Strokes of Organizational Genius? Exploring the Cause of Punctuated Equilibrium in Organizational Learning

Charles Weber
Department of Engineering and Technology Management
Portland State University

ABSTRACT

Do learning organizations have strokes of genius? An empirical study of 34 high technology R&D and manufacturing organizations suggests not. The roots of punctuated equilibrium in organizational learning can be traced to learning activities that occur within organizational subsystems, primarily during R&D. Continuous improvement at the subsystem level contributes significantly to a delayed, rapid surge in organizational performance. Managers coordinate subsystem-level activities to maximize organizational performance by trading off the revenues expected from timely learning against the expected costs. Knowledge accumulated within organizational subsystems can remain hidden from organization-level performance metrics for prolonged periods of time.

1. INTRODUCTION

Organizational learning theory has successfully characterized industrial activities in which unit labor cost or unit cost of production continuously decreases at a decreasing rate as organizations gain production experience (e.g. Argote and Epple, 1990). This phenomenon, which is attributed to increasing skill in production, is generally referred to as learning by doing (Arrow, 1962) or the learning curve. Organizational learning theory has been expanded to cover the observed variability in learning rates (e.g. Dutton and Thomas, 1984; Argote and Epple, 1990; Hayes & Clark, 1985). However, to date, organizational learning theory cannot completely explain radical, discontinuous improvement in organizational performance, which occurs in high technology manufacturing industries such as pharmaceuticals (e.g. Pisano, 1994, 1996), disc drive fabrication (e.g. Bohn and Terwiesch, 1999) and semiconductors (e.g. Terwiesch and Bohn, 2001). In these industries, organizational performance is negligible for a prolonged period of time, rises sharply to high levels in

a relatively short period of time, and (under ideal circumstances) saturates near an optimal level. It appears as if a stroke of organizational genius terminates a long period of organizational ignorance.

In this paper, I investigate radical, discontinuous organizational learning as it relates to the high technology R&D process and its immediate aftermath. After reviewing pertinent literature (§2), I apply the theory of punctuated equilibrium (Abernathy and Utterback, 1978; Tushman and Romanelli, 1985; Gersick, 1991) to organizational learning. I develop a theoretical framework in which continuous improvement performed by organizational subsystems enables rapid surges in organizational performance that punctuate prolonged periods of stagnating performance (§3). In §4, I describe an exploratory empirical study, which I designed to test the validity of the proposed theoretical framework. I use case study research methods (Yin, 1994; Eisenhardt, 1989) to investigate organizational learning in the VLSI (very large-scale integrated) circuit manufacturing industry, in which organizational performance can be decomposed multiplicatively into subsystem-learning activities (Bohn, 1995), and in which rapid surges in organizational performance are known to occur (e.g. Stapper and Rosner, 1995; Weber, et al., 1995; Leachman, 1996; Leachman and Hodges, 1996; Weber, 2004). In §5, I develop an analytical model of the lifecycle of a VLSI circuit manufacturing process from the empirical findings of the study.

Existing organizational learning theory cannot completely explain my empirical findings, which imply that rapid surges in organizational performance are not the consequence of short bouts of intense learning. Instead, managers coordinate subsystem-level learning activities to maximize organizational performance – they trade off the revenues expected from timely learning against the expected costs. Contrary to the observations of Gersick (1988), the rapid rise in organizational performance, which takes place in the final stages of R&D, largely constitutes a delayed reward for prior, prolonged, continuous improvement efforts that transpire within organizational subsystems.

In §6, I discuss the implications of this study's findings, which apply to all organizations in which one weakly performing subsystem can severely constrain the performance of the organization

as a whole. These findings point to changes in R&D practices, which could improve the effectiveness of the R&D process. For example, the model of the VLSI circuit process lifecycle indicates that knowledge, which is accumulated within organizational subsystems during R&D, can remain hidden from organization-level performance metrics for prolonged periods of time. Consequently, managers need to monitor subsystem-level performance metrics to become aware of subsystem-level knowledge – failure to do so may result in gross strategic blunders. To optimize performance under 'urgency' (e.g. Gersick, 1988), learning organizations should deploy performance metrics that contain a time-dependent revenue component, as well as a cost component. Finally, additional research that could lead to a more comprehensive theory of organizational learning is suggested. In particular, further investigation of subsystem-level learning may reveal that organizational subsystems possess more know how and know why (e.g. Bohn, 1994, pp. 62-64) than organization-level performance variables would indicate.

2. PERTINENT PRIOR WORK

Numerous early studies in organizational learning (e.g. Wright, 1936; Searle and Gody, 1945; Alchian, 1963; Rapping 1965; Hayes and Clark, 1985) suggest it to be a continuous process during which organizational performance improves at a decreasing rate as production experience increases. Dutton and Thomas (1984), who analyzed over 200 learning curves, observed a high variability in learning rates, which has been attributed to phenomena such as 'organizational forgetting' (Argote, et al., 1990); employee turnover (e.g. Argote and Epple, 1990); knowledge transfer (Argote, et al., 1990; Hatch and Mowery, 1998); and scale (e.g. Argote and Epple, 1990). However, many studies have shown that organizational learning is not inherently continuous. For example, Hirsch (1952) and Baloff (1970) observed that unit costs were higher after an interruption in production such as a strike. Adler and Clark (1991) enhanced the analysis of learning curves by introducing two managerial variables: the cumulative number of hours that workers spent on training and the cumulative hours an organization spends on engineering changes. In their study of an anonymous high technology manufacturing firm, the authors discovered that cumulative training and engineering could enhance as

well as disrupt total factor productivity. Hatch and Mowery (1998) analyzed quality data from 52 semiconductor processes, showing that cumulative engineering significantly enhanced learning rates. However, when new processes were introduced into manufacturing, cumulative engineering could disrupt ongoing learning in existing processes.

According to Terwiesch and Bohn, (2001, p. 1), "many high tech industries are characterized by shrinking product lifecycles, [as well as] increasingly expensive production equipment and up-front cost. … These forces pressure organizations to cut not only their development times (time-to-market), but also the time it takes to reach full production volume (time-to-volume), in order to meet their financial goals for the product (time-to-payback)." Learning in high technology industries is thus characterized by *a sense of urgency*: it is in the interest of high technology firms to begin the learning process as early as possible and to ramp to production volume as rapidly as possible.

Evidence for an early start in learning comes from studies of process R&D in the pharmaceutical industry (e.g. Pisano, 1994, 1996), which have shown that firms may acquire production skills prior to introducing a product into the factory. This phenomenon, which Pisano (1996) calls 'learning before doing', occurs through computer simulations, laboratory experiments, prototype testing, pilot production runs and other experiments. The intent is to facilitate a seamless transition between research, development and production: if many quality and production issues are settled before product introduction, then the ramp to production can be viewed as an increase in scale during which little, if any, learning is required. Learning before doing primarily occurs in environments such as chemical synthesis, where underlying industrial knowledge is deep. By contrast, organizations in the biotechnology environment, for which the underlying theoretical and practical knowledge is relatively thin, rely on learning by doing for efficient development. However, von Hippel and Tyre (1995) argue that not all learning can occur before doing. In their study of 27 problems that affected two novel process machines in their first years of use in production, the authors discovered that many problems could not be resolved prior to field use, because existing

problem-related information could not be identified in the midst of complexity, and because new problem related-information is introduced by users and other problem solvers, who learn after the machine has been introduced into the field.

Terwiesch and Bohn (2001) investigate learning in semiconductor manufacturing, which like chemical synthesis has a deep underlying knowledge base. The authors have observed that a significant amount of learning occurs during production ramp-up when resources are scarce, production capacity is constrained, the R&D process is not complete, the production process is still poorly understood, but products can be sold at a high price. Under these circumstances, the semiconductor manufacturer has an incentive to learn to improve yield as rapidly as possible, as well as to ramp up to full production capacity at the fastest rate possible. However, these goals may be at cross purposes – ramping rapidly may lower yield, whereas launching many experiments for the purpose of improving yield reduces production capacity. Nonetheless, Terwiesch and Bohn (2001) conclude that during ramp-up, earlier learning through experimentation is more valuable than later learning -- in spite of a high opportunity cost of experimentation -- because the price for semiconductor products tends to erode rapidly.

Gersick (1988) observed that the performance of product-development teams operating under a sense of urgency does not improve continuously. In the early stages of a product-development project's lifecycle, different teams pursued a variety of approaches, which tended not to improve their performance significantly. About halfway through the projects' duration, the teams developed a sense of urgency to complete their respective assigned tasks. Deadline pressure triggered a transition meeting, after which the teams fundamentally changed their mode of operation to solving task-related problems. Organizational performance improved radically. It appears as if a stroke of organizational genius, which terminated a long period of organizational ignorance and enabled very rapid learning, occurred during the transition meeting.

Theories that viewed learning as a continuous process could not explain Gersick's (1988) observations, motivating Gersick (1988, 1991) to apply the theory of punctuated equilibrium (Abernathy and Utterback, 1978; Tushman and Romanelli, 1985) to organizational learning. The punctuated equilibrium model of change assumes that long periods of small, incremental change are interrupted by brief periods of discontinuous, radical change. Fundamental breakthroughs such as DNA cloning, the automobile, jet aircraft, and xerography are examples of radical change (Brown and Eisenhardt, 1997), which can enhance or destroy the competencies of incumbents (Tushman and Anderson, 1986) and fundamentally alter an industry (Gersick, 1991; Utterback, 1994). However, many organizations have learned to "continuously change and thereby to extend thinking beyond the traditional punctuated equilibrium view, in which change is primarily seen as rare, risky, and episodic, to one in which change is frequent, relentless, and even endemic to the firm (Brown and Eisenhardt, 1997, p. 1)." Effective managers link current projects to the future with predictable (time-paced rather than event-paced) intervals, familiar routines and choreographed transition procedures (Gersick, 1991; Brown and Eisenhardt, 1997), enabling organizations continuously improve their performance and to continuously adapt to changes in the environment.

