1

GROWING TO LEARN

As the 2006 war between Israel and Lebanon staggered toward a U.N.-brokered cease-fire, the Iraq war spun into sectarian conflict, and oil prices floated over seventy-five dollars a barrel, BP (once British Petroleum) revealed it was shutting down operations in Prudhoe Bay, Alaska—about 8 percent of U.S. petroleum production capacity. The Organization of Petroleum Exporting Countries eagerly announced it would make up any shortfall.

BP admitted it had not checked some of its North Slope transit pipelines since it ran a "smart pig" corrosion-sensing device through them in 1992. "With hindsight, that's clearly a gap in our program," acknowledged BP corrosion management and chemicals program team leader Bill Hedges.[1] Ignoring the reliability of the pipes carrying 8 percent of U.S. petroleum production is quite a gap. By the same token, you might say forgetting about the Russian winter was a gap in Napoleon's program.

Of course, bad things happen to good people. But BP's run of bad news in 2005 and 2006 started to look systematic. The shutdown occurred five months after an earlier leak from one of the company's Alaska pipelines went undetected and dumped 200,000 gallons of oil on the North Slope.[2] And in June 2006, the Commodity Futures Trading Commission accused BP of rigging the propane market in February 2004.[3] After substantial forward propane purchases—contracts requiring other parties to deliver the fuel at a fixed price in the future—BP allegedly withheld its own refined propane from the market to create an artificial scarcity. That any trader would risk BP's reputation for the caper's paltry $20 million reported yield is astonishing.

The context of these missteps is the real issue, however. Two of them followed an explosion at the firm's Texas City refinery

that killed fifteen and injured one hundred in March 2005. The Occupational Safety and Health Organization fined BP $21 million for "egregious, willful violations of safety standards."[4] The U.S. Chemical Safety and Hazard Investigation Board preliminarily determined in October 2006 that faulty equipment and staff reductions were responsible for the tragedy.[5]

If ever a tragedy put a company on notice to lock down its operations, you would think the worst U.S. industrial accident in a decade would do it. Why didn't BP's extraction businesses respond to the terrible experience of its refining businesses?

This chapter argues that learning from experience requires performance strategies explicit enough to be testable. A strategy is testable if it spells out goals and assumptions about how to achieve them that could conceivably prove wrong. You might ask why learning from experience has anything to do with the strategy you're pursuing. The chapter claims that what you learn depends on what you expect, which is what a testable performance strategy lays out. For example, BP's failure to learn from repeated mistakes in its American operations followed an initiative that devolved strategy setting to operating units. The first section of the chapter argues that the lack of testable clarity in BP's strategies impaired its ability to learn and adapt.

This is important because BP is hardly alone in devolving strategy to independent business units. In fact, the failure to lay out testable strategies is widespread. It afflicts any organization that pursues stagnant or chaotic strategies, including both those with slow-to-mature "hockey stick" profit goals that are impossible to test in the short run (think of profit projections that run flat for three years and suddenly jump up in the fourth) and those that derive their goals from investor requirements without specifying any operating assumptions. The second section of the chapter contends that strategy clear enough to test is extraordinarily rare. That rarity explains the growing difficulty of determining what matters in our voluminous performance reports.

The third section of the chapter shows how setting testable strategies helps firms learn from experience. Alcoa and GE both manage themselves by laying out clear, changing performance goals and strategic assumptions that they regularly put to the test. The section draws links between this practice, Rand Corporation's assumption-based planning, and Rita McGrath's and Ian MacMillan's discovery-driven planning.

Chapter Two continues the argument by proposing a method called *eight-line strategies* for distilling the strategy relevant to a manager at any level of a firm to a short list of testable assumptions. Setting out a testable strategy is the first step in what I call the *guess-test* system of performance management. Starting with guesses in the form of testable strategies, this system lets you pick out facts that can improve decisions from piles of conflicting performance data. In essence, it uses performance results as a means of sharpening strategy continuously and not just of tracking execution.

