feedback', where decision makers do not adequately account for critical feedbacks, time delays, and nonlinearities which condition system dynamics. However, prior studies often failed to vary the strength of these feedbacks as treatments, omitted market processes, and failed to allow for learning. A decision making task portraying new product dynamics is used to test the theory by varying the strength of key feedback processes in a simulated market. Subjects performed the task repeatedly, encouraging learning. Nevertheless, performance relative to potential is poor and is severely degraded when the feedback processes in the environment are strong, supporting the misperception of feedback hypothesis. The negative effects of feedback complexity on performance were not moderated by experience, even though average performance improved. Models of the subjects' decision making heuristics are estimated; changes over trials in estimated cue weights explain why subjects improve on average but fail to gain insight into the dynamics of the system. Though conditions for learning are excellent, experience does not appear to mitigate the misperceptions of feedback or systematic dysfunction they cause in dynamic decision making tasks.

KEYWORDS: Decision making, simulation, feedback, experimental economics, system dynamics

* Department of Economics, Colorado College, Colorado Springs, CO 80903

§ Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139

Please direct correspondence to John Sterman at the address above or jsterman@mit.edu.

The comments of John Carroll, Don Kleinmuntz, Rebecca Henderson, Rogelio Oliva, Anjali Sastry, John Seeger and seminar participants at MIT are gratefully acknowledged. This work was supported in part by the Organizational Learning Center, MIT Sloan School.

Boom and bust is a pervasive dynamic for new products. Sales of new products often grow at rapid exponential rates as word of mouth, advertising, and falling prices attract new buyers. New producers tend to enter the market. But eventually the stock of potential purchasers is depleted and sales fall to an equilibrium determined by replacement needs. During the transition to replacement demand producers often suffer large losses due to excess capacity and falling prices, stimulating exit (Gort and Klepper 1982, Klepper and Graddy 1990).

As a typical example, figure 1 shows the sales and net income of Atari, the leader of the first wave of video games. Atari, then a division of Warner Communications, roughly doubled its sales each year, from $35 million in 1976 to over $2 billion in 1983. Profit from operations reached $323 million in 1983. But within a year sales plummeted as both home and arcade markets became glutted. Atari lost $539 million in 1983, and was sold for just $160 million in debt, a 32% equity stake, and no cash (Petre 1985). Warner took an additional $592 million charge against 1984 earnings for losses related to the sale.

Porter (1980) describes this pattern of boom, bust, price war and shakeout as a generic feature of industrial dynamics:

> "As [a maturing industry] adjusts to slower growth, the rate of capacity addition in the industry must slow down as well or overcapacity will occur. Thus companies' orientations toward adding capacity and personnel must fundamentally shift and be disassociated from the euphoria of the past.... These shifts in perspective rarely occur in maturing industries, and overshooting of industry capacity relative to demand is common. Overshooting leads to a period of overcapacity, accentuating the tendency during transition toward price warfare." (p. 239)

The boom and bust dynamic appears in diverse industries. "Snowmobiles, hand calculators, tennis courts and equipment, and integrated circuits are just a few" examples cited by Porter (1980). To these can be added VCRs and other consumer electronics, personal computers, toys and games (Beinhocker 1991), bicycles and chain saws (Porter 1983), home furnishings (Salter 1969), and numerous other consumer and industrial goods. This study explores the role of cognitive misperceptions in the genesis and persistence of the boom and bust phenomenon.

The boom and bust dynamic provides a typical example of a dynamic decision making system. Decisions made today alter the environment, giving rise to information upon which tomorrow's decisions are based – the evolution of the system is strongly conditioned by the behavior of the decision makers. Recent studies show, with few exceptions, that decision making in complex dynamic environments is poor relative to normative standards or even relative to simple heuristics, especially when decisions have indirect, delayed, nonlinear, and multiple feedback effects (Diehl 1992, Sterman 1989a, 1989b, Kleinmuntz and Thomas 1987, Kluwe, Misiak, and Haider 1989, Brehmer 1990, Smith, Suchanek, and Williams 1988; Funke 1991 reviews the large literature of the 'German School' led by Dörner, Funke, and colleagues). Sterman (1989a, 1989b) argues that the mental models people use to guide their decisions in dynamic settings are flawed in specific ways: they tend to ignore feedback processes which cause side-effects, they fail to appreciate time delays between action and response and in the reporting of information, and they are insensitive to nonlinearities which may cause the relative importance of different feedback processes to change as a system evolves. Sterman argued that such "misperceptions of feedback" cause systematically dysfunctional behavior in dynamically complex settings.

However, prior work is limited in several respects. In many studies feedback complexity was not varied as an experimental treatment; other factors might have been responsible for subjects' poor performance. Market institutions, argued by many to provide incentives and means to overcome individual departures from rationality (Hogarth and Reder 1987), have not been included in most studies of dynamic decision making (but see Kampmann and Sterman 1992). Many studies report the results of first trials in which subjects had little opportunity for learning. In others (Brehmer 1990, Kleinmuntz and Thomas 1987, Broadbent and Aston 1978 and many of the studies surveyed in Funke 1991), subjects had little or no prior training or experience relevant to the task (fighting forest fires, treating disease, running a national economy or managing an ecosystem). While good arguments can be made that 'real life' is more like the first trial in such experiments than the last (Camerer 1987), the robustness of the misperception of feedback phenomenon to opportunities for learning has largely gone untested.

The present experiment addresses many of the limitations of earlier work. The task – the introduction and management of a new product – is realistic and well matched to the interests and training of the subjects – management school students, most with several years of business experience. The simulated environment includes an explicit market mechanism, including competition and consumer response to prices. Powerful incentives are used to motivate performance. The misperception of feedback (MOF) hypothesis is tested directly by varying the strength of key feedback relationships across experimental conditions. If subjects are prone to misperceptions of the feedback environment, performance relative to benchmarks should be systematically worse in conditions with high feedback complexity, since these feedbacks will produce consequences unaccounted for by subjects' mental models, and better in environments with only weak feedbacks, since these environments will more closely coincide with their mental models. Further, the subjects performed the task repeatedly, creating opportunities for learning which might improve performance, particularly in conditions of feedback complexity. We describe the task, protocol, and results, analyze the nature of the learning process, and close with discussion of the implications.

The Task: Managing a new product

The task consists of an interactive computer game or "management flight simulator" (Senge and Sterman 1992 and Graham, Morecroft, Senge and Sterman 1992 discuss the design and use of such simulations and contrast them with traditional business games; Sterman 1988 provides another example). The flight simulator embodies a model representing a firm, its market, and its competition. Subjects must manage a new product from launch through maturity, and make pricing and capacity decisions each quarter year through a ten-year simulation.[1]

Market Sector

The market model is based on well known diffusion models in the tradition of Bass (1969), Kalish and Lilien (1986), Mahajan and Wind (1986), Homer (1987), and Mahajan, Muller, and Bass (1990). The essence of these models are the feedbacks driving the adoption process by which potential purchasers become aware of and choose to buy the product (figure 2). Adoption increases the customer base, generating word of mouth which leads to additional sales (a positive feedback process), but also depleting the pool of potential future customers (a negative feedback). The customer base follows a characteristic s-shaped pattern, while sales rise exponentially, then peak and decline to the rate of replacement purchases as saturation sets in.[2] Key features of the market sector include the following:

- Product price affects the number of potential adopters. The elasticity of industry demand is less than unity, quite typical for many goods (Hauthakker and Taylor 1970).
- The greater the aggregated marketing expenditures of the firm and the competition, the larger the fraction of potential customers who purchase each quarter. Diminishing returns set in for high marketing expenditure levels.
- Demand is also generated by word of mouth. Word of mouth is driven by recent purchasers (people who are still excited by the product and have not yet come to take it for granted). The strength of the word of mouth effect (the number of purchases generated per quarter by each recent purchaser) was a treatment variable in the experiment.
- A fraction of the customer base re-enters the market each quarter to replace worn or obsolete units. The repurchase fraction was a treatment variable in the experiment.
- Total orders for the product are divided between the firm and the competition in proportion to the attractiveness of each product. Attractiveness depends on price, availability (measured by delivery delay), and marketing expenditure. While industry demand is relatively inelastic, firm demand is highly but not infinitely elastic – price is important to consumers but availability and marketing can differentiate the two products.

Firm sector

While many diffusion models implicitly equate shipments with orders, the model here explicitly represents the supply side of the market. The key assumptions of the firm sector are:

- Product is built to order. Customer orders flow into a backlog until they are produced and shipped.[3] The firm will ship the current backlog within one period unless capacity is inadequate, in which case the backlog and delivery delay rise, reducing the attractiveness of the firm's product and the share of orders it receives.
- Subjects set a capacity target each quarter. Actual capacity adjusts to the target with a delay representing the time required to plan for, acquire, and ramp up new production facilities. Capacity adjustments follow a distributed lag with a mean of four quarters. Some investments can be real-

ized sooner than four quarters (purchasing additional equipment), while some take longer (building new plant). For simplicity the delay is symmetrical in the case of capacity reduction.

- The firm benefits from a learning curve which reduces unit costs as cumulative production experience grows. A standard "80%" learning curve is assumed – each doubling of cumulative production reduces unit variable costs by 20%. The competitor's learning curve has identical strength. There are no inter-firm learning spillovers.
- Profit is revenue less total costs. Total costs consist of fixed and variable costs, marketing expenditures, and investment costs. Revenues are determined by the quantity shipped in the current quarter and the average price received for those units. Customers pay the price in effect when they booked their order, even if the price has changed in the interim.
- Fixed Costs are proportional to current capacity. Unit fixed costs are constant. Variable costs are proportional to output. Unit variable costs fall as cumulative production increases. Marketing expenditures are set to 5% of revenues.
- Investment costs represent administrative, installation, training, and other costs of increasing capacity; symmetric decommissioning costs are incurred whenever capacity is decreased. Investment costs are proportional to the magnitude of the rate of change of capacity.
- Subjects may lose as much money as they like without facing bankruptcy. The task is therefore more forgiving than reality since losses which would cause bankruptcy in real life can in the game be offset by subsequent profits.