Both continuous improvement and radical, discontinuous improvement in organizational performance commonly take place in high technology manufacturing industries such as pharmaceuticals (e.g. Pisano, 1994, 1996), disc drive fabrication (e.g. Bohn and Terwiesch, 1999) and semiconductors (e.g. Terwiesch and Bohn, 2001). A comprehensive theory of organizational learning must incorporate therefore both phenomena. It is the purpose of this paper to gather empirical evidence that could lead to the development of such a theory.

3. THEORETICAL FRAMEWORK

In this section, I take a point of view that synergizes punctuated equilibrium theory with continuous improvement. I submit that radical, discontinuous improvement in organizational performance is not necessarily a consequence of a short period of intense learning. Instead, I propose

that this phenomenon *is largely caused by continuous improvement efforts that are performed by the subsystems of an organization*. I argue that this guiding proposition is consistent with existing theories of continuous improvement (e.g. Zangwill and Kantor, 1998; Lapré et al., 2000), and I suggest that it can be tested empirically.

3.1 Continuous Improvement through Waste Reduction

Zangwill and Kantor (1998) present a framework for continuous improvement and the learning curve, which is based on a series of head-to-tail learning cycles in which each cycle contributes incrementally to the reduction of "errors, wastes and other inefficiencies that impair the operations of the [production] process" (ibid, p. 911). According to Zangwill and Kantor (1998, pp. 917-918), the performance metric of a continuous improvement process M(q) can be expressed in terms of the differential equation

$$dM(q)/dq = -c(M(q) - M^*)^{\kappa+1} \qquad (1),$$

where M* designates the metric's optimal value; c is a coefficient; 'κ' represents a parameter that reflects the effectiveness of management; and 'q' denotes an experience variable. In equation (1), the quantity '|M(q) – M*|' represents the magnitude of the non-value-added (NVA) or 'wasted' component of M(q), a performance gap that closes with increasing production experience. The shape of M(q) depends upon the value of κ. When $\kappa<0$, M(q) takes the shape of the power function associated with the traditional learning curve (e.g. Argote and Epple, 1990). When $\kappa=0$, M(q) becomes an exponential function that approaches M* asymptotically. When $\kappa>0$, equation (1) generates exponential functions, which reach their optimal value M* at a finite amount of production experience q.

Lapré et al., (2000) have conducted a study in a factory that produces tire cord, where performance is characterized in terms of yield rates and waste rates whose range is restricted from naught to unity. A yield rate Y(q) is defined as $M(q)/M_{max}$, the ratio of the value of the performance

metric at a production experience q to the performance metric's theoretical maximum value. The waste rate of is given by the quantity '1-Y(q)', which approaches naught as continuous improvement efforts reduce waste and drive Y(q) towards unity. If continuous improvement causes '1-Y(q)' to decrease at a decreasing rate with increasing production experience, then Y(q) should be a concave, monotonically increasing function of q. Thus observing a concave, monotonically increasing yield rate can be considered evidence that a continuous improvement effort may cause the increase in the yield rate. If, on the other hand, Y(q) is not a concave, monotonically increasing function of q, then the behavior of Y(q) is unlikely to be caused by continuous improvement. If the benefits of continuous improvement must be traded off against the costs of achieving them, as Lundvall and Juran (1974) and Chase and Acquilano (1981) have argued for the cost of quality, and the costs of continuous improvement are significant, then the maximum may not be the optimum. The optimal value for a yield rate 'Y*' may be less than unity, and the optimal value for the waste rate '1-Y*', may be greater than naught. The optimal value for a yield rate would reside at some point beyond which the marginal costs of its continuous improvement would exceed the marginal revenues that further continuous improvement would generate.

The reduction of errors, waste and inefficiencies can take many forms. It may consist of a variety of activities such as removing the defects in an automobile production line using the Kaizen approach; applying Total Quality Methods in a factory that manufactures tire cord to reduce the frequency of cord fractures that occur at each process step (Mukherjee et al., 1998; Lapré et al., 2000); improving the yield of a manufacturing process by reducing process noise (Bohn, 1995); increasing the production rate by reducing equipment downtime; replacing a policy of sacrificing product for analytical purposes with inspection policies for quick feedback on the quality of the manufacturing process (e.g. Tang, 1991); reducing excess work-in-progress inventory to improve yield (e.g. Wein, 1992); or maximizing a factory's contribution to the bottom line by improving the

criteria as to whether to pass a partially defective batch of products, to rework the batch or to remove the batch from the manufacturing line (Bohn and Terwiesch, 1999).

3.2 Subsystem-Level Learning and Organization-Level Performance

Zangwill and Kantor's (1998) theory of continuous improvement by waste reduction may provide a framework for explaining punctuating surges in organizational performance, if it is applied to the level of organizational subsystems. Zangwill and Kantor (1998, p. 918) postulate that that all solutions to equation (1) can be decomposed additively into sub-metrics whose sum equals M(q), and that sums of finite exponential forms ($\kappa>0$) can approximate all solutions to equation (1). Therefore, according to Zangwill and Kantor (1998), the organization-level performance of any continuous improvement effort can be estimated from the sum of the performance of subsystem-level continuous improvement efforts. The weakly performing subsystems prevent the organization as a whole from performing at its optimal level, but no single subsystem can restrict organizational performance to negligible levels.

In industries such as tire cord production (Mukherjee et al., 1998; Lapré et al., 2000), disc drive fabrication (e.g. Bohn and Terwiesch, 1999) and semiconductor manufacturing (e.g. Bohn, 1995), organizational performance is decomposed multiplicatively, i.e. the organization-level performance metric $Y_{org}(q)$, a yield factor, is the multiplicative product a set of sub-metrics, where each sub-metric $Y_k(q)$ measures the performance of a subsystem 'k' as a yield factor, and 'k_{total}' denotes the total number of subsystems in the organization. $Y_{org}(q)$ is given by the expression

$$Y_{org}(q) = \prod_{k=1}^{k_{total}} Y_k(q) \qquad (2).$$

If all $Y_k(q)$ in equation (2) measure continuous improvement efforts that take the shape of concave, monotonically increasing functions, then the multiplicative product of these functions will not be a concave, monotonically increasing function (Weber, 2003, Ch. 5). Instead, continuous improvement at the subsystem level has the potential of generating a punctuated equilibrium at the

organization-level – i.e. a prolonged period of weak organization-level performance during R&D and a prolonged period of strong organization-level performance during volume production are interspersed by a short period of rapid improvement in performance that occurs during the very late stages of R&D. External observers, who have no knowledge of the organization's internal learning mechanisms, would detect a surge in organization-level performance, and perhaps interpret it as the result of period intense learning that occurs during the final stages of the R&D process or simply a stroke of organizational genius. Instead, I submit that *in an industry in which organizational performance can be decomposed multiplicatively*, *a punctuated surge in organization-level performance constitutes a delayed reward for a prolonged investment in continuous improvement at the subsystem level,* which can occur during research and development. This assertion is confirmed, if the following propositions are confirmed.

Proposition 1a: $Y_{org}(q)$ remains relatively close to naught, until all its constituent yield factors $Y_k(q)$ achieve values that significantly exceed naught, and $Y_{org}(q)$ cannot approach unity (or its optimal level) until all its constituent yield factors approach unity (or their optimal levels). Consequently, $Y_{org}(q)$ inherently lags behind its constituent yield factors.

Proposition 1b: The constituent yield factor with the weakest performance will have the greatest (limiting) impact on $Y_{org}(q)$, whereas relatively well-performing subsystem level learning efforts have proportionally less impact. This proposition is supported, if $Y_{org}(q)$ does not lag significantly behind the constituent yield rates (factors) with the weakest performance, but lags significantly behind the constituent yield rates (factors) with the strongest performance.

In an urgent environment, in which the unit price of the product to be sold deteriorates over time, as many units of product need to be produced as soon as possible (Terwiesch and Bohn, 2001). The performance metric for an organization operating in such an environment can be given in terms of the yield rate 'Q(t)', which is defined as the number of units of product that are produced per unit time at a particular point in time 't' divided by the maximum number of units of product per unit time

that the organization can possibly produce. (Using time as an experience variable has been shown to be particularly effective in characterizing learning processes that to a great degree depend upon learning how to execute tasks involving complex operations (e.g. Fellner, 1969; Dudley, 1972).) If the performance of said organization can be decomposed multiplicatively, equation (2) can be rewritten as

$$Q(t) = \prod_{k=1}^{k_{total}} Y_k(t) \qquad (3),$$

where the constituent yield factors $Y_k(t)$ are concave, monotonically increasing functions of t, of which each represents the performance metric of a subsystem-level continuous improvement effort. Equation (3) suggests that *an organization with a sense of urgency pursues its various subsystem-level continuous improvement efforts in parallel and coordinates them in a manner that causes Q(t) to surge towards unity as early as possible (**Proposition 2**)*. This proposition is supported if the constituent yield factors $Y_k(t)$ are observed to be concave, monotonically increasing functions of t that run in parallel, and Q(t) surges towards its optimal value Q* once all $Y_k(t)$ are nearing the completion of gradual approaches to their optimal values.