Accountability at the Expense of Firmwide Learning at BP

BP is a good example of this chapter's theme that lack of testable clarity in a company's strategy can impair its ability to learn and adapt. The run of accidents and scandals in BP's American operations constituted a firmwide learning failure, and the setbacks followed an initiative that devolved strategy setting to operating units.

The learning failure is well documented. "It is a very significant finding that BP does not effectively investigate incidents throughout the corporation," warns U.S. Chemical Safety Board spokesman Daniel Horowitz. "If you're not learning from near misses, you're not in a position to prevent major disasters like the one in Texas City."[6] What impeded learning at BP?

The company underwent an organizational change starting in 1995 large enough to account for its uncharacteristic swerve into a series of apparently systematic mistakes a decade later. As John Roberts tells it in *The Modern Firm*, the company reduced staff

in its corporate headquarters by 80 percent under CEO Robert Horton's "Project 1990."[7] At the same time, the upstream business cluster BP Exploration devolved strategy to its business units in a disaggregated model called "asset federation." After leading BP Exploration's transformation, John Browne became Group CEO in 1995.[8]

That transformation transferred responsibility for performance down to the managers of forty individual sites or fields from the regional operating companies that had previously run them. Browne eliminated the regional operating companies, cut back the management of BP Exploration to an executive committee of himself and two others, and transferred key technical staff to the fields.[9]

Managers of the fields and other assets started to sign performance contracts with Browne's executive committee. The contracts held them responsible for production volumes and expenditures but left them "empowered to figure out how to achieve their promised performance. They could decide on outsourcing and choose suppliers, do their own hiring, and determine where and how to drill."[10]

Institutional memory and learning are at risk in this sort of disaggregated organizational design. In fact, Roberts makes the point that any organizational design trades off initiative—maximized under Browne's plan—and cooperation—for which, like institutional learning, Browne's system needed to find other solutions. "But the changes," he writes of Browne's design, "also created a great need for the business units to cooperate in sharing best practice and in supporting one another in solving technical and commercial problems—activities that were previously handled by the center, but that it now lacked the resources to undertake."[11]

To fill this need, BP Exploration created a system of "peer assists" by which a unit could call on any other unit in its group of peers (defined as those with assets at a similar life stage) for help with commercial and technological problems. The delegation of responsibility for performance to individual asset units, supported

by the system of peer assists, proved so successful in streamlining operations and cutting costs that Browne implemented it across the entire company when he became Group CEO.[12]

The groups of similar asset-based business units that facilitated peer assists also facilitated "peer challenges." In a peer challenge, a business unit questions the performance goals another unit has negotiated with the executive committee.[13] A peer group's collective responsibility for meeting each member unit's goals and for allocating resources among the members is supposed to motivate thoughtful challenges.

Neither peer assists nor challenges, however, were enough to prevent calamity. In investigating the Texas City explosion, a panel headed by James Baker found that workers at a sister refinery in Whiting, Indiana, reported that "preventive maintenance was seldom practiced, the refinery had a 'run until it breaks' mentality, and the workforce had a great deal of experience running equipment with 'Band-Aids.'"[14] In other words, the Whiting plant ran under conditions like those at Texas City. Peer assists failed the two refineries in that they did not usefully pool their experience. And peer challenges failed them in that lax safety and upkeep practices prevailed despite worker concerns.

BP's system of independent business unit strategy setting combined with peer challenges failed to produce uniformly explicit, testable strategies when it came to performance sustainability. Insofar as peer challenges brought objective scrutiny to bear on performance goals, they were a brilliant innovation. They nevertheless have two disadvantages from the perspective of organizational learning. Their resource implications introduce an element of competition among peer group units that can undermine peer-to-peer learning. At the same time, patterns of cooperative laxity can emerge that weaken the search for continuous improvements.

These two disadvantages may seem contradictory, but both are often on display in the behavior of boards of directors. For example, CEOs who sit on boards compete with one another informally

for the honors that go with top pay from their own companies. The record of these boards, however, as rehearsed in literally hundreds of studies, reports, and articles, has been one of at least equal informal collaboration to relax CEO pay package discipline.[15] So competition and collaboration can go hand in hand.