Competitor Structure and Strategy

The subject's firm faces competition from another firm which has launched a similar product at the same time. The playing field is level – the structure and parameters for the firm and its competitor are identical. But while the subjects make price and target capacity decisions for their firm, the competitor's price and target capacity decisions are simulated with simple rules.

The competitor sets target capacity to meet expected orders for its product and to control delivery delays by reducing excessive backlogs. Expected orders are determined by the competitor's current order rate and the expected growth rate of orders. Extrapolative expectations are assumed: the recent growth rate of orders is projected four quarters ahead – the length of the capacity acquisition lag and thus the relevant planning horizon – to account for the growth in demand likely to occur while awaiting delivery of capacity ordered today. To this forecast of future demand is added an adjustment proportional to any excess backlog. If desired production were higher than current capacity, additional capacity would be ordered to reduce the backlog. The decision rule for competitor capacity acquisition has been used extensively in simulation models and is well supported both empirically and experimentally (Senge 1980, Sterman 1987a, 1987b).

Competitor price is set to equal unit costs multiplied by a fixed margin to cover marketing and investment costs, and to provide a normal return when capacity is well utilized. As competitor production experience grows, unit costs fall, and competitor price falls proportionally.

Note that these decision rules are extremely simple. Consistent with theories of bounded rationality (Simon 1982, Morecroft 1985) and experimental evidence (cited above), the competitor relies on locally available information and uses simple rules of thumb. No optimization of investment costs versus opportunity costs of lost revenue is performed in selecting the path of capacity, much less any strategic or game-theoretic reasoning about competitive reactions. One might expect that subjects would easily find ways to exploit the competitor and achieve excellent results. On the other hand, Hogarth and Makridakis (1981) found that subjects in a management game could not differentiate between simulated and human competitors and attributed complex strategic reasoning to 'competitors' whose decisions were largely random.

Hypotheses and Experimental Design

The central issue is the extent to which subject behavior and performance depend on the feedback complexity of the environment. In markets for new products two critical feedback processes involve word of mouth and the average lifetime of the product. Word of mouth creates a powerful positive feedback loop by which recent purchasers of the product generate new purchasers. The stronger the word of mouth feedback, the faster demand grows, the higher it peaks, and the sooner and more suddenly the market declines as the nonlinear transition to saturation sets in. The longer the lifetime of the product, the lower the replacement demand and the larger the "bust" or decline in demand from its peak to equilibrium value as the market for new customers is saturated. We seek to understand whether subjects employ capacity expansion and pricing strategies that are internally consistent and that function well in the rich feedback environment surrounding the diffusion of new products. Are subjects' strategies sensitive to the important feedbacks, time delays, and nonlinearities in the environment? Or do people approach the task with simple mental models which do not adequately account for the these features?

Prior work on misperceptions of feedback (Sterman 1989a, 1989b, Diehl 1992) and on intuitive extrapolation of exponential growth (Wagenaar and Timmers 1979, Wagenaar and Sagaria 1975, Andreassen 1990a, b) predict that performance *relative to potential* should deteriorate as the word of mouth feedback increases in strength and as the product lifetime lengthens. Performance relative to potential is hypothesized to decline as these factors increase not merely because the market will have higher variance and will thus engender larger forecast errors. On the contrary, the MOF hypothesis suggests subjects will approach the decision task with mental models that are insufficiently sensitive to these feedbacks, nonlinearities and delays, and thus make decisions that intensify the problems created by strong word of mouth and long product lifetimes. For example, a learning curve strategy (low prices and aggressive capacity expansion to seek market share advantage and push costs down the learning curve faster than the competitor) may be quite effective in an environment where word of mouth is weak and replacement demand strong, since there will be little overshoot of peak demand relative to equilibrium. However, when these feedbacks are strong, the same strategy might lead to disaster. Low prices induce more customers to enter the market, further accelerating demand growth. Likewise, aggressive capacity expansion accelerates growth of the customer base, further strengthening the word of mouth feedback. By augmenting the growth-producing feedbacks such a strategy increases peak demand and forces saturation to occur more rapidly. The resulting excess capacity and losses may overwhelm any cost advantage gained through the learning curve. The example illustrates a general point: subjects may exacerbate or moderate the forces which create boom and bust, depending on how well they understand the feedback processes in the environment. The misperceptions of feedback hypothesis thus predicts strong main effects of the treatments, with performance relative to potential degraded significantly by stronger word of mouth and longer product lifetimes.

Yet performance should improve with experience. Improvement might arise for two reasons with quite different implications. First, all features of the task other than the treatment variables remain constant over successive trials. Subjects can be expected to improve simply because they become increasingly familiar with the task and information display. Later trials will be informed by knowledge of the magnitudes and timing the variables achieved in prior trials. Beyond learning from these surface features, however, we hope and expect subjects will gain a deeper appreciation for the dynamics of the system and the feedback processes which produce them, resulting in changes in strategy which allow them to perform better in complex feedback environments. The distinction between these two modes of learning is critical. Since real situations vary in more dimensions than the task, improvement based on knowledge that, e.g. "last time demand reached about two million units/quarter" will not transfer well from one actual new product setting to another. Insight into the feedback structure and dynamics of such settings, however, can be applied to situations with very different numerical values. Learning derived from surface features of the task is expected to improve performance on average, but differences in performance across feedback conditions would remain. Insight into the feedback structure of the task, in contrast, should help subjects improve performance more in conditions of high feedback complexity. Such learning would manifest as a significant interaction between treatments and trial in which the negative effects of strong word of mouth and long product lifetimes are moderated by experience.

We created five scenarios identical in all respects except for the strength of the word of mouth feedback and average lifetime of the product (replacement fraction). These parameters are varied from half to double the base case values. Figure 3 shows the pattern market demand would take in the five scenarios if capacity were always equal to orders and assuming price equals unit costs plus a fixed gross margin of 25%. In all cases the characteristic pattern of growth, peak, and decline to equilibrium is present, but the growth rate, peak value and timing of orders, and equilibria vary. Of course, the actual order pattern in any trial is strongly influenced by the subject's decisions, both directly, through the influence of price and capacity decisions on customer purchases, and indirectly, through the subject's influence on the competitor's price and capacity decisions.

The 122 subjects were students in two sections of an elective class on system dynamics for corporate strategy at MIT's Sloan School of Management. About 35% and 40% were first and second year MBA students, respectively. Roughly 10% were mid-career managers in an executive MBA program, 7% were undergraduates and the rest were graduate students from other MIT departments. We used a Latin square design with five sequences of the five scenarios. Though a few subjects failed to complete all five trials, the design was quite well balanced. The number of trials in each of the 25 cells of the design ranged from 19 to 27, with a mean of 23.

The task was assigned as homework to be done individually within ten days. Subjects received a detailed written description of the task describing their firm, the market, the competitor, the cost structure, information available, and so on (Paich 1992). The software was demonstrated in class, and the simulated environment discussed. Subjects were randomly assigned to one of the five sequences by randomizing the floppy disks prior to distribution. Subjects could take as much time as they wished for each decision, and could suspend play between trials, resuming it later, moderating fatigue effects and encouraging reflection between trials. The instructions directed subjects to keep a log during each trial, including the strategy they intended to follow and evaluation of their results. Subjects were told these write-ups would be graded for the quality of the analysis, providing an incentive to formulate a sensible strategy and evaluate its effects carefully. Subjects also received "bonus points" in proportion to their cumulative profits for all five trials (grades were never reduced no matter how poorly subjects did). Students' evident concern with grades and the large number of questions received about the bonus point system suggest the bonus provided a powerful incentive to perform well.[4] These conditions allowed individual effort to vary, possibly introducing additional between-subject variance, but increased the incentives and opportunity for subjects to perform well and learn from their experience.

Each trial consisted of a 40 quarter market. The first two quarters of data were provided to orient the subjects, who then made 38 sets of target capacity and price decisions. After each decision, outcome feedback was provided showing the results for the quarter. A spreadsheet display was used since the subject pool, management students with excellent computer skills, were all experienced in interpreting such displays. The screen presented 19 variables, including a complete description of the firm's operations and finances, extensive market data and competitor intelligence. The display normally showed the current and three prior quarters. Subjects could at any time scroll through the entire history with a few clicks of the mouse. In addition, subjects could select up to four variables to display graphically the entire history of the trial to date. Any of the variables in the spreadsheet could be so displayed; graphs could be constructed at any time and as often as desired. The software automatically recorded the results.

Results

Before presenting the statistical results it is useful to examine the dynamics produced by the subjects. Figure 4 shows a typical first trial (overall this subject's profits were 114% of the grand mean). The subject's log (table 1) records his strategy before playing: "grow at market pace. Price follower". However, the subject faced the most difficult condition (strong word of mouth; low re-purchase fraction). Orders grow rapidly. The subject comments at time two "Need much more capacity" and raises target capacity. However, due to the acquisition lag, capacity constrains shipments, backlog grows, and delivery delay rises. The subject tends to follow competitor price moves with a lag, even though the subject is unable to fill incoming orders throughout the growth phase. The subject continues to increase target capacity to a peak of 5 million units in quarter 16 – reflecting rapid industry growth and the large backlog of unfilled orders. However, orders peak in quarter 12 at about 2.7 million units, and by quarter 14 capacity has risen enough to work off all excess backlog. Shipments fall precipitously to the rate of new orders. The subject dramatically cuts target capacity, commenting "I've gotta cut fixed costs", but actual capacity lags behind, peaking at 3.7 million units in quarter 18 just as orders fall to their low point. The subject is unable to cover the fixed costs of his excess capacity and experiences large losses. He writes "fire the CEO" as cumulative losses reach $475 million in period 21. By quarter 25 orders stabilize at the replacement equilibrium. The subject gradually reduces capacity and uses price to manage utilization.