4. RESEARCH METHODS

In this section, I describe an empirical study that investigates punctuated equilibrium in organizational learning. Beginning with the propositions from §3, I use the case study research method (Yin, 1994; Eisenhardt, 1989) to explore how, why and under which circumstances subsystem-level learning can generate rapid surges in organizational performance. Organizations and their subsystems consequently constitute the units of analysis of the study. I have chosen VLSI circuit R&D and manufacturing organizations as the setting of the study because the relationship between subsystem performance and organization-level performance is clearly defined in the VLSI circuit manufacturing industry (Bohn, 1995, p. 33), and because performance at both levels varies substantially over the lifecycle of a VLSI circuit production process. Thus the effect of subsystem-

level activities should yield clearly discernable patterns at the organization level, which can hopefully be linked to the aforementioned theoretical propositions from §3.

4.1 The Research Setting: VLSI Circuit Manufacturing

VLSI circuit manufacturing (like most other forms of semiconductor manufacturing) relies on batch processing (e.g. Bohn, 1995, p. 33). Silicon wafers act as batches for products, which are known as integrated circuits (ICs), "chips" or "dice". The number of products per batch 'N' can vary from as few as 50 to as many as 400. The manufacturing line (as well as the R&D line on which a production process is developed) consists of a series of fabrication, metrology and inspection steps, which are executed on process, metrology and inspection equipment, respectively; there are no assembly steps in the line. Integrated circuits can be destroyed (wasted) by process parameters that are out of control, as well as by 'micro-contaminants' – particulate matter that can be observed in volumes smaller than 1/1000 of a cubic micrometer. Due to the threat of micro-contaminants, all three of the abovementioned types of equipment, reside inside a clean room fabrication facility called a 'fab'. All chips are tested for functionality and electrical characteristics at the end of the manufacturing line, when they are still part of a silicon wafer but can be safely removed from the fab. A chip is declared to be non-functional, if one of the many variables that are tested exceeds specified limits. Such an anomalous electrical signal is known as a 'fault'; its detection precludes an integrated circuit product from being sold.

The VLSI process lifecycle bears significant resemblance to process lifecycles in the pharmaceutical industry (Pisano, 1994). In both industries "process development occurs somewhat in parallel with product development," and "processes go through three development phases – process research, pilot development commercial start-up" – prior to commencing with volume production (ibid, p. 90). Just as in the pharmaceutical industry, *process research* (PR) in VLSI circuit manufacturing "involves defining the basic structure of the process. … The goal of process research is to define the basic process architecture rather than the details." (ibid, p. 90) This goal is generally

achieved by running a set of small-scale experiments in a laboratory setting to select reaction sequence from a set of theoretically feasible alternatives. "*Pilot development* (PD) involves scaling up the process to some intermediate scale and selecting reaction parameters (such as timing, temperature, pressure), which optimize the efficiency of the process." (ibid, p. 90) Pilot development is much more empirical in nature than process research because it relies on the analysis of the output of pilot production runs, which are subjected to conditions that reflect actual production environment more accurately. *Commercial startup* (CS), the last phase of VLSI circuit research and development, involves ramping up the VLSI circuit manufacturing process to commercial scale. "How smoothly this phase goes, depends upon how well problem solving during research and pilot development have integrated knowledge about the factory environment" (ibid, p. 91). Similarly, subsequent *volume production* (VP) is less problematic if problem solving has been effective during the CS phase.

VLSI circuit manufacturing exhibits a series of attributes that makes punctuated learning an attractive proposition for firms who compete in that industry. Firstly, the lifetime of market opportunities for particular VLSI products is rather short, and they face eroding unit prices for the goods to be sold. As a consequence, VLSI manufacturers are under time-to-market, time-to-volume and time-to-payback pressure (Terwiesch and Bohn, 2001). They tend to operate in an environment of capacity constraint (Bohn, 1995) for a significant portion of their lifecycles, during which they have an incentive to learn as early as possible (Terwiesch and Bohn, 2001) and during which organizational performance can be equated with the product output rate. Secondly, VLSI circuit manufacturing, like chemical synthesis, relies on a deep underlying theoretical and practical knowledge base, which enables VLSI circuit manufacturers to engage in learning-before-doing during research and development (Pisano, 1996). Thirdly, VLSI circuit manufacturing, like disc drive fabrication, can be yield driven (Bohn and Terwiesch, 1999), a happenstance from which VLSI circuit manufacturers can derive significant contributions to the bottom line by engaging in waste reduction efforts (such as the ones described in Lapré et al., 2000). Last but not least, the technical

requirements of VLSI circuit manufacturing advance along a set of dimensions of merit known as the Moore's Law trajectory.[1] These requirements and the dates by which they need to be fulfilled are detailed and forecasted in an industry-wide publication known as the International Technology Roadmap for Semiconductors (ITRS).[2] Following the guidelines of this biennially published document gives VLSI circuit manufacturers and their suppliers a multi-year technology-planning horizon, which facilitates the execution of prolonged, coordinated subsystem-level learning activities.

Organizations that partake in VLSI circuit manufacturing or process development typically engage in three functions that involve learning within organizational subsystems. The first function, *process learning,* is generally performed by process engineering subsystems. It primarily consists of improving the quality of the manufacturing process by increasing control over key process parameters and eliminating the causes of faults. The second, *production quality learning*, increases the amount of material that survives the manufacturing line, whereas the third, *production volume learning*, increases the line's throughput. Both production quality learning and production volume learning require the acquisition of logistical skills and good plant-wide operating practices, especially in material handling (Bohn, 1995, p. 33); both are generally considered the responsibility of production subsystems.

Learning in VLSI circuit manufacturing and process development can be expressed in terms of subsystem-level performance variables, which use time as the experience variable (e.g. Stapper and Rosner, 1995; Weber, et al., 1995; Leachman, 1996; Leachman and Hodges, 1996; Weber, 2004). Process learning tends to lead to an increase in *batch yield* '$Y_B(t)$', which is defined as the fraction of products within a batch that function to specification at the end of the manufacturing line or the fraction of products within a batch that was not wasted in the completion of the industrial process.[3] An increase in batch yield is correlated with a reduction in *batch fault density* '$F_B(t)$', which is defined as the number of faults contained in a batch. Successful production quality learning tends to increase *line yield* '$Y_L(t)$', which is defined as the fraction of batches that survive the manufacturing

(or R&D) line; the fraction of batches that is not wasted in the line; or "the fraction of [batches] that go all the way through the process without being irreparably damaged by breakage, gross processing errors or other systematic problems (Bohn, 1995, p. 33)." Successful product volume learning increases the *line throughput rate* '$T_L(t)$', which is defined as the number of batches that exit a manufacturing (or R&D) line per unit time, assuming that no batches have been lost during the manufacturing process, divided by the maximum number of batches that can exit the line per unit time.[4] The product output rate 'Q(t)' – the primary organization-level performance metric of VLSI circuit manufacturing – can be decomposed multiplicatively into $Y_B(t)$, $Y_L(t)$ and $T_L(t)$ (Bohn, 1995, p. 33).

$$Q(t) = Y_B(t)\, Y_L(t)\, T_L(t) \qquad (4).$$

In equation (4), $Y_B(t)$, $Y_L(t)$, $T_L(t)$ and Q(t) denote yield rates whose range is restricted from naught to unity for all t. The quantities '$1-Y_B(t)$', '$1-Y_L(t)$', '$1-T_L(t)$' and '1-Q(t)' represent waste rates or non-value-added complements of $Y_B(t)$, $Y_L(t)$, $T_L(t)$ and Q(t), respectively. Therefore, waste is reduced when $Y_B(t)$, $Y_L(t)$, $T_L(t)$ and Q(t) move away from naught and towards unity (e.g. Lapré et al., 2000, pp. 603-605). Batch fault density can be expressed as a yield rate for purpose of comparison to other yield rates, if it is given in terms of the *batch fault yield rate* $Y_F(t) = 1-F_B(t)/F_{B\_max}$, where F_{B_max} denotes the maximum batch fault density that can be measured. The NVA component of the batch fault yield rate is given by '$F_B(t)/F_{B\_max}$', a term that goes from unity to naught as waste is reduced.

4.2 Data Sources, Data Collection and Data Analysis

The primary source of data for this study consists of 34 cases of organizational learning and problem solving, which transpired between 1982 and 1999 in 34 VLSI circuit process research-, development- and manufacturing organizations located in Asia, Europe and North America. The organizations under study can be considered wholly owned assets of 15 VLSI circuit manufacturing

firms from Asia, Europe and North America. Thirty-one respondents who were personally involved in the cases recounted them in personal, one-on-one interviews.

The respondents answered specific questions concerning the numerical values of subsystem-level and organization-level performance metrics at the time their cases transpired. For the purpose of generating an analytical model, the responses to these questions were classified as ***low***, if they resided close to naught (yield rate ≤ 0.1); as ***high***, if they resided near unity (yield rate ≥ 0.85); or as ***medium***, if they were neither near naught nor unity ($0.1 <$ yield rate < 0.85). The respondents were also asked questions concerning the trends and the shape of the curves of all performance metrics. For the period of time that they witnessed these cases, the respondents were allowed to answer whether 'on balance' the yield rate trajectories would be rising, falling or flat, and whether they were rising or falling at an increasing or decreasing rate. This line of questioning would elicit responses that focus on long term trends instead of process noise, which tends to be prevalent in semiconductor manufacturing (Bohn, 1995). (Serious setbacks in process learning, production quality learning and production volume learning can cause the yield rate trajectories to drop below their expected values for periods of time that can potentially be measured in weeks. Conversely, the solution of a major problem can create small surges in the yield rates.) Consequently, when the phrase 'on balance' is used in §5 of this paper, it refers to an observable long-term trend from which an analytical model of the VLSI circuit process lifecycle can be built.