Board behavior shows why BP should not have relied on peer challenges to drive its field managers to learn from one another's experience. Small peer groups don't consistently raise the stringency of one another's performance goals because cooperative leniency keeps breaking out. Even without explicit cooperation, you can always find out by trial and error which of your peers tend to reciprocate when you give them the benefit of doubt on a lenient goal. Over time, lenience becomes the norm.

BP's system of independent business unit strategy setting combined with peer challenges also failed to produce uniformly explicit, testable strategies when it came to risk management. The manager of a business unit will question peer levels of operating risks less closely than a corporate center risk manager will, if only because individual units experience mishaps less frequently than peer groups as a whole. For example, if a group of ten similar operating units experiences one major mishap every ten years and one early warning sign every year, then each operating unit will experience a major mishap only once a century on average and an early warning sign only once a decade. A central planner for any group of units will try to aggregate their risk experience and expose her individual managers to it. But the delegation of strategy review to operating units may reduce managers' sensitivity to the group's catastrophic risks because individual units can run so long without seeing any evidence of them at all.

Something may seem jarring about the proposition that BP's delegation of strategy setting to operating units impaired their ability to learn from experience and adapt to changing conditions. Surely Browne had adaptation foremost in mind when he gave the managers of individual operating assets more responsibility for how they met their goals. You can't blame a corporate learning failure

on an organizational innovation that moves decisions closer to the people with the information to make them.

The answer to the conundrum may be that Browne's was an incomplete revolution. He may well have been right that BP's operating units—or at least its upstream units in exploration and production—could operate independently. If so, they may prove most valuable under independent ownership. The parts of BP Exploration may someday prove more valuable than the whole.

The reason stems from the very weakness of running completely independent extraction operations. A stand-alone operation's lack of resilience after major mishaps heightens its managers' sensitivity to any kind of latent risk. Independent operating units have to learn from a narrow base of business experience, but they may compensate with a heightened appetite for relevant outside information. Operating units that can fall back on the financial strength of a firm like BP have access to more useful peer information on risks but less incentive to scrutinize it.

We may never know whether John Browne was right. He was committed to his management revolution at BP but did not take its Texas City tragedy and other setbacks lightly. He is the type of systematic thinker who might even have considered the radical step of divestitures. But he resigned on January 12, 2007, earlier than he had planned.

The idea that businesses that don't benefit from strategy coordination should be independent raises a broader theme explored in this book. According to classical industrial economics, the cost of doing business defines the natural scope of a company. The example of BP suggests this may be wrong. The mutual relevance of the various activities of a business may define its natural scope instead.

The classic idea that transaction costs determine the natural scope of a firm follows from an argument of British Nobel prize winner and University of Chicago professor Ronald Coase. He claimed that a market with clear property rights will produce the same efficient outcome no matter who owns the property unless there are high transaction costs in the market.[16]

For example, it should be just as efficient for two business activities to trade supplies within a single firm as it is for a similar pair of activities in separate firms to trade supplies—so long as transaction costs are low. When transaction costs are high, however, it may be more efficient to keep the two business activities within a single firm. In this case, the firm may be able to cut the cost of the activities by mandating the terms of their interactions and exchanges.

This section provides a reason for thinking that high transaction costs are a special case of situations where companies can coordinate activities more efficiently than markets can. It suggests companies may coordinate activities more efficiently than markets when the results of one activity are relevant to the plans of another, regardless of transaction costs. The case where one business unit's experience applies strongly to the management of another unit's risks is an example. More than internal transaction costs, the mutual relevance of divisions' results may shape the modern firm.