Profitability is restored, and the cumulative loss is cut to 'only' \$268 million by the end of the trial.

Performance Measures

Since the profit potential of each scenario depends on the strength of word of mouth and the product lifetime, subjects' raw performance – cumulative profit – confounds their relative ability to manage the situation with absolute profit potential.[5] We therefore assess subject performance relative to benchmarks to remove the effect of the treatments on potential profits. Profit equals the product of unit sales and the profit margin (price - unit costs). Sales depend multiplicatively on the strength of word of mouth, w, and the replacement fraction, r. Thus cumulative profit is a multiple of function of w and r, and the appropriate performance measure, Π, is the ratio of cumulative profit for each subject i in each trial t, $\pi_{it}(w,r)$, to cumulative benchmark profit $\pi^*(w,r)$:[6]

$$\Pi_{it}(w,r) = \pi_{it}(w,r)/\pi^*(w,r). \tag{1}$$

Π adjusts raw profit for the intrinsic profit potential of the task and allows us to measure the effects of the experimental treatments beyond changes in intrinsic task difficulty. The benchmark reported here is provided by simple behavioral rules for both price and target capacity[7]:

$$C^*_t = s^* D_{t-1} (1+g_{t-1})^{\alpha_2} (B_t/C_t)^{\alpha_3} \tag{2}$$

$$g_{t-1} = (D_{t-1} - D_{t-2})/D_{t-2} \tag{3}$$

where C^* is target capacity, s^* is target market share, D is total industry sales, g is the fractional growth rate of industry sales over the most recently available period, B is the current backlog and C is current capacity. The target market share is set to 50%, with $\alpha_2 = 2.88$ and $\alpha_3 = .83$.[8] The capacity rule seeks to capture 50% of expected demand, where demand is forecast by extrapolating current industry sales at the current growth rate. In addition, target capacity is increased (decreased) relative to the demand forecast when capacity is insufficient (excessive) relative to desired production.

The benchmark strategy assumes cost-plus pricing with a constant gross margin:

$$P_t = (1+m)c_t \tag{4}$$

where c = unit costs and m = gross margin, set to .25. Price in the benchmark strategy simply follows costs down the learning curve, with a markup sufficient to cover marketing expense, investment costs, and provide a reasonable profit (at normal capacity utilization).

The behavioral benchmark is a simple, even naive, rule. It involves no game-theoretic reasoning. There is no explicit consideration of investment costs, competitor price or capacity, nor any market information, much less of the competitor's strategy. It utilizes only four cues (costs, industry sales, backlog, and current capacity) rather than full information. The rule naively extrapolates demand growth and does not anticipate market saturation. It does not use pricing to clear the market, control profit margins, or signal intentions. The behavioral benchmark should therefore be a floor on subjects' performance.

We next estimate general linear models to test the hypotheses above. We first tested for the effect of the sequence of scenarios. No sequence effect was found (p=.27 with Trial (T), Word of mouth (w), and Replacement fraction (r) as explanatory variables). Sequence was therefore not included in subsequent models. We next estimated a model with Subject (S), Trial, w, and r as explanatory variables along with the interactions w x r, T x w, T x r, and T x w x r (table 2).

Overall, subjects performed very poorly relative to the benchmark. Despite the naivete of the benchmark rule, the estimated mean ratio of profit to benchmark profit was under 24%. The benchmark beats the subjects 492 trials to 73: in 87% of the trials the naive benchmark was a ceiling, not a floor, on performance.

The main effects of the treatments are highly significant and consistent with the MOF hypothesis. Over and above changes in intrinsic profit potential, performance relative to the benchmark is severely degraded as feedback complexity increases (Table 2; figure 5). Specifically, the stronger the positive feedback loops which produce growth (the stronger the word of mouth effect), the worse people do even relative to the naive benchmark. Why? The faster the growth in orders the harder it is for subjects to match capacity with demand. Excess backlogs build up, increasing delivery delays and reducing market share, all reducing profit. The stronger the positive growth loop, the sooner and higher demand will peak and the larger the decline in demand when demand drops to replacement rates, leading to huge losses when the fixed costs of peak capacity cannot be covered by low demand. Also, unnecessary investment costs are incurred to the extent subjects respond to faster demand growth with faster growth of capacity.

Most importantly, subjects seem insensitive to, and their behavior exacerbates, the critical feedbacks which couple the firm to its market and competitor. The word of mouth feedback depends on the customer base, which can only increase as fast as shipments. The faster the subjects increase capacity the faster the customer base grows, and the stronger the word of mouth loop will be. And to the extent subjects seek to gain market share by pricing below the competition they speed demand growth as lower prices draw more people into the market. As will be seen below, subjects consistently act to strengthen these positive feedbacks, speeding and increasing the severity of market saturation and resulting in larger losses during the 'bust' phase.

Also consistent with the MOF hypothesis, the longer the useful life of the product the worse subjects perform relative to the naive benchmark (figure 5). The longer the useful life, the lower the replacement demand and the steeper and deeper the drop from peak orders as the market saturates. Subjects are not able to track demand even as well as the naive strategy which forecasts by univariate extrapolation and has no knowledge of the lifecycle. The highly significant interaction between w and r shows that the negative effects of feedback complexity on performance are compounded when the word of mouth feedback is strong *and* the replacement fraction small. Frequent repurchase implies demand overshoots its equilibrium value only slightly – the nonlinear transition from growth-generating positive feedback to decline caused by the negative feedback of market saturation is gradual and mild. Subjects who are overly aggressive in acquiring capacity during the boom are not punished too harshly. Thus, as shown in figure 5, average performance is significantly better and changes in word of mouth have but little effect when the replacement fraction is high. When there is little replacement demand, however, the nonlinear transition from boom to bust is sudden and sharp. Thus average performance is significantly worse and changes in word of mouth have stronger effects when the replacement fraction is low. In the most difficult case (w=2, r=.5), mean subject profits are *negative* $121 million while the benchmark is *positive* $229 million; in the condition with the weakest feedbacks (w=.5, r=2), mean performance is $605 million while the benchmark is $922 million.

Learning

As expected, subjects improve with experience. Mean performance is just 13% of the benchmark in trial 1 but rises to 63% in trial 5 (figure 6). The win ratio for the subjects rises from under 4% in trial 1 to 17% in trial five. However, improvement is slowing by the final trial. Mean subject profits in trial 5 are the same as in trial 4, and the subject win ratio actually drops between trials 4 and 5. Performance seems to be saturating well below even the naive benchmark.

We had hoped that, beyond the general improvement shown by the main effect of trial on performance, subjects would be improving their understanding of the environment with experience and so would develop heuristics which produce better relative performance in the conditions with high feedback complexity. Such learning would be reflected in significant interactions between the treatments and trial. However, none of the interactions between treatments and trial are even remotely significant. Though subjects improve on average there is no evidence to suggest they are learning to cope better with environments involving strong feedback processes.

The failure of subjects to improve their ability to manage complex feedback environments is an important, and somewhat unexpected, result. The conditions for learning are excellent: Subjects receive immediate, comprehensive and accurate outcome feedback. In the course of their five trials they made nearly 200 sets of decisions, were under no time pressure, and faced strong incentives for performance. While prior research shows that learning from outcome feedback is difficult in the presence of noise and nonlinearity (e.g. Brehmer 1980), the present task is completely deterministic and subjects had extensive knowledge of the causal structure of the system. Further, despite the general learning effect, performance after five trials remains significantly worse than the naive benchmark.

Modeling subjects' decision rules

To understand both the weakness of the overall learning effect and the failure of subjects to improve performance across feedback conditions we next test several behavioral decision rules for target capacity and price. The rules postulated here were suggested by consideration of the subjects' written reports of their strategies, prior models of similar decisions in the literature, and the feedback structure of the task. Other rules are of course possible. The models are designed to indicate the importance of different cues in subjects' decisions; changes in the weights across trials yield insight into learning and its limitations (Einhorn, Kleinmuntz, and Kleinmuntz 1979).

Generalizing the behavioral benchmark, we postulate decision makers who select the target share of the market their firm seeks to capture, estimate future market demand from prior information, current demand, and the recent growth rate of demand, and invest to balance capacity with demand. The rule thus combines feedforward or forecasting (estimation of future demand) with a feedback component to correct errors in the forecasts (the response to excess or insufficient capacity). Specifically,

$$C^*_t = s^*[D^e_0{}^{(1-\alpha_1)} D_{t-1}{}^{\alpha_1}](1+g_{t-1})^{\alpha_2}(B_t/C_t)^{\alpha_3} \tag{5}$$

$$g_{t-1} = (D_{t-1} - D_{t-2})/D_{t-2} \tag{6}$$

where s^* is target market share (assumed constant), D^e_0 is the prior expectation of average industry demand, D is actual demand, g_t is the expected fractional growth rate of demand, B is the backlog (desired production), and C is actual capacity.[9] The rule assumes subjects seek to capture a certain share of the forecasted market demand. Forecasted demand is modeled as a weighted geometric average of current demand and a prior expectation. That prior belief is likely to be strongly conditioned by the demand observed in earlier trials. Subjects with little initial idea of the size of the market, as before the first trial, would most likely follow a demand tracking strategy with a high α_1. Subjects whose prior expectation is never modified by actual experience would have $\alpha_1 = 0$, while $0 \le \alpha_1 \le 1$ indicates a conservative strategy in which target capacity falls increasingly short if orders exceed the value of demand the subject expects before the trial begins. Such behavior could be intentional, if subjects fear overshooting the equilibrium as demand becomes large, or the result of inadvertent anchoring on prior beliefs. In addition to estimating current demand, the capacity acquisition lag requires subjects to account for likely growth in demand. Subjects who extrapolate recent changes in demand would have $\alpha_2 > 0$; extrapolative expectations based on the most recent demand data are assumed (this cue is reported on the information display and is commonly available in actual markets). Target capacity should also respond to the demand/supply balance as measured by the ratio of backlog (desired production) to production capacity.