To calibrate the position of a case within its specific VLSI process lifecycle, the respondents were requested to estimate the time that had elapsed since (or was expected to elapse until) the anticipated release date for the first product to be realized by processes of the kind that were under study. In addition, respondents were asked open-ended questions whose answers provided a more detailed explanation of how the organization in each case learned and how the context of each case – a specific production environment during a particular phase of the VLSI circuit process lifecycle –

affected organizational learning. (Please see appendix A for a more detailed description of the questions that the respondents were asked.)

To avoid undesired confusion between the unit of data collection (individuals) and the units of analysis (the learning organization and its subsystems) (Yin, 1994, pp. 75-76), data from the above case studies have been compared to data from secondary sources. These include published semiconductor benchmarking studies (Leachman, 1996; Leachman and Hodges, 1996); historical performance data of individual VLSI circuit manufacturing firms (Stapper and Rosner, 1995; Weber, et al., 1995); and projections of technical trends, which have been printed in the more recent editions of the Semiconductor Industry Association's technology roadmaps for semiconductor manufacturing.[2]

The case study data have been coded and displayed in a manner that reveals how subsystem-level and organization-level performance metrics vary as a function of maturity of the manufacturing process (please see table 1). Two of the organizations under study were engaged in process research; the respondents describing these organizations expected to release their first product into production within two to three years. Six organizations were pursuing pilot development; in these organizations the first product release was expected within less than a year. Nine organizations were involved in commercial startup; their first product release was less than a year behind them. Seventeen organizations were engaged in volume production; their first product release was more than eighteen months behind them.

Table 1 indicates that theoretical saturation was achieved relatively easily for cases pertaining to the early phases of the process lifecycle. Within-case analysis provided information regarding the relative value of different yield rates within a short time interval. Cross-case analysis revealed consistent patterns regarding the behavior of the particular subsystem-level and organization-level performance metrics over time, as well as insight into how organizational learning practices changed as the VLSI circuit manufacturing process matured. The picture was less clear during volume

production primarily because in many instances the respondents had difficulty determining precisely when volume production began or what constituted the optimal level for the previously discussed performance metrics during volume production. However, all 17 respondents reciting cases in which the initial product release was at least 18 months behind them agreed that all of the abovementioned performance metrics had achieved or had begun to converge on some optimal value, even though the vast majority of organizations under observation were still practicing some form of continuous improvement.

An analytical model of organizational learning of the kind that VLSI circuit manufacturers of the 1980's and 1990's were likely to experience has been derived from within-case analysis and cross-case analysis. The model, which is displayed in Figure 1, tracks the previously identified subsystem-level and organization-level performance metrics throughout the lifecycle of a VLSI circuit manufacturing process, from the early stages of process research through mature volume production. Product release in this model is defined to occur at t=0. The model has been validated in interviews with 61 experts in VLSI circuit manufacturing, VLSI circuit process development and related technical fields. These experts were recruited by recommendations from within their respective peer groups. In addition, some of the respondents that have recited a case have acted as experts for cases other than their own. The technical complexity of VLSI circuit manufacturing and process development, which results from integrating a multitude of scientific disciplines (Iansiti, 1998), has warranted such a large number of expert interviews. For example, an expert in materials engineering would not necessarily be able to validate the relevance of a case that revolved around optics and polymer chemistry. Similarly, cases in which production management skills were germane to performance improvement required validation by experts in operations management, logistics or supply chain management, rather than validation by technical experts.

{Please insert table 1 and figure 1 about here.}

5. EMPIRICAL FINDINGS AND ANALYTICAL MODEL

In this section, I present significant results from the study that has been described in §4, and I compare the empirically grounded analytical model from figure 1 to the theoretical propositions from §3.2. Patterns suggesting agreement between the empirically derived analytical model and the theoretical propositions imply that these propositions possess explanatory power. Patterns suggesting disagreement between the empirically derived analytical model and the initial trial propositions are considered evidence that further theory needs to be developed. Comments by respondents provide insight into the reasons for both similarities and discrepancies between theory and observation, which can hopefully assist in the development of new theory.

Empirical evidence from the study in this paper suggests that the subsystem-level yield rates behave very differently from each other throughout the VLSI circuit process lifecycle (see figure 1). For example, batch yield does not act as if it were driven by continuous improvement. During process research its magnitude is negligible, and the shape of the batch yield trajectories is flat. Batch yield begins to rise in the late pilot development phase in nine out of nine cases under observation and takes the shape of an S-curve during commercial startup. However, evidence from this study suggests that process learning (elimination of faults, waste reduction) begins early in the pilot development phase at a point in time 't_{PD}' when batch fault density has been reduced to a level '$F_{B\_Max}(t_{PD})$' at which it can be quantified and the analysis of the output of pilot production runs (Pisano, 1996, p. 90) can commence. (According to the respondents who documented the eight cases that transpired in PR and PD phases, batch fault density can definitely be measured at 5000 faults per batch (wafer) and perhaps as high as 20 000 faults per batch (wafer). Thus, $20\,000 < F_{B\_Max}(t_{PD}) < 5000$. This level of process quality is typically achieved between 1.5 and 1 year prior to product release. Thus, -1.5 years $< t_{PD} <$ 1.0 years.) At $t=t_{PD}$, the process engineering subsystem launches a fault reduction effort that results in a subsequent exponential decay in batch fault density. This decay translates into a concave, monotonic increase in batch fault yield rate '$Y_F(t)=1-F_B(t)/F_{B\_max}(t_{PD})$', which lasts throughout pilot development and commercial startup; $Y_F(t)$ saturates near unity during volume production.

The shapes of the $Y_F(t)$ and $Y_B(t)$ trajectories in figure 1 suggest that the reduction of batch fault density constitutes a continuous improvement effort that maintains highly *nonlinear learning leverage* over batch yield, which results from an exponential relationship between batch yield and batch fault density.[5] In the domain in which $F_B(t) >> N$, which was observed in 8 organizations in this study, each unit of product is likely to contain multiple faults; batch yield is negligible; and batch fault density has low leverage over batch yield because removing just one fault from a product is not likely to cause the product to function to specification. In all 9 organizations under observation around or shortly before t=0, when $F_B(t)$ approaches N from above, respondents reported a high batch fault yield rate and a low batch yield. In the domain in which $F_B(t) \simeq N$ (the CS phase), which was observed in 9 organizations, each unit of product is likely to contain approximately one fault, and removing one fault from a product is highly likely to cause the product to function to specification. Batch yield rises dramatically in this domain, and batch fault density exerts strong leverage over batch yield. *The rapid rise in batch yield represents the culmination of a prolonged, continuous improvement effort*, which has brought batch fault density from $F_B(t)>>N$ to $F_B(t)\simeq N$. Batch yield lags 9 to 15 months behind the batch fault yield rate, suggesting that the rise in batch yield which begins slightly before product release rises sharply shortly thereafter to a large degree constitutes *a delayed reward for process learning* that has taken place prior to product release. In the domain in which $F_B(t) << N$ (the VP phase), batch fault density approaches naught, the batch fault density approaches unity, and faults become more difficult to find. On balance, $Y_B(t)$ saturates; the effect on batch yield of reducing batch fault density becomes negligible once more.

The cost of increasing batch yield is viewed as very high because identifying faults the size of a tenth of a cubic micrometer is associated with many in-line inspection steps (Tang, 1991) on very expensive equipment with an inherently low throughput rate. A yield manager from a VLSI-circuit manufacturing facility explains.

"An optical inspection tool costs as much as $3 million, and it inspects at a rate of about one wafer per hour. ... We need to purchase many of these tools to conduct inspections at a rate that will increase our [batch] yield. ... We have to amortize [these diagnostic tools] just like regular fabrication tools, even though they do not improve anything on the wafers. On the contrary, the inspections generally teach us that we may have to scrap some badly contaminated wafers."

VLSI circuit manufacturers of the 1990's operated in an environment in which reaching VP six months early could generate more than $ 1 billion in additional profits over the lifecycle of a VLSI circuit manufacturing process (Weber, 2004). They consequently had an incentive to move forward in time the anticipated surge in batch yield upon which a timely surge in product output rate and, by extension, profitability are contingent. The highly nonlinear relationship between batch yield and batch fault density enables VLSI circuit manufacturers to generate an earlier surge in batch yield by accelerating the rate of batch fault reduction. VLSI circuit manufacturers are thus inclined to pay for technology that accelerates a decline in batch fault density, and focus on process learning during pilot development. The yield manager from above explains.

"We try to get rid of as many faults as soon as we can. We try to solve the problems that we believe cause most of the faults first. We continue to solve problems until [batch fault] density is low enough for [batch] yield to rise significantly. Then we ramp up to volume production. ... This fault reduction process takes many months. ... [Our company] spends tens of millions [of US dollars] on diagnostic equipment and fault reduction practices. The investment is worth it, if it accelerates fault reduction by a few weeks.