The Rarity of Strategic Clarity

It may seem surprising that a company like BP could have failed to lay out group strategies clearly enough to tell from experience which parts worked and which parts did not. But strategies clear enough to test are extraordinarily rare. Clarity is the victim any time an organization pursues *stagnant* or *chaotic* strategies. Stagnant strategies include the hockey-stick projections in naively optimistic venture capital presentations because they are impossible to test in the short run. Typical examples of chaotic strategies are goals derived from investor requirements that don't specify operating assumptions. The rarity of strategic clarity helps explain why it's so hard to tell what matters in our burgeoning performance reports.

The failure to lay out testable strategies is widespread. A strategy is testable if it spells out goals and assumptions about how to achieve them that could conceivably prove wrong. A strategy

could prove wrong if the performance results of an organization pursuing it are capable at least in principle of revealing its flaws. It's hard to imagine what, for instance, could count against the vague strategy to increase sales by hiring more salespeople. But results could well refute the clearer strategy of increasing sales 20 percent by hiring 10 percent more sales staff. What makes such a strategy clear—or at least clear enough to test—is the possibility of showing it's not quite right.

In some ways, this kind of clarity is a low bar. Plenty of strategies are clear in this sense and yet still not advisable. Even so, clear strategies are few and far between. For example, both chaotic and stagnant strategies fail to clear the clarity bar. And yet most strategies, as we'll see, fall into one of these two categories.

Stagnant Strategies

Stagnant strategies are unclear the way habits are unclear. Habits get things done, but it's usually not clear how well. We may as a matter of habit get through all seventy e-mails we receive every day, for example, but it's not clear we do it efficiently and it's not clear which of our e-mail habits are particularly effective. We set no expectations for getting through those e-mails, so we don't look for better ways to process them.

Strategies often become stagnant because their goals are stagnant. Such strategies may suffice to meet their flat-line goals. But since it's not clear by how wide a margin they should meet those goals, just passing doesn't really tell you much about them. A great example is the notorious hockey-stick graph endemic to project plans and new business proposals that forecasts a great long-term gain after a string of break-even years. It's a warning sign of a strategy free to stagnate until the curve finally turns up. Hockey-stick projections are supposed to clinch business cases for new initiatives, but they really say, "Trust me." They offer no way to test a strategy to see whether it's working until the end of the forecast period—usually too late.

Readers who, like me, may have used a hockey-stick projection or two in a moment of weakness in the past will be tempted to protest that these projections at least allow long-term thinking and planning. Some ideas take a long time to bear fruit, goes the argument. But since all of our financial reporting focuses on quarterly results, it militates against the very long-term plans we may most need. The least we can do is to allow our best new initiatives enough time to gestate.

The problem with this defense of long-term goals is that it affords no way to tell exactly which one really is your best new initiative. Calling an initiative a best bet is, after all, a theory. And theories are worthless if there's no way to test them. To see whether one of several possible initiatives really is most promising, you need to devise some near-term indicator that it's working. Such an indicator lets you test and adapt the strategy behind the initiative. It keeps the strategy from stagnating until the time that the upswing in a hockey-stick projection eventually puts it to the test.

If lax and static goals really do allow strategies to become stagnant, there may be more to the pursuit of aggressive growth goals than wishful thinking and willful ignorance of the law of averages. To meet high-growth goals, according to this view, means more than to win at the expense of competitors. It means that the organization pursuing those goals is continually testing its strategy and sensing change. Instead of saying organizations must learn in order to grow—following Nelson and Winter, who called growth a process of pure selection when they launched evolutionary economics in 1982—it may be truer to say that organizations must grow to learn.[17]

Chaotic Strategies

Chaotic strategies, in contrast, are unclear because they're too fragmented or volatile to attribute results to any one aspect of them. Several years ago, for example, I asked my staff in the Corporate Executive Board's subscription-based research program for

treasurers to do the following in a single fiscal year: develop a new online bank credit tool; segment the program's market according to capital structure; analyze our marketing success with each segment; emphasize the most enthusiastic segments in scheduling marketing calls; make quarterly checkup calls on each member; analyze treasurers' business processes; and reorganize our content offerings in terms of the treasury processes that each offering supported.