The proposed decision rule for price P assumes subjects use markup pricing:

$$P_t = UVC_t \cdot M^*_t \tag{7}$$

where UVC = unit variable cost and M* = gross margin. Gross margin depends on the subject's response to demand/supply balance and the policy for passing cost reductions on to the consumer:

$$M^*_t = M_0 \cdot (UVC_t/UVC_0)^{\beta_1} \cdot (B_t/C_t)^{\beta_2} \tag{8}$$

As the firm moves down the learning curve, the subject must decide how much of the cost reduction to pass on to consumers. All cost reductions are passed into price when $\beta_1 = 0$, while $-1 \le \beta_1 \le 0$ indicates price falls less than costs. Positive values of β_1 indicate price falls faster than costs, perhaps indicating an attempt to build market share and move more rapidly down the learning curve than the competitor.[10] We further expect that the gross margin will increase when desired production (backlog) is high relative to capacity ($\beta_2 > 0$).

Collecting constant terms, assuming independent multiplicative errors, and taking logs yields the form in which the decision rules were estimated:

$$\log(C^*_t) = a_0 + a_1\log(D_{t-1}) + a_2\log(1+g_{t-1}) + a_3\log(B_t/C_t) + \varepsilon_1 \tag{9}$$

$$\log(P_t) = b_0 + b_1\log(UVC_t) + b_2\log(B_t/C_t) + \varepsilon_2 \tag{10}$$

Each rule was estimated separately for each of the five trials of each subject. OLS regression revealed positive autocorrelation in the residuals, so the Cochrane-Orcutt procedure was used.

The proposed rules capture the bulk of the variance in subject's decisions (table 3). For target capacity the mean $\overline{R}^2$ is .87, exceeds .90 for more than two-thirds of the trials, and is less than .50 for only 5%. The mean $\overline{R}^2$ is .95 for price, with $\overline{R}^2 > .90$ for more than 87% of the trials and less than .50 for just 1%. The coefficients generally have the expected signs. For target capacity, the constant a_0 and the parameter a_1 usually are large and statistically significant. The mean estimate of $\alpha_1 = .38$ indicates subjects based their capacity decisions primarily on their prior estimate of market demand and only secondarily on actual market demand. To assess the prior expectation of market demand, we note from equations 5 and 9 that $\alpha_0 = (1-\alpha_1)\ln(s^*D^e)$. After obvious outliers are eliminated[11], the mean value of $\ln(s^*D^e) = 13.32$, indicating subjects' initial goal for capacity averaged roughly 600,000 units. In contrast, the parameters a_2 and a_3 are generally small – roughly half are not statistically different from zero – showing subjects are quite insensitive to the recent growth of demand and the demand/supply balance. For the price equation, the constant b_0 and the parameter b_1 are usually statistically significant while b_2 is very small and not significant in nearly two-thirds of the cases. Subjects generally price to follow costs (or competitor price) down the learning curve. The supply/demand balance has but little effect on price.

The estimated decision rules provide insight into subjects' poor overall performance relative to the benchmark rules. Subjects generally set target capacity equal to their initial goal for capacity, are only partially responsive to actual market demand, and are quite insensitive to the growth in demand. Given the capacity acquisition lag such conservative demand forecasts ensure that actual capacity will be grossly inadequate during the boom phase, causing high backlogs, long delivery delays, and market share erosion. The subjects' insensitivity to demand growth further exacerbates capacity shortfall during the boom and is consistent with prior work (Wagenaar and Timmers 1979). However, the low weight on the supply/demand balance in both the capacity and pricing decisions is quite surprising. Hogarth (1981) argues that dynamic decision making might be better than one-shot static decision making, since subjects can review and revise prior decisions as outcome feedback becomes available, gradually correcting errors. Here, however, this postulated adaptation does not operate well. Subjects fail to increase target capacity sufficiently as backlogs accumulate, and they fail to cut target capacity aggressively when faced with excess capacity. The result is lost profit during the growth phase and larger losses during the bust phase. Similarly the near-zero weight on the backlog/capacity balance in the pricing rule shows that subjects price too low during the boom phase, even though they cannot possibly satisfy the current demand; likewise, subjects generally fail to cut prices to stimulate demand during the bust despite huge amounts of excess capacity. Not only are subjects insufficiently adaptive, but their capacity and pricing decisions are inconsistent.

We next investigate how subjects' decision weights change over trials. In particular, we seek to explain the weak overall improvement in performance and the failure of subjects to improve their relative performance across different feedback complexity conditions in terms of changes in their cue weights. To do so we explore models of each estimated parameter with subject, trial, and the word of mouth and replacement fraction treatments as explanatory variables (table 4).

All the coefficients of the target capacity rule change significantly over the five trials, indicating experience caused subjects to alter their forecasts of demand and their responsiveness to market growth and demand/supply balance. The estimated elasticity of target capacity with respect to current market demand (a_1) falls from .49 in trial one to .32 by trial five. Before their first trial subjects have little prior knowledge of likely demand, so have little choice but to follow actual demand (even so there is substantial conservatism in their forecasts). As suggested by the poor results and the subject logs (e.g. table 1), subjects find it difficult to forecast the boom and bust pattern. With experience, however, subjects learn approximately how big the peak and equilibrium values of demand will be. They rely increasingly on their knowledge of the demand levels reached in prior trials and become less responsive to the actual demand in the current trial. Similarly, initial demand and capacity levels are low compared to the peak and equilibrium values of demand, so initial expectations of demand in trial one, before any experience is gained, should be low but rise over trials. Indeed, the imputed prior expectation $s^{*}D^{e}$ averages about 300,000 units in trial one; it doubles by the second trial, and rises to nearly 900,000 by trial five. The estimate also depends significantly on equilibrium demand (indicated by the significant effect of replacement fraction on $s^{*}D^{e}$) since the true equilibrium demand becomes clear about half-way through (see figures 3-4).

The capacity acquisition lag requires subjects to forecast demand well into the future. But the regression results show that on average subjects are virtually unresponsive to the growth rate of demand. While subjects may choose not to respond fully to demand growth, ignoring it inevitably results in insufficient capacity during the boom and larger surpluses during the bust. Table 4 shows that subjects do learn to respond to demand growth with successive trials. However, the effect is small. In the first trial the mean estimate of the elasticity of target capacity with respect to market growth is negative, indicating subjects expect demand to regress to prior values rather than growing further. By the fifth trial most have learned to extrapolate recent demand changes, but the mean elasticity has risen only to about .10 (similar shifts from regressive to extrapolative expectations were found by Andreassen (1990a, 1990b) in stock market trading experiments). While experience does lead subjects to anticipate market growth to some degree, the improvement is so small that even after five trials subjects show little ability to forecast changes in demand or account for the lag in acquiring capacity. The result is massive backlogs and lost revenue during the growth phase, and slow reductions in excess capacity during the bust phase.

The response of target capacity to the demand/supply balance also increases with experience. The mean elasticity rises from .18 in trial one to about .5 in the last two trials. However, even with experience the average response to supply and demand is much less than optimal.

Worse, there is no evidence that subjects' pricing strategies evolve. In particular, the responsiveness of price to the demand/supply balance does not increase with experience. Subjects learn slowly to adjust capacity to demand, but do not learn to use price to clear the market, nor do they alter the average level of price. During the boom, price can be raised to both reap higher profit and to slow the growth of demand; during the bust, price can be cut to boost market share and increase capacity utilization, reducing losses. Subjects in general do neither, foregoing a critical opportunity to moderate the severity of the boom and bust dynamic and boost profitability.

Thus experience does not move subjects towards a greater appreciation of the feedback processes which create market dynamics. Rather, subjects seem to find the dynamics so hard to anticipate or understand that they move towards a 'ballistic' strategy in which prior beliefs about equilibrium demand are increasingly influential while outcome feedback about actual demand is increasingly ignored. The subject logs confirm the regression results. The following, describing a subject's strategy for the final trial, is typical:

> Since I *know* I can't beat my competitor during peak sales, I am going to ramp up early to the level that I believe might represent replacement sales. I can hopefully then capture that much of the market during the peak..., and the majority of the replacement market after sales peak. Price will be adjusted as necessary to maintain market share.

The subject immediately boosted capacity to about 2 million and held it constant through quarter 30, even though orders reached more than 5 million. Price was increased above the competitor during the boom, but not enough to prevent the long delivery delays. During the bust, prices were cut, but not enough to prevent significant excess capacity.

Discussion

Prior work in dynamic decision making suggested that subjects' mental models of dynamic environments are generally poor. Specifically, prior work suggested subjects do not account well for feedback loops, time delays, accumulation processes, and nonlinearities. However, much prior work did not explicitly vary the strength of feedback processes as treatments to establish how dynamic complexity influenced performance, nor did they include market institutions and opportunities for learning which might mitigate the errors. Many prior experiments used abstract tasks or tasks not relevant to the subjects' training and experience. The present experiment presents subjects with a common and realistic management task. Extensive opportunities for learning were provided – fifty years of simulated industry experience with perfect, immediate outcome feedback.