Figure 1 suggests that line yield is driven by continuous improvement. On balance, line yield rises throughout process research and pilot development in all eight out of eight cases under observation, reaching a high level at the beginning of commercial startup. Line yield continues to increase during the CS phase, but in no case does line yield reach unity. In all cases, on balance, the shapes of the line yield trajectories are concave and monotonically increasing throughout the PR, PD and CS phases. Line yield saturates during volume production, either by flattening out completely or increasing at the margins. In 41 out of 41 synchronous, pair-wise comparisons between line yield,

batch yield and line throughput rate, which span the complete VLSI process lifecycle, $Y_L(t)$ is greater than $Y_B(t)$ and $T_L(t)$. Increases in line yield primarily result from improvement in equipment maintenance and wafer handling procedures, which do not require significant physical experimentation or investment in diagnostic tools. A senior technician in a VLSI circuit R&D facility explains.

"In the early stages of process development, we have to make sure that equipment does not destroy, damage or seriously contaminate wafers while they are being processed. This is done through rigorous preventative maintenance. Human factors like sloppy wafer handling can damage wafers, too. Therefore, better training and wafer handling procedures will increase line yield.

The skills required to improve line yield can be acquired to prior to product release, an accomplishment that is rewarded during the CS phase. The technician from above explains.

It is best to learn them [these skills] as soon as you can. When the engineers [process engineering subsystem] start to increase the chip yield [batch yield], then we [in the production subsystem] do not want become the limiting factor on the fab's output. We want the line yield to be high by then. ... Equipment maintenance skills learned during process development will help you in volume production. You are adding more units of the same equipment types during the ramp to volume production, so you tend to know what you are doing when [volume] production starts.

Figure 1 suggests that the line throughput rate does not behave as if it were driven by continuous improvement. Throughout process research and pilot development, the eight organizations observed in this study operate at relatively constant line throughput rate that is more than an order of magnitude smaller than the line throughput rate of the volume production line at which the process and its products are supposed to be manufactured (i.e. $T_L(t)$ scores a 'low' in table 1). (In the two PR-phase cases $T_L(t)$ was even lower than in the six PD-phase cases.)

In all nine cases that occurred during the CS phase, the line throughput rate required 12 to 18 months to rise to volume production levels. This prolonged period of time suggests that not all skills required for the commercial startup were acquired during pilot development, i.e. not all learning can

be done before doing. Some aspects of production volume learning have to occur in a use environment that reflects realistic production conditions on real production equipment (von Hippel and Tyre, 1995). In the words of an expert in semiconductor diagnostics:

"You cannot just crank up a wafer fab like the volume knob on your stereo. It requires some learning. You will add more equipment. You may have to add and manage additional shifts in maintenance and production. Equipment problems that do not occur when you run at low volume are likely to appear. For example, robotic loading equipment is more likely to fail if you run it perpetually without maintenance. ... You may also have to remove a few unnecessary [diagnostic] steps from the process to minimize your WIP [work-in-progress] inventory. You will have to learn how to run the fab without the information that these [diagnostic] steps reveal. ... All of this [learning] takes time."

In principle, a VLSI circuit manufacturer can save time by simulating a realistic production conditions on the factory floor prior to product release. The manufacturer could move production volume learning forward in time by ramping up to volume production conditions during the PD phase or earlier. Production volume learning would run in parallel with attempts to continuously improve the batch fault yield rate and $Y_L(t)$. A multi-year planning horizon enables the learning organization to coordinate the subsystem-level continuous improvement efforts such that $Y_L(t)$ and $T_L(t)$ could achieve their optimal values when $Y_B(t)$ begins to surge. $Y_L(t)$ and $T_L(t)$ would not delay Q(t); a surge in batch yield would translate directly into a surge in the product output rate.

Evidence from the empirical study in this paper suggests that production volume learning does not occur in parallel with process learning and production quality learning prior to commercial startup. $T_L(t)$ consistently exceeds $Y_B(t)$ in all pair-wise comparisons from the eight cases that transpired in PR phase or the PD phase, indicating that process learning rather than production volume learning is VLSI circuit manufacturers' primary source of concern during these phases. The concave, monotonic increase of the batch fault yield rate during pilot development suggests that a continuous improvement effort drives process learning during that phase, and continuous improvement line yield occurs throughout the PR, PD and CS phases. A flat line throughput rate

during pilot development suggests that process learning and production volume learning are not occurring in parallel during that phase because production volume learning is not occurring at all.

At t=0 (the beginning of the CS phase; the time of product release), nine out of nine pair-wise comparisons between $Y_B(t)$ and $T_L(t)$ revealed that $Y_B(t) > T_L(t)$, suggesting that production volume learning has replaced process learning as the primary source of concern. Production volume learning occurs during commercial startup, but the effort is not driven by continuous improvement. The shape of the line throughput rate trajectory is not concave in any of the nine cases under observation; respondents have consistently indicated a convex rise or the shape of an S-curve for $T_L(t)$. A manager at a firm, which provides yield management consulting services to many VLSI circuit manufacturers, explains why this is so.

"From what I observe at our customers, the initial ramp to volume is very painful, but gets a lot easier as time progresses. ... You are removing bottlenecks at every process step in the line. You remove the worst bottlenecks first. In the beginning, there are many bottlenecks, so your efforts do not make much difference. Once there are only a few bottlenecks left, your efforts begin to pay off. Your throughput rises dramatically."

Respondents primarily attribute the belated effort at production volume learning to a deliberate practice that is driven by high materials costs and the inability to amortize the full set of plant equipment that is required for volume production. The expected costs of production volume learning before doing may exceed the expected revenues generated by an earlier ramp to volume production. A vice president of a major VLSI circuit manufacturer, who has led many process development efforts, provides insight into the materials costs associated with learning.

"During research and development you are producing at very low [batch] yield, so you are not generating any revenue. It is as if you were producing for the garbage can. Everything you do just costs. And it costs very much. ... During R&D, processing one single wafer probably costs more than $10,000. If you have fewer than 10 good chips on a wafer -- that amounts to more than $1000 per chip. Very few chips bring in more than $100 in revenues. ... You want to run as few wafers as possible during R&D. You cannot afford to ramp up to production until your yield is moderately

high, so you have to learn how to produce at high volumes when your [batch] yield is moderately high."

A manager in a fab that runs multiple VLSI circuit manufacturing processes at various levels of maturity explains how a premature ramp to volume production impacts the utilization of fixed assets.

"You have to install the full equipment set when you ramp up to volume production. If we were to ramp up to volume production before we have significant yields, then we would not be generating much revenue. Our production equipment would be exercising its capabilities without producing anything. From an accounting point of view, the equipment is essentially idle, or worse – it is consuming without producing. … A full set of production equipment costs hundreds of millions – we cannot afford to install it unless we produce chips. … In our particular case, some production equipment runs multiple processes. If we were to ramp up a new process prematurely, we would crowd out capacity that we use for existing processes. We would be interfering with our existing production lines and lose some revenue (e.g. Hatch and Mowery, 1998). *… Accelerating the [ramp to volume production] rate by a few weeks is not worth the cost.*

According to the respondents in this study, who observed a total of 34 organizations at various stages of the VLSI circuit manufacturing lifecycle, the product output rate tends, on balance, to behave as if it were undergoing punctuated equilibrium. During PR and PD, Q(t) exhibits negligible values, and the shape of its trajectory was essentially flat. During the CS phase, Q(t) rises significantly and rapidly; in all 9 observed CS cases its trajectory takes the shape of an S-curve. Q(t) is either flat or flattening out during volume production.

The results of study in this paper dispel the notion that punctuated organizational learning results from a stroke of organizational genius, which terminates a prolonged period of organizational ignorance. Instead, empirical evidence suggests that subsystem-level continuous improvement efforts influence the behavior of organization level performance functions. Two subsystem-level continuous improvement efforts were identified: 1) production quality learning, which causes line yield to continuously increase at a decreasing rate throughout PR, PD and CS; and 2) process learning, which

causes the batch fault yield rate to continuously increase at a decreasing rate throughout PD and CS. Production volume learning has not been identified as a continuous improvement effort.

The findings of this study suggest that subsystem-level learning efforts exert highly nonlinear leverage over the organization-level performance metric -- the product output rate. In §3.2, multiplicative decomposition of Q(t) into constituent yield rates has been proposed as a mechanism for nonlinear learning leverage. Forty-one within-case comparisons of Q(t) to $Y_B(t)$, $Y_L(t)$ and $T_L(t)$ validate this proposition. In forty-one out of forty-one within-case comparisons Q(t) remains relatively close to naught until the CS phase, when its constituent yield factors achieve values that significantly exceed naught (Proposition 1a); Q(t) approaches its optimal level when its constituent yield factors approach their respective optimal levels (Proposition 1a); and Q(t) lags behind every one of its constituent yield rates in a manner that is consistent with equation (4). Figure 1 shows that this lag is asymmetric (Proposition 1b). Q(t) lags months behind $Y_B(t)$ and $T_L(t)$, the constituent yield rates that tend to limit Q(t) the most. Q(t) lags years behind $Y_L(t)$; $Y_L(t)$ is not a limiting factor.

A highly nonlinear relationship between batch fault density and batch yield has been identified as an additional mechanism for nonlinear learning leverage that was not anticipated in §3.2. The highly nonlinear relationship between $F_B(t)$ and $Y_B(t)$ postpones the rewards of process learning for prolonged periods of time. $Y_B(t)$ remains near naught until $F_B(t)$ approaches N from above. Multiplicative decomposition of the product output rate prevents Q(t) from rising above negligible levels until $Y_B(t)$ does so, which occurs when $F_B(t) \underline{\sim} N$. Consequently, process learning exerts enormous leverage over organizational performance during commercial startup.