This, it's embarrassing to realize, is a good example of a chaotic strategy. It contains seven new initiatives, reflecting seven ways the program's prior strategy might have been wrong about how the business should really work. Moreover, the initiatives struck out in different directions. The multiplicity and incoherence of the initiatives made it hard to tell at the end of the year which ones to keep and which to adjust. Would slow sales, for example, mean that the bank credit tool underperformed or was hard to explain; that the market segmentation missed the market's real fault lines; that the segment analysis drew mistaken conclusions; that call schedulers had placed prospects in the wrong segments; that our treasury business process analysis was wrong; or that our content reorganization was opaque? Would disappointing current subscriber renewals mean any of these or, alternatively, that checkup call execution was poor?

This kind of chaotic strategy is common because it's a natural response to a tough goal or an urgent need to close a challenging performance gap. In these situations, there is a strong temptation to make lots of changes and "let a thousand flowers bloom." The advantage of such an energetic response is that it increases the likelihood that at least one of the changes you make will actually improve results. The disadvantage, characteristic of all chaotic strategies or chaotic strategic shifts, is that you won't know what worked.

Chaotic strategies also include those derived by reasoning backward from investor requirements—without specifying the operating assumptions behind the reasoning. Many companies derive their budgets this way, determining how much profit each unit must contribute to meet a promise to the market. But unless

such a strategy lays out how to achieve its goals, there's no way to tell what went wrong if it misses them.

Even planning systems designed to generate testable strategies fall into this trap. For example, the reverse income statements in Rita McGrath's and Ian MacMillan's discovery-driven planning, described in the next section, work back from a profit target to generate the line items of an income statement that would hit it.[18] These line items might include segment revenue, direct costs, and indirect costs. They're supposed to suggest the key assumptions of a testable strategy, but they don't tell you how to hit the strategy's target. There might be hundreds of ways to do so—all consistent with the line item requirements.

More broadly, finance executives are increasingly adopting versions of aspiration-based planning to focus their efforts to improve operating decisions. The idea is to ask operating managers where they hope to take their divisions in three or five years—their aspirations. The next step is to assess gaps between current capabilities and what those aspirations require. Some finance teams are even gearing up to fill those gaps.

Aspiration-based planning lends itself to capability analysis, but it doesn't necessarily generate testable strategies. It may elicit admirably clear and ambitious goals. And it may not be hard to derive requirements from those goals. But even if you meet every requirement you can imagine, you may not hit a goal. Requirements are much broader than specific prescriptions for achieving a goal. That's why requirements are relatively easy to list, while testable strategies can be fiendishly hard to devise. For example, it's obvious that you have to schedule at least ten visits with prospective customers to have a hope of selling ten insurance policies. That's a requirement. But scheduling those visits hardly guarantees the sales. A sales strategy requires more knowledge, such as research into your prospects' needs. We'll return to the sharp distinction between requirements and strategies in Chapter Three.

Strategies that specify a final goal but not how to achieve it resemble those that strike out in too many directions at once. Both

kinds of strategy may have stringent goals, but neither helps you find out from experience what achieves them.

Whether an organization's strategy is chaotic because it specifies too many new activities or none, its marching orders boil down to scrambling. Often that's all you can do when you have a target you must meet in any way you can. But it's hard to learn what works in a chaotic scramble to hit a target. And if you face a competitor who, instead of scrambling, can test alternative means for achieving his goals, you'll learn more slowly.

This may be why so many venture capitalists bring in managers with clear, simple plans to turn around start-ups that founders have made frantic in a never-ending scramble to meet investor targets. There's nothing wrong with the joyous frenzy of a start-up to hit a target. But there's something wrong with a start-up that can't learn.

This quick sketch of stagnant and chaotic strategies and the kind of clarity they both lack shows why clear strategies are rare. It's because they're hard to specify. They need to set aggressive goals to avoid stagnation. And they need to propose a way of achieving them that experience can help refine over time. Clear strategies are hard because they really force you to find out, if not to know, what you're talking about.