Despite these opportunities for excellent performance, the vast majority of subjects are outperformed by a naive behavioral rule. The benchmark rule utilizes a small subset of available information and combines these cues in simple ways, without recourse to optimization or game theoretic reasoning. As in many prior static judgment and decision making tasks, what should have been a floor on performance turned out to be a very high ceiling.

Consistent with the misperceptions of feedback hypothesis, the stronger the feedback processes in the environment the worse people do *relative to potential.* In particular, subjects fail to account for a fundamental structural feature of durables markets: *ceteris paribus*, the faster sales grow, the sooner and more suddenly the market must saturate. The effect is not mere forecasting error. Subjects' own actions exacerbate the problem by strengthening the positive feedback loops, generating a more vigorous boom and a more severe bust. Further, subject behavior is riddled with inconsistencies. Most subjects maintain price close to or less than the competitor price, even during the growth phase when capacity lags orders and the availability of their product plunges.

They cut prices to match the competitor even though they are unable to meet the current demand. Low prices also intensify the boom and bust dynamic by stimulating demand, reinforcing the word of mouth loop and leading to faster growth and steeper decline.

Most disturbing, there is no evidence that subjects improved their ability to manage the environments with high feedback complexity as they gained experience, despite improvement on average. Analysis of estimated cue weights revealed the sources of subjects' failure to improve their ability to manage complex dynamic environments. Based on their experience of previous games, the subjects altered their strategies to use less, not more, outcome feedback on demand. The subjects learned the basic pattern of market dynamics. They predicted the approximate size of the replacement market and quickly increased capacity to that level, largely ignoring the level or growth rate of market demand in the specific scenario. Because experience provided useful information on likely equilibrium demand, performance on average improved with trials. Because subjects were insufficiently responsive to the level and growth rate of the market and to the demand/supply balance, they did not improve their ability to handle the difficult feedback environments. The nature of subjects' learning is particularly remarkable in light of their poor overall performance: after five trials mean performance is just 60% of the naive benchmark, and only 17% of subjects beat the benchmark in their final trial.

Conditions for learning in the experiment are superior those in the real world, where outcome feedback is often missing, noisy, and significantly delayed; where other confounding factors are more numerous and demands on managerial attention are greater; and where the time scale for change can exceed the tenure of managers (Kahneman and Tversky 1987, Thaler 1987). Further, the learning that did occur in the experiment would be much less successful in the real world. By design, the experiment simplified the environment by holding the potential market constant across scenarios. In real situations, the variation is much greater. Few decision makers could ignore market feedback, determining capacity primarily from the history of prior products and then failing to revise that decision despite the huge backlogs, angry customers and other pressures engendered by rapid growth. Without mental models which account for the feedback processes, time delays, and nonlinearities which create boom and bust, outcome feedback is likely to lead to behavior such as observed in subjects' first trials. The striking correspondence between many first trials and the behavior of numerous actual firms suggests these misperceptions of feedback may play a significant role in the real world. Empirical studies contrasting the dynamics of firms facing high and low feedback complexity should be undertaken to test the generality of these findings. Existing studies (e.g. Zarnowitz 1985, Mosekilde, Larsen, Sterman and Thomsen 1992) do show that markets characterized by long lags, strong positive feedbacks, accumulations, and nonlinearities (e.g. commercial real estate, shipping, capital goods) do suffer from more instability than those with less feedback complexity (e.g. soft goods, services). Experiments underway with seasoned managers as subjects and field study could also provide important insight into processes of individual and organizational learning which might enable high performing firms in complex environments to avoid the problems observed in the laboratory.

The results also have pedagogical and prescriptive implications. The recognition that traditional management pedagogy, stressing lectures and cases, often does not lead to improved decision making ability has long motivated the use of management games. However, evidence on the effectiveness of traditional business games is mixed at best (see Graham et al. 1992 for a review). The failure of subjects to learn from experience in the present study offers an explanation and suggests principles for the design of more effective management simulations and computer-supported learning environments. Specifically, learning processes based on simulations must include the use of conceptual schema and tools capable of representing and interpreting feedback complexity. Action and reflection must be integrated in an iterative, interactive learning cycle (see Morecroft and Sterman 1992 for illustrations and evidence). Experiments with these tools are also underway with a number of organizations, where speeding individual and organizational learning is a growing challenge for firms facing shorter product lifecycles and other changes which intensify the feedback complexity of their environment.

Notes

1 See Paich (1992) for full documentation of the model, methods and results. The game, revised for educational use with full documentation and instructions, is available from John Sterman.

2 In reality additional feedbacks exist involving e.g. changes in technology, line extensions, cannibalization of sales by new generations of the product, network externalities, and so on. To keep the task manageable these effects are not treated.

3 Experimental evidence shows that inventories would substantially destabilize the system and make the player's task much harder (Sterman 1989b, Diehl 1992). Beinhocker (1991) discusses the destabilizing role of inventories in the collapse of toymaker Worlds of Wonder.

4 The role of incentives, monetary and nonmonetary, in decision making performance is complex (see Hogarth et al. 1991 for a review and experimental evidence). There is no evidence to suggest subjects did not take the task seriously or attempt to do their best.

5 Theoretically, the net present value of profits should be the performance measure. However, experimental studies (see Prelec and Loewenstein 1991 and references therein) show people's subjective time preferences often do not follow standard discounted utility models. We therefore used undiscounted cumulative profits to simplify the cognitive burden of the task and to avoid confounding the results with various anomalies in intertemporal choice. In fact, the bulk of the profits earned by subjects occur in the final phase of the simulated markets, after equilibrium has been reached, while the losses occur early in the product life cycle.

6 Consider equilibrium, when sales = orders ≈ capacity and $\pi(w,r) \approx$ orders*(price - unit costs). In equilibrium orders are determined by the product of total market size and the replacement purchase fraction. Thus both subject profit and benchmark profit are multiples of the replacement fraction, and the ratio then allows comparison of profit relative to potential across scenarios with different repurchase fractions.

7 We have also analyzed other benchmarks, including a 'perfect foresight' rule in which target

capacity is always set to provide capacity exactly equal to desired production (but maintaining the naive pricing strategy). This rule substantially outperforms the behavioral rule and dominates the subjects in 98% of the trials. Cumulative profit relative to cumulative sales was also used to remove the effect of the treatments on total market size. The statistical significance and magnitudes of the treatment and learning effects are robust to these alternative benchmarks (Paich 1992).

8 The values of α_1 and α_2 were chosen by grid search of the parameter space to maximize cumulative profit over the five scenarios conditional on the assumption that target market share is 50%. We stress, however, that the simplistic behavioral benchmark rule is far from optimal.

9 Demand is measured by industry sales (shipments), as reliable industry aggregate order information is generally not available in actual markets. Note, however, that sales may be (and in many trials are) constrained by capacity during the boom, so that sales data measure not demand but capacity, a subtle but important way in which nonlinearity influences task complexity.

10 Gross margin should also depend on competitor price. However, competitor cost (and therefore price, since competitor margin is constant) is highly correlated with the firm's own costs since the learning curve for each has the same strength. Therefore subjects' responses to costs and competition cannot be independently estimated. Substituting competitor price for costs in the price equation does not substantially change the results reported here.

11 Values such that $0 \leq \ln(s^*D^e) \leq \ln(20e6)$ were retained (410 out of 565). This range easily encompassed all actual market demand levels attained in the task.

References

Andreassen, P. (1990a) Judgmental extrapolation and the salience of change. *Journal of Forecasting*. 9, 347-372.

Andreassen, P. (1990b) Judgmental extrapolation and market overreaction: on the use and disuse of news. *Journal of Behavioral Decision Making*. 3, 153-174.

Bass, F. M. (1969) A New Product Growth Model for Consumer Durables. *Management Science*. 15, 215-227.

Beinhocker, E. (1991) Worlds of Wonder (A) and (B). Case Study available from John Sterman, Sloan School of Management, MIT, Cambridge MA 02139.

Brehmer, B. (1980) In One Word: Not From Experience. *Acta Psychologica*, 45, 233-41.

Brehmer, B. (1990) Strategies in Real Time, Dynamic Decision Making, in Hogarth, R. (ed) Insights in Decision Making. Chicago: University of Chicago Press, 262-279.

Broadbent, D., & Aston, B. (1978) Human Control of a Simulated Economic System. *Ergonomics*, 21(12), 1035-43.

Camerer, C. (1987) Do biases in Probability Judgment Matter in Markets? Experimental Evidence, *American Economic Review*. 77, 5 (December), 981-997.

Diehl, E. (1992) Effects of Feedback Structure on Dynamic Decision Making. PhD. Dissertation, MIT Sloan School of Management.

Einhorn, H. J., Kleinmuntz, D., & Kleinmuntz, B. (1979) Linear Regression and Process-Tracing Models of Judgment. *Psychological Review*, 56(5), 465-85.

Funke, J. (1991) Solving Complex Problems: Exploration and Control of Complex Systems, in R. Sternberg and P. Frensch (eds.), *Complex Problem Solving: Principles and Mechanisms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gort, M. and S. Klepper (1982) Time paths in the diffusion of product innovations. *Economic Journal* 92, 630-653.

Graham, A. K., Morecroft, J. D., Senge, P. M., & Sterman, J. D. (1992) Model Supported Case Studies for Management Education. *European Journal of Operational Research*, forthcoming.

Hauthakker, H. S., & Taylor, L. C. (1970) *Consumer demand in the United States*. Cambridge, MA: Harvard University Press.

Hogarth, R. (1981) Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics, *Psychological Bulletin*. 90, 197-217.