Proposition 2 states that an organization with a sense of urgency pursues its various subsystem-level continuous improvement efforts in parallel and coordinates them in a manner that causes Q(t) to surge towards unity as early as possible. The continuous improvement efforts that drive process learning and production quality learning occur in parallel (a practice that is consistent with urgency) because the relatively low costs of early production quality learning are expected to be

less than the additional revenues generated from the resulting earlier ramp to volume production. By contrast, production volume learning is consciously deferred to the CS phase even though a deep theoretical and practical knowledge base underlies VLSI circuit manufacturing. This practice, which is apparently inconsistent with urgency, is justified because the high costs of earlier production volume learning are expected to be greater than the additional revenues generated from the resulting earlier ramp to volume production.

6. DISCUSSION

The exploratory study in this paper, which has investigated punctuated equilibrium in organizational learning, has provided some new insight into how organizational learning impacts the R&D process and its immediate aftermath. The study's findings suggest that relatively weakly performing organizational subsystems constrain organization-level performance to negligible levels of performance for a prolonged period of time, i.e. for most of the R&D process. A rapid surge in organizational performance occurs in the last stages of R&D, when no subsystem-level variable performs weakly anymore. Saturation of organization-level performance at an optimal level occurs only if and when every subsystem-level performance variable is approaching its respective optimal level. In §3.2 of this paper, these results had been predicted for industrial processes in which organizational performance can be decomposed multiplicatively into subsystem-level performance variables (e.g. Bohn, 1995; Mukherjee, et al., 1998; Lapré et al., 2000). However, the discovery of a more complex mechanism of organizational learning – one that involves highly nonlinear relationships between performance variables; performance lags; and conscious, coordinated, economically motivated delays of learning activity – suggests the results of the study in this paper apply to all organizations in which a weakly performing subsystem severely constrains the performance of the organization as a whole.

The results of the study in this paper suggest that learning organizations can enhance future organizational performance, if they are able to take advantage of the inherently nonlinear

relationships between the variables by which organizational subsystems measure their own performance (e.g. batch fault density, line yield and line throughput rate) and variables through which the learning organization as a whole measures its performance (e.g. product output rate), as well as understanding the role intermediate variables (e.g. batch yield). For example, the process engineering subsystem can engage in process learning during pilot development without achieving any near-term organization-level success, yet maintain confidence that batch yield will surge once batch fault density drops to the enabling level. Organizations that operate in an urgent economic environment can coordinate subsystem-level learning activity during R&D to deliver an optimal level of output in a timely manner. For example, the production organization can acquire production skills by engaging in production quality learning 'before doing', while, in parallel, the process engineering subsystem conducts process learning. An in depth understanding of the tradeoffs between the additional cost incurred and the additional revenue generated by a timely optimal organization-level performance allows organizations to optimally time particular learning activities. For example, the high cost of production volume learning causes VLSI circuit manufacturers to defer said activity to commercial startup, the final stage of the R&D process, during which rising batch yield generates an increasing revenue stream that renders a significant increase in the line throughput rate as economically justified.

The abovementioned conclusions impact the development of more comprehensive theories of organizational learning, which incorporate potential organizational responses to observable phenomena such as urgency (e.g. Gersick, 1988), time-sensitive market windows, shortening product lifecycles and volatile unit prices for products to be produced (e.g. Bourgeois and Eisenhardt, 1988). Organization-level performance metrics that are able to encompass these phenomena must contain a revenue component that is sensitive to time as well as a component that reflects the cost of learning. The time dependent revenue component captures the sense of urgency that drives the organizations under observation in this study to *learn to generate revenue sooner rather than later*. A performance variable containing both a time-dependent revenue component and a component that expresses the

cost of learning allows managers to make rational decisions pertaining to tradeoffs between the costs of and revenues generated by earlier learning.

The empirical findings in §5 of this paper suggest that high technology manufacturing organizations accumulate a significant portion of their production knowledge prior to product release, a conclusion that is consistent with Pisano's (1994, 1996) observations from the pharmaceutical industry. For example, figure 1 illustrates that line yield comes very close to its optimal value prior to t=0, implying that most knowledge resulting from process learning and production quality learning accumulates within organizational subsystems before a factory scales up to volume production. However, line yield, the relatively best-performing subsystem-level yield rate, barely contributes to the organization-level metric, product output rate, for most of VLSI circuit process lifecycle. The majority of all batches survive the manufacturing line for more than a year before any significant product output rate can be detected. The production quality knowledge that is required to get line yield to such high levels is not reflected in the organization-level performance variable until batch yield and line throughput rate -- the weaker subsystem-level variables -- approach their optimal levels. Similarly, the overwhelming majority of all faults are removed from the VLSI circuit process before the product output rate deviates substantially from naught. The process knowledge required to remove the majority of batch faults does not manifest itself in the organization-level performance metric. Both forms of subsystem-level knowledge remain *hidden* from organization-level performance metrics until all critical subsystem-level performance variables approach their optimal values. As a consequence, the organizational knowledge accumulated during R&D is underestimated.

Hidden subsystem-level knowledge potentially leads practitioners to major judgment errors. Figure 1 shows that, shortly prior to t=0, a high level manager who looks exclusively at organization-level metrics such as the product output rate may correctly observe that the organization as a whole is not performing to satisfaction, but may falsely conclude that the R&D process has failed and that underperformance will continue for a prolonged period of time. The manager could shut the

organization down prematurely, never knowing that strong performance would have been less than 6 months away. Had the manager also looked at batch fault density and line yield at t=0, he/she would most likely not have made such a strategic blunder.

During the process research phase of VLSI circuit process development all organization-level knowledge and some subsystem-level knowledge is 'pre-technological' (Bohn, 1994, p. 63-64). Due to low levels of process quality, neither batch yield nor batch fault density can be measured with any accuracy through physical experimentation that uses a VLSI circuit product; thus product output rate cannot be determined with any accuracy. A characterization of the full VLSI circuit process ('know how') or a scientific explanation of how the process operates over a broad region ('know why') appears elusive. Yet, figure 1 shows that by the end of the PR phase, production quality knowledge advances to the point where most batches that have been introduced into the production line are likely to complete manufacturing process without being irreparably damaged. Figure 1 also suggests that no process quality learning occurs until the end of the PR phase, a point in time by which the basic architecture of the process has been defined.

It is difficult to imagine that these achievements could have been accomplished by serendipity, by making analogies to unrelated processes or by bringing in knowledge from outside the organization, which Bohn (1994, p. 63) suggests is how pre-technological knowledge is acquired. Pisano (1994, p. 90) documents process modeling through 'thought experiments' during the early stages of process research in the pharmaceutical industry, which are followed by physical experimentation on a very small scale. In the semiconductor industry, Thomke (1998) has observed electronic simulation activities, which are replaced by physical experimentation once that mode of learning is proven to be more effective method of improving process and product quality. Both modes of experimentation occur during process research, indicating that fault reduction does occur during that stage of the process lifecycle. Significant production quality know how (and perhaps 'know why') should thus be present within the production subsystem, and significant process know how

(and perhaps 'know why') should be present in the process engineering subsystem at the end of the PR phase. Further investigation of these phenomena is warranted before a comprehensive theory of organizational learning can be built and the R&D process in high technology manufacturing can be completely explained.

The findings of the study in this paper are of importance to high technology manufacturing industries in which market timing represents a key success factor, the relationships between subsystem-level performance variables and organization-level performance variables are highly nonlinear and the cost of continuous improvement can be very high. However, the study has also raised a series of general issues concerning the R&D process, organizational learning and the acquisition of technological knowledge. These include the need to incorporate time dependence, a revenue component and a cost component into comprehensive organizational performance variables; the significance of knowledge that resides in organizational subsystems but is hidden from organization-level metrics; and the potential ability to bring subsystem-level knowledge to the 'know how' and 'know why' stage before organizational performance can be measured. The author of this paper calls for empirical research into these topics in the hope of advancing the field organizational learning towards a more comprehensive understanding of technological knowledge and the R&D process.

ACKNOWLEDGEMENTS

I would like to thank Prof. Stefan Thomke of Harvard Business School and Prof. Roger Bohn of UC San Diego, as well as Prof. James M. Utterback, Prof. Nelson Repenning, Prof. Michael Cusumano, Prof. Eric von Hippel and Prof. Edward B. Roberts, all of MIT's Sloan School of Management, for volunteering their insight into subject matter related to this paper.

LIST OF REFERENCES

Abernathy, W. J., and Utterback, J. M. 1978. Patterns of industrial innovation. ***Technology Review*** 80, 40-47.

Adler, P. S., and Clark, K. B., 1991. Behind the learning curve: A sketch of the learning process. ***Management Science*** 37(3), 267-281.

Alchian, A. 1963. Reliability of progress curves in airframe production. ***Econometrica*** 31, 679-693.

Argote, L., Beckman, S. L., and Epple, D. 1990. The persistence and transfer of learning in industrial settings. ***Management Science*** 36, 140-154.

Argote, L., and Epple, D. 1990. Learning curves in manufacturing. ***Science*** 247, 920-924.

Arrow, K. 1962. The economic implications of learning by doing. ***Review of Economic Studies*** 29, 155-173.

Baloff, N. 1970. Startup management. ***IEEE Transactions on Engineering Management*** 17(4), 132-141.

Bohn, R. E. 1994. Measuring and managing technological knowledge. ***Sloan Management Review*** 36(1), 61-73.

Bohn, R. E. 1995. Noise and Learning in Semiconductor Manufacturing. ***Management Science*** 41(1), 31-42.

Bohn, R. E., and Terwiesch, C. 1999. The economics of yield-driven processes. ***Journal of Operations Management*** 18(1), 41-59.