How Setting Testable Strategies Helps Alcoa and GE Learn from Experience

However rare, clear and testable strategies appear to be a requirement for firms to learn from experience. For example, both Alcoa and GE manage themselves by laying out clear, changing performance goals and strategic assumptions that they regularly put to the test. This section draws links between this practice, Rand Corporation's assumption-based planning, and Rita McGrath's and Ian MacMillan's discovery-driven planning.

Before he became secretary of the U.S. Treasury, Paul O'Neill served as CEO of Alcoa from 1987 until 2000. Alcoa revenue grew

on average 23 percent per year from $1.5 to $23 billion over the period.[19] The record is especially remarkable in light of the fact that inflation was low at that time, no major, discontinuous new uses of aluminum emerged, and the company nevertheless grew largely organically. More remarkable still, O'Neill compiled the record over a period that saw fairly sharp commodity market downturns in 1991, 1993, and 1998.[20] Alcoa has proven a deft giant in a sector whose risk management needs have become paralyzingly complex.

After seven years at Toyota and several as a consultant, John Marushin became director of an internal consulting group that helped operating units implement the Alcoa Business System (ABS). He gives an anecdote illustrating the high adaptability of Alcoa operations. It also illustrates how Alcoa's adaptability depends on an environment that lets managers speculate about what matters most and test their ideas.[21]

The difficulty of meeting surges in customer needs through new ingot production led a smelting unit to hold thirty thousand tons of aluminum ingots in inventory. But the cost of financing that inventory made the smelter's ingots uncompetitive. Marushin's ABS team started by focusing the unit on what was critical for customer needs: on-time production of varying numbers of ingots, in this case. Then it tried to rank possible obstacles. Finally, it laid out detailed cause-and-effect scenarios for overcoming the most likely obstacles in enough detail to test them. Those scenarios were testable strategies.

Speeding up the changeover time for casting equipment in the smelter's ingot pit from six hours to twenty minutes turned out to have a big impact on production schedules. It let the operation produce all five of its products every day rather than just one type each week. And that let the unit meet changing customer needs without much inventory. "The process improvement was changeover in the pit," explains Marushin. "The system improvement was . . . the cash cost of the business."[22] He might have added plant flexibility, happier customers, and less pressure for smelter consolidation.

You might wonder whether the firm tried to match the adaptability of ABS at the level of its management system by systematically articulating testable corporate and divisional strategies. That turns out to be a good description of how Alcoa worked.

When Paul O'Neill became CEO, for example, he drafted the strategic vision that Alcoa would be "the best aluminum company in the world." And it would achieve this by being "a growing worldwide company dedicated to excellence through quality."[23] Instead of a bromide, the statement put forward two ideas that experience might have proved wrong. It asserted that quality was a route to excellence, which might have come as a surprise to many in a basic metals sector beset by declining costs in a globalizing market. And it asserted that growth was necessary to be best in class, suggesting growth was a necessary ingredient of not just size but quality. Results might have proved this vision wrong, and yet growth made the firm more agile than its competitors and quality reduced cost by reducing defects and waste.

What applied to the apex of the company applied to its operating units: every Alcoa operating manager articulated a clear aspiration for what he or she ran, together with a sense of how to achieve it. An example from a mine manager was to become "the most customer-oriented, quality-focused mine in Australia."[24] Marushin's consulting team asked each internal client what would have to happen for the unit to grow and gain market share according to its marketing plan and aspirations.[25]

Coincidentally Paul O'Neill served on the Rand Corporation's board of directors and later as its chairman at the time it published Jim Dewar's 1993 paper on assumption-based planning (ABP) for the U.S. Army.[26] Those initials may look familiar. In fact, ABP and ABS are close cousins.

Jim Dewar and his coauthors at Rand developed ABP to deal with highly uncertain environments. While U.S. military planners in the Cold War were able to work around a single future scenario, they point out that "during very uncertain times, such as those of

today . . . [plans] that assume the likelihood of one particular world run the risk of being seriously wrong."[27]

For uncertain environments, they propose a planning process that starts with the key assumptions needed to project an outcome. Subsequent steps are to identify vulnerabilities of the key assumptions, indicators of those vulnerabilities, actions to shape outcomes favorably, and actions to hedge against unfavorable outcomes.