Hogarth, R. M., & Reder, M. W. (1987) *Rational Choice: The Contrast between Economics and Psychology*. Chicago: Univ. of Chicago Press.

Hogarth, R. M., & Makridakis, S. (1981) The Value of Decisionmaking in a Complex Environment: An Experimental Approach. *Management Science*, 27(1), 93-107.

Hogarth, R., B. Gibbs, C. McKenzie, and M. Marquis (1991) Learning from Feedback: Exactingness and Incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 17(4), 734-752.

Homer, J. (1987) A Diffusion Model With Application to Evolving Medical Technologies. *Technological Forecasting and Social Change*, 31, 197-218.

Kalish, S. and Lilien, G. (1986) Market Entry Timing Entry for New Technologies, *Management Science* 32(2) 194-204.

Kampmann, C. and J. Sterman (1992) Do Markets Mitigate Misperceptions of Feedback in Dynamic Tasks? Proceedings of the 1992 International System Dynamics Conference, Utrecht, The Netherlands.

Kleinmuntz, D., and J. Thomas (1987) The value of action and inference in dynamic decision making, *Organizational Behavior and Human Decision Processes,* 39(3), 341-364.

Klepper, S. and E. Graddy (1990) The evolution of new industries and the determinants of market structure. *RAND Journal of Economics.* 21(1), 27-44.

Kluwe, R. H., C. Misiak, and H. Haider 1989, Modelling the Process of Complex System Control, in Milling, P. and E. Zahn (eds), *Computer Based Management of Complex Systems.* Berlin: Springer Verlag, 335-342.

Mahajan, V., E. Muller, and F. Bass (1990) New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing* 54(1), 1-26.

Mahajan, V. and Y. Wind, eds. (1986) *Innovation Diffusion Models of New Product Acceptance.* Cambridge, MA: Ballinger.

Morecroft, J. (1985), "Rationality in the analysis of behavioral simulation models", *Management Science* 31(7), 900-916.

Morecroft, J. and J. Sterman (eds.) (1992) Modeling for Learning. Special issue of *European Journal of Operational Research.* Spring.

Mosekilde, E., E. Larsen, J. Sterman, & J. Thomsen (1992) Nonlinear Mode-Interaction in the Macroeconomy. In Feichtinger, G. (Ed.), *Annals of Operations Research* 37.

Paich, M. (1992) PhD Thesis, MIT Sloan School of Management.

Petre, P. (1985) Jack Tramiel is back on the warpath. *Fortune.* 4 March, 46-50.

Porter, M. (1980) *Competitive Strategy.* New York: Free Press.

Porter, M. (1983) *Cases in Competitive Strategy.* New York: Free Press.

Prelec, D. and G. Loewenstein (1991) Decision Making over Time and under Uncertainty: A Common Approach. *Management Science.* 37(7), 770-786.

Salter, M. (1969) Tensor Corporation. Case 370-041. Harvard Business School Publishing Division, Boston MA 02163.

Senge, P. (1980) A System Dynamics Approach to Investment Function Formulation and Testing, *Socioeconomic Planning Sciences*, 14, 269-80.

Senge, P., & Sterman, J. D. (1992) Systems Thinking and Organizational Learning: Acting Locally and Thinking Globally in the Organization of the Future. In Kochan, T. & Useem, M. (Eds.), *Transforming Organizations* Oxford: Oxford University Press.

Simon., H. A. (1982) Models of Bounded Rationality, The MIT Press, Cambridge.

Smith, V., Suchanek, G., and A. Williams (1988) Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets, *Econometrica*, 56(5), 1119-1152.

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. (1988, *People Express* Management Flight Simulator. Simulation Game (software), Briefing Book, and Simulator Guide, Available from author, MIT Sloan School of Management, Cambridge, MA 02139.

Sterman, J. D. (1987a) Testing Behavioral Simulation Models by Direct Experiment. *Management Science*, 33(12), 1572-1592.

Sterman, J. D. (1987b) Expectation Formation in Behavioral Simulation Models. *Behavioral Science,* 32, 190-211.

Thaler, R. (1987) The psychology of choice and the assumptions of economics. In Roth, A.E., ed. *Laboratory experiments in economics: Six points of view.* Cambridge: Cambridge University Press.

Tversky, A. and Kahneman, D.H. (1987) Rational choice and the framing of decisions. In Hogarth, R.M. and Reder, M.W., ed. *Rational choice: The contrast between economics and psychology*. Chicago: Univ. of Chicago Press.

Wagenaar, W., and H. Timmers (1979) The pond-and-duckweed problem: Three experiments in the misperception of exponential growth. *Acta Psychologica.* 43, 239-251.

Wagenaar, W. and S. Sagaria (1975) Misperception of Exponential Growth. *Perceptions and Psychophysics*. 18, 416-422.

Zarnowitz, V. (1985) Recent Work on Business Cycles in Historical Perspective: A Review of Theories and Evidence, *Journal of Economic Literature*. 23, 523-80.

Table 1. Subject log: Trial 1, Scenario D: Strong word of mouth, low repurchase rate.

Q	Comments Before Entering Decisions	Comments After Entering Decisions
1	Strategy: grow @ market pace; price follower; assume capacity growth is linear.	Sales WAY up !!!!!
2	Grow faster than the market. Need much more capacity. Expand to 55,000. P[rice] constant.	Sales WAY up !!!!! Try and be patient.
3	Hold capacity target fixed. P to 81 to follow competitor.	Sales WAY up ! Way above capacity. Behind. EXPAND ! ! !
4	Want 50% of market. Build to 150,000. Believe market max. = 300,000.	Sales WAY up. ! Way above capacity. Behind.
5	Want more. Build to 250,000 NOW. P up to 82.	
6		
7	Revise market max est. Build to 500,000 and vow never to add.	
8		
9	Build to 1,000,000 & vow never to add again.	Sales flattening therefore drop price.
10	Build to 1,500,000 & vow never to add again. P to 70.	Market demand up to 7,000,000.
11	Build to 5,000,000.	Sales fell. HELP !!!!!!
12	Destroy CAPACITY to 2,500,000. P to 65 to up demand.	LOOK AT NET INC[OME]. I got suckered !!!!
13		
14	P to 55	
15		
16	Target capacity = 90,000 ONLY. I've gotta cut fixed costs.	
17	P to 52	
18		
19		
20		
21	Capacity "OK". P up to 56.	A new high in lows. Fire the CEO.
22		
23		
24		
25	Capacity = orders = backlog GOOD.	Let's fool around with Price.
26	P to 58.	
27	P to 60.	
28	P to 62	
29	P to 62	
30	P to 61	Zero-sum game.
.		
.		
.		
40		That's all folks. This was humiliating.

Table 2

Analysis of cumulative profit relative to the behavioral benchmark.

Dependent Variable: $\Pi_{it}(w,r)$ = cumulative subject profit relative to benchmark profit as functions of subject i, trial t, (log) strength of word of mouth feedback w and (log) replacement fraction r.

N = 565, $\overline{R}^2 = 0.39$

Explanatory Variable	SS	df	Mean Square	F	p
Subject	182.13	121	1.51	1.26	0.05
Trial	30.29	4	7.57	6.31	0.000+
w	14.07	1	14.07	11.73	0.001
r	79.51	1	79.51	66.29	0.000+
w x r	9.93	1	9.93	8.28	0.004
Trial x w	3.87	4	0.97	0.81	0.52
Trial x r	2.38	4	0.60	0.50	0.74
Trial x w x r	4.73	4	1.18	0.99	0.42
Error	508.52	424	1.20		

Estimated Treatment and Trial Effects.

The interactions of trial with treatments are not shown as they are not significant.

Constant	0.235				
Trial:	1	2	3	4	5
	-0.353	-0.118	-0.053	0.216	0.308
w	-0.257				
r	0.613				
w x r	0.313				

Table 3

Means and standard deviations of estimated parameters for subjects' capacity and pricing rules (eq. 9-10). The parameter ρ is the estimate of the first-order autoregressive term. The final column shows the percentage of estimates which were not significantly different from zero.

Parameter	μ	σ	**% NS**
Capacity Rule:			
a_0	8.414	6.450	22
a_1	.383	.433	30
a_2	.036	.533	56
a_3	.318	.715	43
ρ	.560	.328	21
$\overline{R}^2$	.872	.183	
Pricing Rule:			
b_0	3.125	2.553	3
b_1	.259	.337	28
b_2	.016	.067	62
ρ	.781	.215	3
$\overline{R}^2$	.947	.095	

Table 4.

Dependence of estimated parameters on trial and treatments. The model in all cases is:

$$p_{it} = \text{Constant} + \text{Subject}_i + \text{Trial}_t + w_{it} + r_{it} + w_{it} \cdot r_{it} + \text{error},$$

where p_{it} is the estimated parameter for trial t of the ith subject; w and r are the (log) strengths of the word of mouth parameter and replacement fraction, respectively. Significance levels (p-values of the F-statistic) for the effects are given in parentheses. The subject factor was significant for all estimated parameters at better than the .001 level.