Bourgeois, III, L. J. and Eisenhardt, K. M. 1988. Strategic Decision Processes in High-Velocity Environments: Four Cases in the Microcomputer Industry, ***Management Science*** 34(7), 816-835.

Brown, S. L. and Eisenhardt, K. M. 1997. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations, ***Administrative Science Quarterly*** 42(1), 1-34.

Chase, R. B., and Acquilano, N. J. 1981. ***Production and Operations Management***, third ed. Richard D. Irwin, Inc., Homewood, IL.

Cunningham, J. 1990. The use and evaluation of yield models in integrated circuit manufacturing. ***IEEE Transactions on Semiconductor Manufacturing*** 3(2), 60-71.

Dudley, L. 1972. Learning and productivity change in metal products. ***The American Economic Review*** 62(4), 662-669.

Dutton, J. M., and Thomas, A. 1984. Treating progress functions as a managerial opportunity. ***Academy of Management Review*** 9(2), 235-247.

Eisenhardt, K. M. 1989. Building theories from case study research. ***Academy of Management Review*** 16, 620-627.

Fellner, W. 1969. Specific interpretations of leaning by doing. ***Journal of Economic Theory*** 1, 119-140.

Gersick, C. J. G. 1988. Time and transition in work teams: Toward a new model of group development. ***Academy of Management Journal*** 31(1), 9-41.

Gersick, C. J. G. 1991. Revolutionary Change Theories: A multilevel Exploration of the Punctuated Equilibrium Paradigm. ***Academy of Management Review*** 16(1), 10-36.

Hatch, N. W., and Mowery, D. C. 1998. Process innovation and learning by doing in semiconductor manufacturing. ***Management Science*** 44(11), 1461-1477.

Hayes, R. H. and Clark, K. B. 1985. Exploring the source of productivity: Differences at the factory level, in: Clark, K. B., Hayes, R. H. and Lorentz, C. (eds.), ***The Uneasy Alliance: Managing the Productivity-Technology Dilemma***. Boston, MA, Harvard Business School Press, pp. 151-188.

Hirsch, W. 1952. Manufacturing progress functions. ***Review of Economics and Statistics*** 34, 143-155.

Iansiti, M. 1998. ***Technology Integration: Making Critical Choices in a Dynamic World***, Harvard Business School Press, Cambridge, MA.

Lapré, M. A., Mukherjee, A. S., and Van Wassenhove, L. N. 2000. Behind the learning curve: linking learning activities to waste reduction, ***Management Science*** 46(5), 597-611.

Leachman, R. C. 1996. Competitive manufacturing survey: Third report on the results of the main phase. ***UC Berkeley Report CSM-31***, University of California, Berkeley, CA.

Leachman, R. C., and. Hodges, D. A. 1996. Benchmarking Semiconductor Manufacturing. ***IEEE Transactions on Semiconductor Manufacturing*** 9(2), 1158-1169.

Lundvall, D. M., and Juran, J. M. 1974. Quality Costs. In: Juran, J. M., Gryna, F. M., and Bingham, R. S., Jr. (eds.), ***Quality Control Handbook***, third ed., McGraw-Hill, Inc., New York, NY.

Moore, G. 1975. Progress in digital integrated circuits. ***IEDM Technical Digest*** 1975, 11.

Mukherjee, A. S., Lapré, M. A., and Van Wassenhove, L. N. 1998. Knowledge-driven quality improvement. ***Management Science*** 44 (11), 35-44.

Pisano, G. P. 1994. Knowledge, integration and the locus of learning: An empirical analysis of process development. ***Strategic Management Journal*** 15, 85-100.

Pisano, G. P. 1996. Learning before doing in the development of new process technology. ***Research Policy*** 25, 1097-1119.

Rapping, L. 1965. Learning and the World War II production functions. ***Review of Economics and Statistics*** 48, 98-112.

Searle, A. D. and Gody, C. S. 1945. Productivity increases in selected wartime shipbuilding processes. ***Monthly Labor Review*** 60, 1132-1147.

Stapper, C. and Rosner, R. 1995. Integrated circuit yield management and yield analysis: Development and implementation. ***IEEE Transactions of Semiconductor Manufacturing*** 8(2), 95-101.

Tang, C. S. 1991. Designing an optimal production system with inspection. ***European Journal of Operational Research*** 52, 45-54.

Terwiesch, C., and Bohn, R. E. 2001. Learning and process improvement during production ramp-up. ***International Journal of Production Economics*** 70(1), 1-19.

Thomke, S. H. 1998. Managing experimentation in the design of new products. ***Management Science*** 44(6), 743-762.

Tushman, M. L., and Anderson, P. 1986. Technological discontinuities and organizational environments. ***Administrative Science Quarterly*** 31, 439-465.

Tushman, M. L., and Romanelli, E. 1985. Organizational evolution: A metamorphosis model of convergence and reorientation, in: Cummings, L. L., and Staw, B.M. (eds.), ***Research in Organizational Behavior*** 7, JAI Press, Greenwich, CT, pp. 171-222.

Utterback, J. M., 1994. ***Mastering the Dynamics of Innovation***, Harvard Business School Press, Cambridge, MA.

Von Hippel, E., and Tyre, M. 1995. How leaning by doing is done: Problem identification in novel process equipment. ***Research Policy*** 24, 1-12.

Weber, C. M. 2003. ***Rapid Learning in High Velocity Environments***. Doctoral Dissertation, MIT Sloan School of Management, Cambridge, MA, pp. 220-282.

Weber, C. M. 2004. Yield learning and the sources of profitability in semiconductor manufacturing and process development. ***IEEE Transactions on Semiconductor Manufacturing*** 17(4), 590-596.

Weber, C., Moslehi, B., and Dutta M. 1995. An integrated framework for yield management and defect/fault reduction. ***IEEE Transactions on Semiconductor Manufacturing*** 8(2), 110-120.

Wein, L. M. 1992. Random yield, rework and scrap in a multistage batch manufacturing environment. ***Operations Research*** 40, 551-563.

Wright, T. P. 1936. Factors affecting the cost of airplanes. ***Journal of Aeronautical Science*** 3, 122-128.

Yin, R. K. 1994. ***Case Study Research***, Sage Publishing, Newbury Park, CA.

Zangwill, W. I., and Kantor, P. B. 1998. Toward a theory of continuous improvement. ***Management Science*** 44(7), 910-920.

APPENDIX A: QUESTIONS FOR THE RESPONDENTS

Questions Regarding Special Events: Did you witness any of the following special events regarding the VLSI circuit process that pertained to the particular case that you are recounting: the beginning of physical experimentation; the first physical experiment that was performed on a VLSI circuit product prototype ($t=t_{PD}$); the beginning of commercial startup ($t=0$); the cessation of the increase in the line throughput rate? If yes, at what calendar date did the event occur? Please estimate the numerical value of the following performance metrics at that point in time: batch fault density '$F_B(t)$'; batch yield '$Y_B(t)$'; line yield '$Y_L(t)$'; line throughput rate (in wafer starts per month) and product output rate (in chips per month). (Batch fault density was converted into batch fault yield rate by substituting observed values for batch fault density into the quantity '$1-F_B(t)/F_{B\_max}$', where $F_B(t)$ represents batch fault density at the point in time at which the special event occurs. Given that $F_B(t)$ cannot be measured at $t<t_{PD}$, the highest level of batch fault density that could be measured acted as a proxy for F_{B_max}.)

Questions Regarding Maximum Performance: What was the maximum value for line throughput rate (in wafers per month) and product output rate (in chips per month) that your company could possibly achieve in the venture that you have described in your case? (Line throughput rates and product output rates were converted into yield rates by dividing observed values by maximum values.)

Questions Regarding VLSI Circuit Process Technology: How many instances of a VLSI circuit product (prototype) design did the batches (wafers) in your case contain? What were the minimum

features of the VLSI circuit products in your case? What were their dimensions? How many clock cycles per second (key performance indicator for VLSI technology) does a leading-edge device realized by the process technology in your case achieve? If your case transpires in the early phases of the VLSI process lifecycle, what are the minimum feature sizes and maximum device clock speed that the VLSI circuit process under observation is intended to realize?

The Timing of the Case: What are the key calendar dates that pertain to your case? When did learning and problem-solving activity pertaining to your case begin? When was it completed?

Reciting the Case: Please recite the case as you believe it transpired. Please answer the following questions in the process. How was the problem in your case detected initially? How was it localized to a specific technology? How was the root cause of the problem identified? What was the fix? Was the fix implemented? Why or why not? How? How was the fix confirmed?

The Case – Performance variables (Numerical Values): Please estimate the numerical value of the following performance metrics at the time your case transpires: batch fault density '$F_B(t)$'; batch yield '$Y_B(t)$'; line yield '$Y_L(t)$'; line throughput rate (given in wafer starts per month) and product output rate (given in chips per month). How did the actions taken in your case affect these values?

The Case – Performance Variables (Long-Term Tendencies, Shapes of Curves): Please indicate which of the following best describe the long term tendencies of batch fault density, batch yield, line yield, line throughput rate and product output rate at the time your case took place: a) decreasing at an increasing rate; b) decreasing at a decreasing rate; c) flat – little change; d) increasing at a decreasing rate; of e) increasing at an increasing rate. How did the actions taken in your case affect these rates?

The Cost of Learning: Approximately, how many instances of each equipment type were present in the fab the time the case you are reciting transpired? What was the cost of a silicon wafer at the time of your case? What was the cost of a fully processed wafer? What are the specific costs associated

with increasing batch yield, line yield and the line throughput rate? Please rank these performance parameters with respect to cost of improvement.