Parameter	**w**	**r**	**w*r**	**Constant + Trial Effect**					$\bar{R}^2$
				1	**2**	**3**	**4**	**5**	
Capacity Rule:									
$\ln(s^*D^e)$§	.077 (.55)	.453 (.000+)	.35 (.06)	12.6	13.3	13.6 (.000+)	13.5	13.7	.36
a_1	.006 (.82)	-.032 (.22)	-.06 (.14)	0.49	0.39	0.40 (.011)	0.33	0.32	.40
a_2	.171 (.000+)	.038 (.26)	.08 (.115)	-0.06	-0.02	0.05 (.018)	0.14	0.09	.36
a_3	.003 (.94)	-.031 (.47)	.05 (.43)	0.18	0.40	0.34 (.000+)	0.57	0.43	.40
Pricing Rule:									
b_0	-.126 (.42)	.288 (.07)	.08 (.72)	3.39	3.01	3.00 (.502)	3.27	2.92	.39
b_1	-.004 (.86)	-.028 (.18)	.01 (.76)	0.26	0.27	0.24 (.489)	0.30	0.23	.38
b_2	.018 (.000+)	-.007 (.12)	.01 (.063)	0.01	0.01	0.02 (.381)	0.02	0.02	.34

§ $\ln(s^*D^e)$ is the (log) of the imputed expectation of the share of equilibrium demand the subject seeks. It represents the subject's prior belief about the size of the market they wish to capture in the trial, and is calculated from the estimated coefficients as $\ln(s^*D^e) = a_0/(1-a_1)$; see eqs 5 and 9.

Figure 1. Boom and Bust: Sales and Net Income of Atari, Inc. Sources: 1976-1983: Warner Communications Annual Reports. 1984-1985: Atari, Inc. Company Reports; Investext. In 1984 Warner sold Atari; 1984 results cover the period from sale (5/17/84) through end of year. Operating income does not show 1984 charges against Warner profit associated with the sale of $592 million.

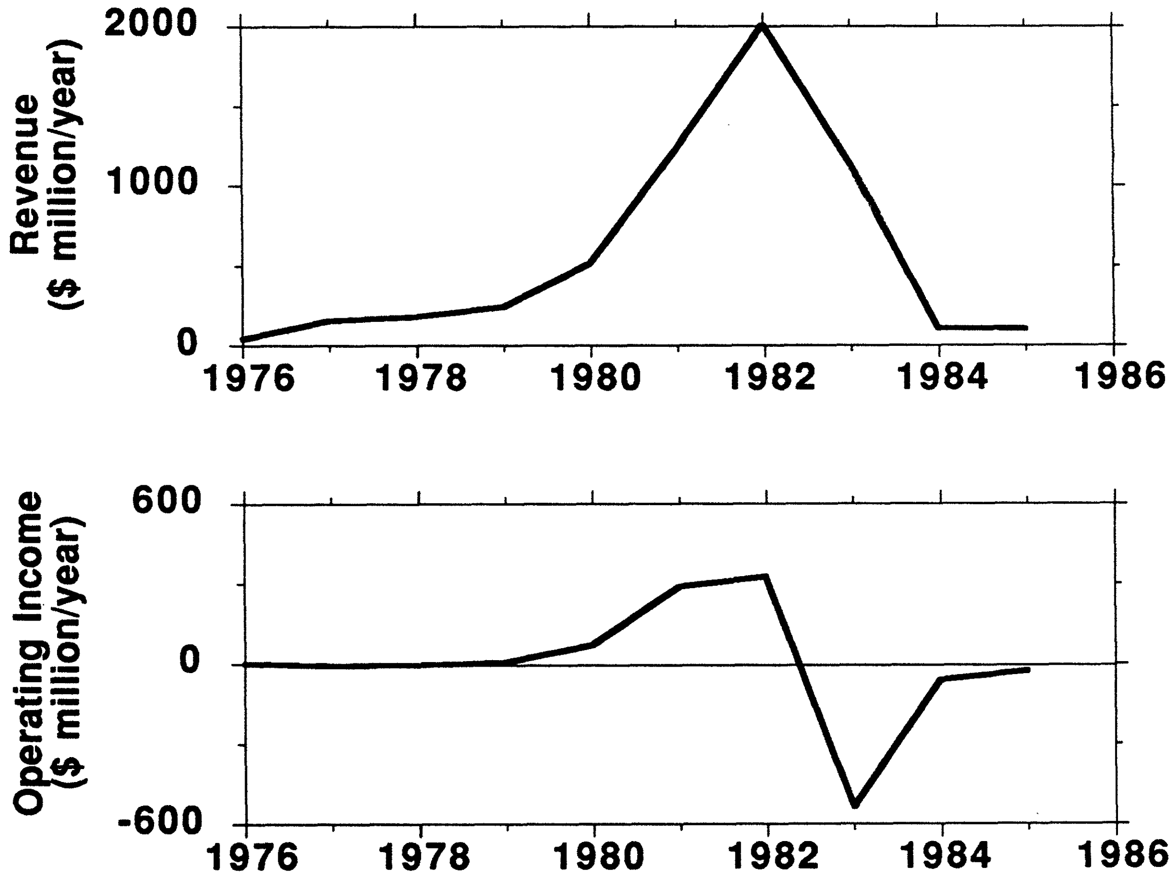

Figure 2 Causal Structure of the Market Sector.

Potential customers may adopt either through the effects of marketing or word of mouth, flowing into the customer base. When the decision to discard and repurchase is made, customers re-enter the potential pool. Changes in product price change the size of the potential market. The self-reinforcing (positive) word of mouth feedback loop promotes growth early in the product's life. As the pool of potential customers is depleted, the negative feedback of market saturation constrains adoption. In equilibrium adoption equals replacement demand. Not shown is the determination of market share between the two firms represented in the simulation. Though not shown, many additional loops are created by the coupling of the market to the subjects' decisions.

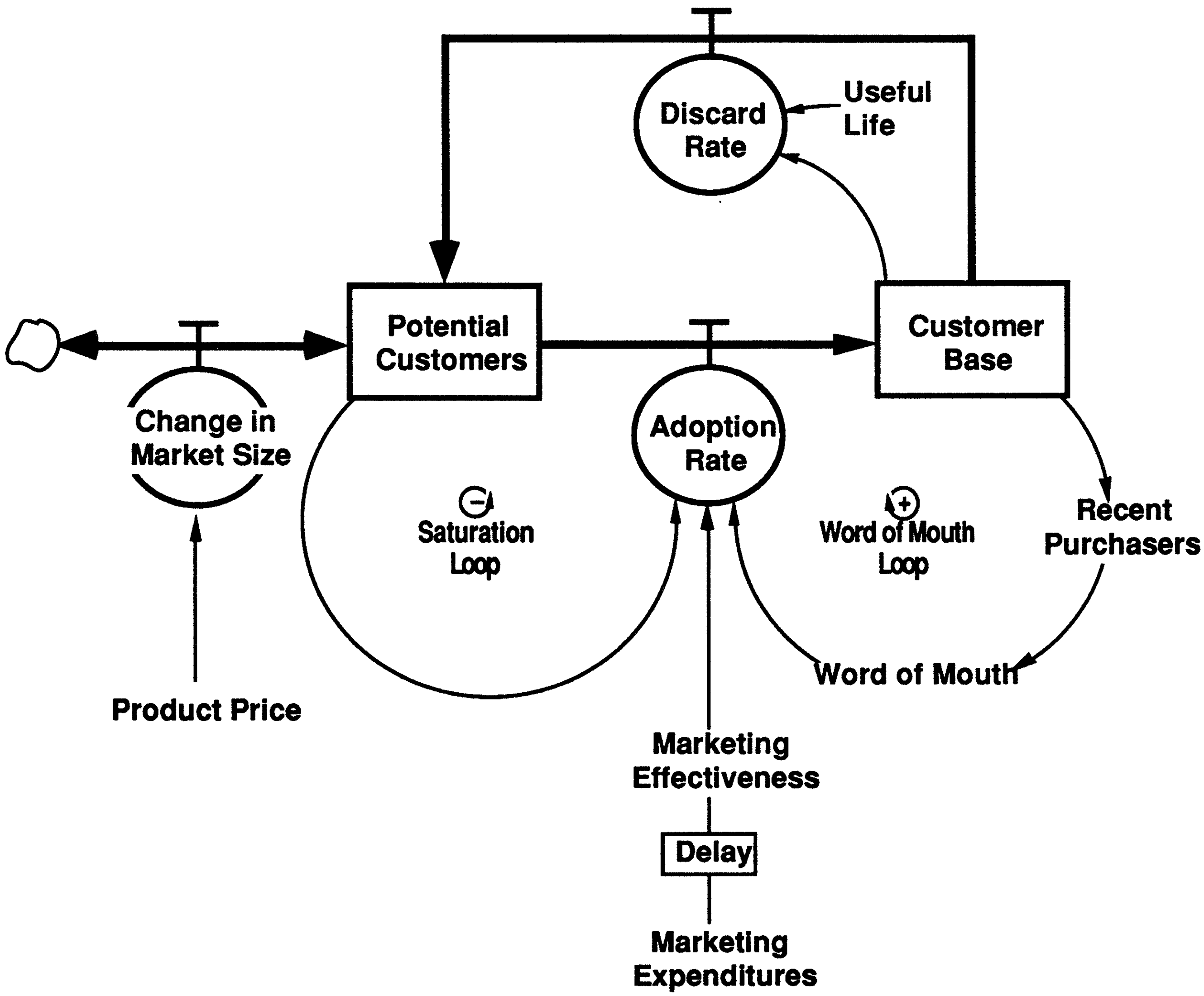

Figure 3.

The two treatment variables, the strength of the word of mouth feedback and the repurchase fraction, were varied from half to double their base case values to produce five scenarios (A-E) with orthogonal (log) values of the two treatments. The figure illustrates the impact of these treatments on market dynamics through simulations assuming no capacity constraints and constant margin pricing; actual demand patterns depend on subject decisions. When word of mouth is strong (scenarios B and D) demand grows rapidly, peaks at a large value, and declines sharply to equilibrium compared to those cases where word of mouth is weak (C and E). When the replacement interval is long (D and E), the drop to equilibrium is large, while frequent replacement leads to modest declines (B and C). The slight variation in final orders between scenarios D and E and between scenarios B and C is due to small differences in cumulative production, and hence costs, price, and demand, despite the same repurchase interval.