Optimal Performance Levels: What was the optimal level of performance for batch fault density, batch yield, line yield, line throughput rate (in wafer starts per month) and product output rate (in chips per month) that the organization in the case was trying to achieve?

Financial Performance: Approximately what was the (expected) unit price of VLSI circuit product of the type your venture would produce? At the time the case transpired, did you believe the venture was (going to be) profitable? Did your expectations materialize? If your case occurred after the release of the first major product do expect the unit price of your product to increase, stay the same or decrease over time?

Temporal Calibration: When (calendar date) do you believe the organization in your case had to achieve optimal performance levels for maximum profitability to occur? At the time the case transpired, did you believe that you were ahead of Moore's Law, in synch with Moore's Law or behind Moore's Law? By how much time?

CAPTIONS

Table 1: Yield Rate Data

Figure 1: Empirically based model the lifecycle of a VLSI circuit manufacturing process. CI stands for continuous improvement.

FOOTNOTE TABLE

[1] For a detailed description of Moore's Law, please see Moore (1975).
[2] Semiconductor Industry Association (1994, 1997). *The National Technology Roadmap for Semiconductors*. Semiconductor Industry Association (1999, 2001, 2003). *The International Technology Roadmap for Semiconductors*.
[3] Batch yield is colloquially known as 'chip yield', 'die yield' or 'die-sort yield' in the semiconductor industry (Bohn, 1995, p. 33).
[4] In the semiconductor industry, line yield is frequently referred to as 'survival yield' because it represents the fraction of batches that survive the manufacturing line or the fraction of batches that is not wasted in the line. The line throughput rate is colloquially known as 'wafer starts' because it roughly corresponds to the number of wafers that entered the fab a few weeks earlier.
[5] The relationship between batch yield and batch fault density varies from process to process and from product to product. Most process engineering subsystems tend to characterize this relationship empirically as a highly nonlinear function known as a yield model (e.g. Cunningham, 1990; Stapper and Rosner, 1995).

FIGURE 1

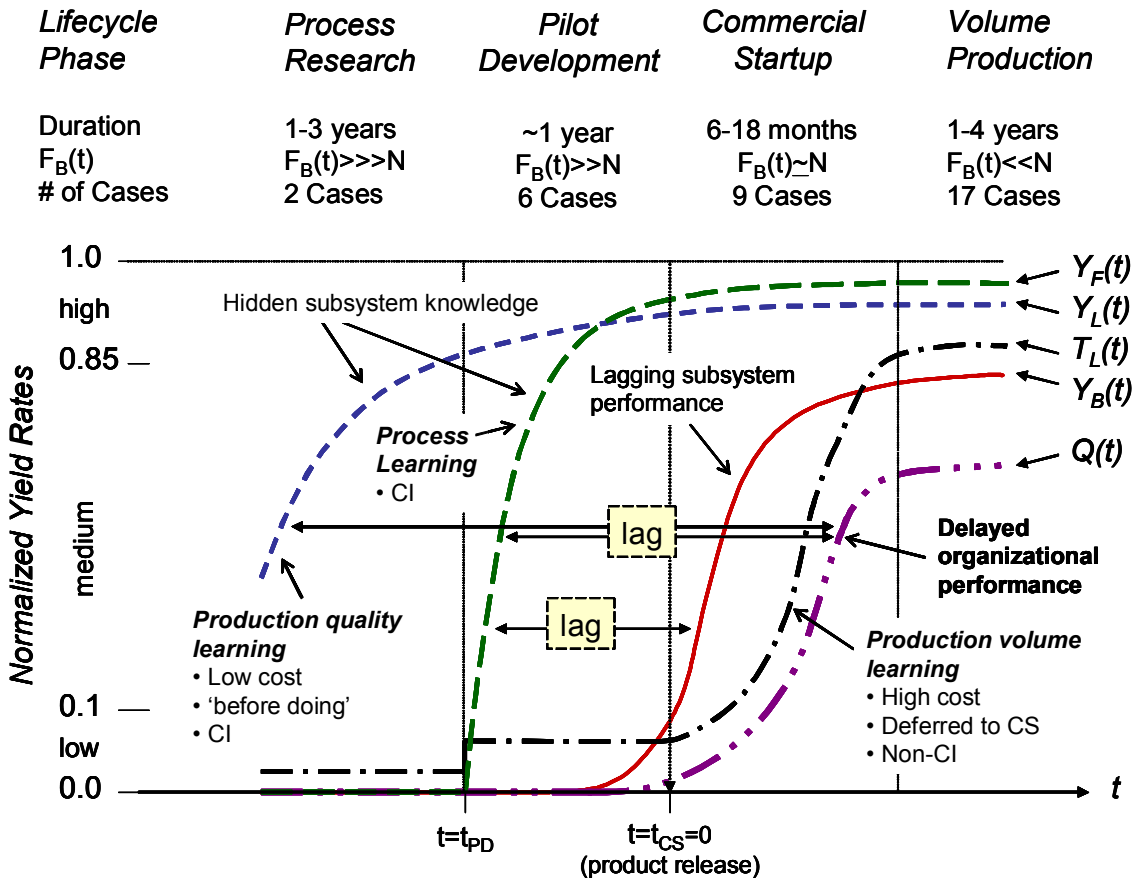
Lifecycle Phase
Process Research
Pilot Development
Commercial Startup
Volume Production
Duration
$F_B(t)$
# of Cases
1-3 years
$F_B(t)>>>N$
2 Cases
~1 year
$F_B(t)>>N$
6 Cases
6-18 months
$F_B(t)\sim N$
9 Cases
1-4 years
$F_B(t)<<N$
17 Cases
1.0
high
0.85
medium
0.1
low
0.0
Normalized Yield Rates
Hidden subsystem knowledge
Process Learning
• CI
Lagging subsystem performance
lag
lag
Delayed organizational performance
Production quality learning
• Low cost
• 'before doing'
• CI
Production volume learning
• High cost
• Deferred to CS
• Non-CI
$Y_F(t)$
$Y_L(t)$
$T_L(t)$
$Y_B(t)$
$Q(t)$
t
$t=t_{PD}$
$t=t_{CS}=0$
(product release)

TABLE 1

Phase	Timing	$Y_F(t)$ Level, Trend and Shape	$Y_F(t)$ Frequency of Observation	$Y_B(t)$ Level, Trend and Shape	$Y_B(t)$ Frequency of Observation	$Y_L(t)$ Level, Trend andShape	$Y_L(t)$ Frequency of Observation	$T_L(t)$ Level, Trend and Shape	$T_L(t)$ Frequency of Observation	$Q(t)$ Level, Trend and Shape	$Q(t)$ Frequency of Observation
Process Research (PR)	-4 yrs.< t <-1.0 yrs.	Level: low	(2/2)	Level: low	(2/2)	Level: medium	(2/2)	Level: low	(2/2)	Level: low	(2/2)
		Trend: ??	(2/2)	Trend: flat	(2/2)	Trend: rising	(2/2)	Trend: flat	(2/2)	Trend: flat	(2/2)
		Shape: ??	(2/2)	Shape: flat	(2/2)	Shape: concave	(2/2)	Shape: flat	(2/2)	Shape: flat	(2/2)
Pilot Development (PD)	-1.0 yrs.< t <0.0 yrs.	Level: high	(6/6)	Level: low	(6/6)	Level: high	(6/6)	Level: low	(6/6)	Level: low	(6/6)
		Trend: rising	(6/6)	Trend: flat	(6/6)	Trend: rising	(6/6)	Trend: flat	(6/6)	Trend: flat	(6/6)
		Shape: concave	(6/6)	Shape: flat	(6/6)	Shape: concave	(6/6)	Shape: flat	(6/6)	Shape: flat	(6/6)
Product Launch Date	t=0.0 yrs.	Level: high	(9/9)	Level: medium	(9/9)	Level: high	(9/9)	$T_L(0)<Y_B(0)$	(9/9)	Level: low	(9/9)
								Level: low	(9/9)		
Commercial Scale-up (CS)	0.0 yrs.< t <1.0 yrs.	Level: high	(9/9)	Level: medium	(9/9)	Level: high	(9/9)	Level: medium	(9/9)	Level: low	(9/9)
		Trend: rising	(9/9)	Trend: rising	(9/9)	Trend: rising	(9/9)	Trend: rising	(9/9)	Trend: rising	(9/9)
		Shape: concave	(9/9)	Shape: S	(9/9)	Shape: concave	(9/9)	Shape: conv. or S	(9/9)	Shape: S	(9/9)
Volume Production (VP)	1.5 yrs.< t	Level: high	(17/17)	Level: medium	(17/17)	Level: medium	(17/17)	$Y_B(t)<T_L(t)$	(12/17)	$Q<Y_B,Y_B,T_L$	(17/17)
								Level: high	(1/17)		
								Level: medium	(16/17)	Level: medium	(17/17)
		Trend: flat	(1/15)	Trend: flat	(2/15)	Trend: flat	(12/15)	Trend: flat	(4/15)	Trend: flat	(0/15)
		Trend: rising	(14/15)	Trend: rising	(13/15)	Trend: rising	(3/15)	Trend: rising	(9/15)	Trend: rising	(15/15)
		Shape: flat	(1/15)	Shape: flat	(2/15)	Shape: flat	(12/15)	Shape: flat	(4/15)	Shape: flat	(0/15)
		Shape: concave	(14/15)	Shape: concave	(13/15)	Shape: concave	(3/15)	Shape: concave	(9/15)	Shape: concave	(15/15)