Repurchase Fraction/ Base Case Value \ Strength of Word of Mouth/ Base Case Value	0.5	1	2
0.5	E		D
1		A	
2	C		B

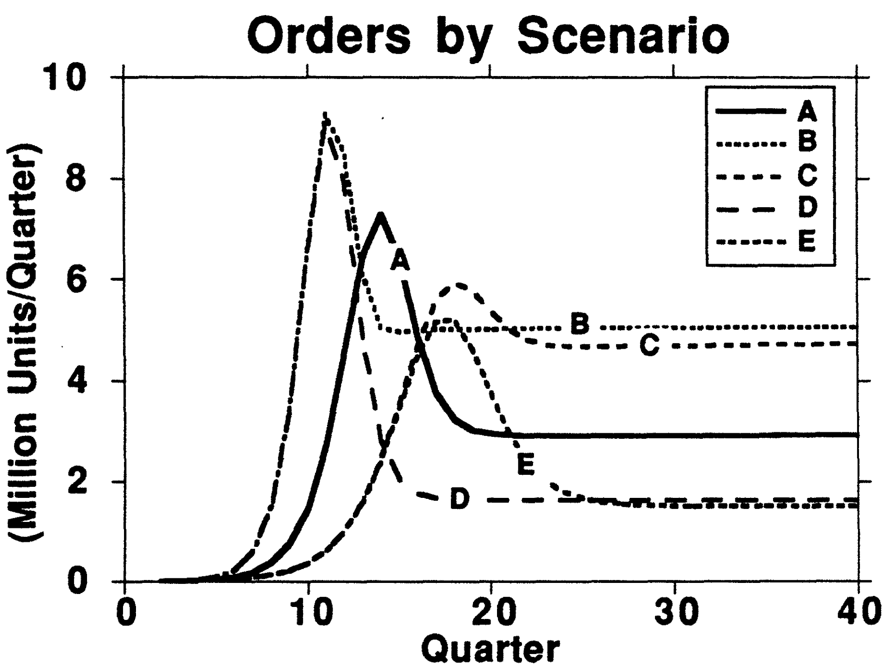

Figure 4

Behavior of typical subject in the first trial. The subject faces the most difficult scenario (D) with strong word of mouth and low replacement demand. Compare to Figure 1.

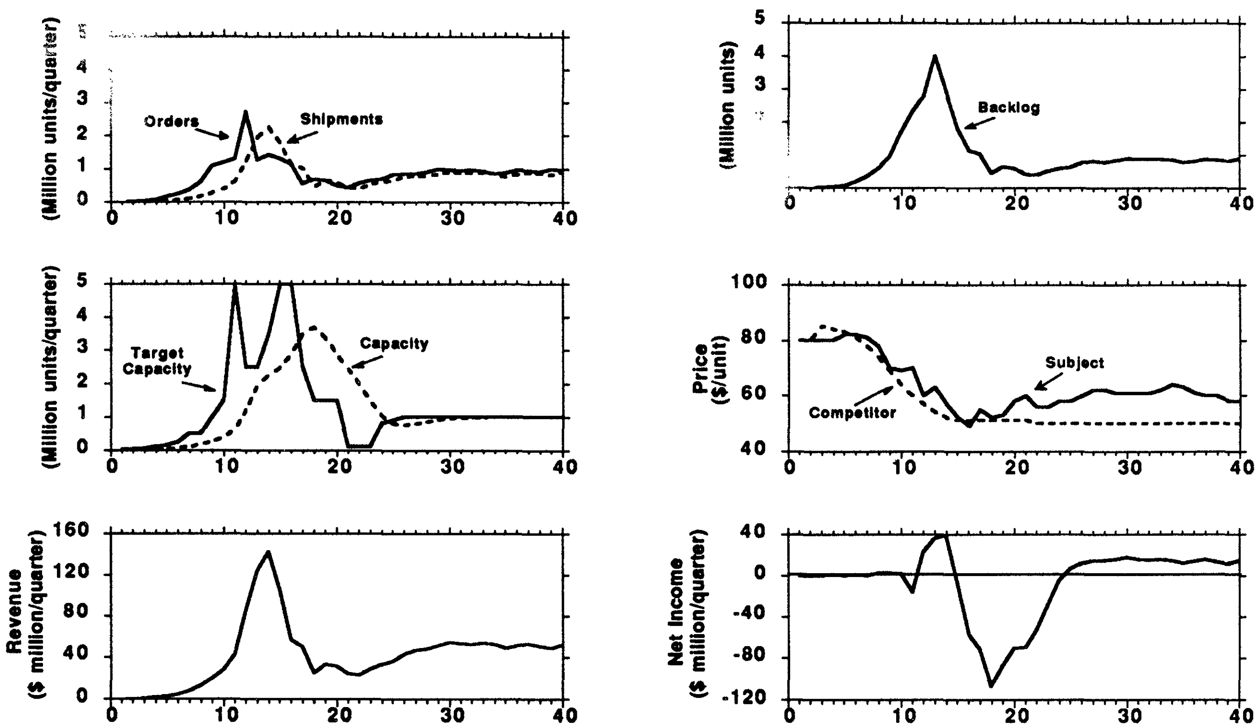

Figure 5. Effect of Word of Mouth and Replacement Fraction treatments on relative performance. Strong word of mouth and low replacement rates degrade subject performance relative to the benchmark. The significant interaction shows that the combination of strong word of mouth and low replacement is particularly troublesome.

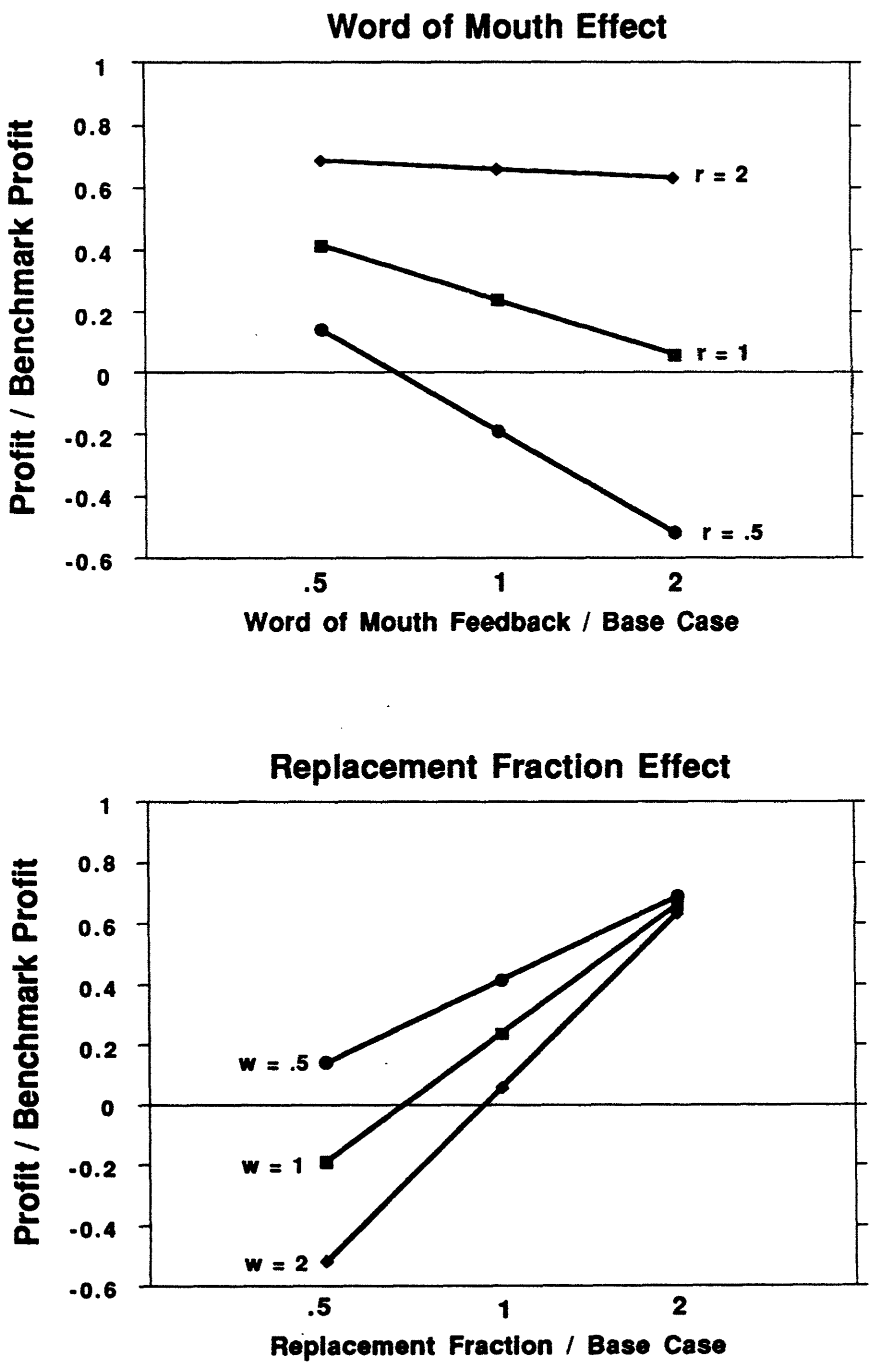

Figure 6. Effect of experience on performance. Average performance increases with experience from 13% to 63% of the behavioral benchmark. Also shown is the "perfect foresight" benchmark which results from assuming perfect knowledge of future demand in the capacity decision rather than the demand forecast produced by the behavioral rule. (The slight variation in the values of the benchmarks over trials arises because the design was not perfectly balanced.)