The Rand authors focus on what they call important assumptions. They define an assumption as "an assertion about some characteristic of the future that underlies the current operations or plans of an organization."[28] An assumption is important if "its negation would lead to significant changes in . . . current operations."[29] In other words, the assumptions of interest here could be wrong, and being wrong matters to results.

The hardest assumptions to identify are those we make implicitly. A simple example of an implicit assumption is, "The enemy cannot possibly approach by that route."[30] An example of one of the actual assumptions the Rand writers identified for their project was, "The Army will continue to play a primary role in maintaining global stability across the operational continuum."[31] It's a good example because there was little doubt about its importance, but it may no longer be true.

Of course, nothing in ABP can guarantee that you will identify all of the most important assumptions you're making, especially the implicit ones. This has been a recurring objection to ABP, and there's still disagreement whether it needs an answer. As Dewar and his coauthors note, the best remedy is repeated application of the process.[32]

Rand's ABP and Alcoa's ABS share more than a common patron in Paul O'Neill and a casual family resemblance. Both start by laying out the guesses we have to make in forming an expectation of future results that could lead to the biggest errors. For Alcoa managers, these guesses are strategic assumptions. Managers at different levels of the company will focus on assumptions about different things. But the assumptions must all be precise enough for performance

results to test them. Alcoa is an example of a firm that makes itself adaptable by laying out strategies clear enough to be testable.

Rita McGrath and Ian MacMillan draw an even clearer connection between learning from experience and testable assumptions in a variation on ABP that they describe in *The Entrepreneurial Mindset*.[33] Their starting point is the refrain of nearly every manager they met who was responsible for a new business initiative subject to the intricate budgets of a modern enterprise planning system: "The ink would barely be dry when unfolding experience revealed the numbers to be wrong."[34]

What comes after the refrain tells you all about the health of the company. If top management asks, "What will you change going forward?" it's probably fine. If top management asks, "How can I rely on your numbers?" the firm may be in trouble. "An organization in which [the latter] is happening," write the authors, "is an organization deprived of permission to learn."[35]

McGrath and MacMillan call their version of Paul O'Neill's and Rand's solution *discovery-driven planning* (DDP).[36] DDP starts with a reverse income statement that forces you to work backward from an income target to the main assumptions at the level of income statement line items that you must make to arrive at it. Although I'll criticize the adequacy of reverse income statements for generating testable assumptions in the next chapter, they at least force you to make assumptions more specific than your final goal. The method's next step generates a more detailed list of assumptions about the project. Its last step defines milestones at which you should test the assumptions.

Two features of DDP stand out. Although it pushes you to proceed with a project without waiting for perfect information to fall into your lap, it never mistakes the assumptions that let you proceed for facts. And it focuses your energy on testing those assumptions, starting with the ones that could have the biggest impact on your target.

McGrath and MacMillan suggest that a form of assumption-led planning resembling DDP made it possible for the National

Aeronautics and Space Administration (NASA) to simplify and streamline the enormously complex problem of getting to the moon. "Through three programs," they write, "the manned Mercury program, Project Gemini, and, finally, the Apollo missions, which eventually led to the successful lunar landing, NASA's rocket scientists faced a massive, ongoing learning challenge."[37] The key was to specify for the milestones of each program "what its staff would have to learn to acquire maximum confidence to go on to the next stage."[38]

NASA's milestones represented major assumptions about what it would have to achieve to land someone on the moon. The most prominent ones were developing reliable launch and recovery technology and orbiting the earth. And the focus of each stage of the programs was to minimize uncertainty around the remaining unsettled assumptions that had the largest impact on program goals. "The organization specified, in other words, what assumptions needed to be tested and validated at each stage so that NASA staff would have the knowledge needed to go to the next stage."[39]

The methods and examples of *Entrepreneurial Mindset* show how managers can adapt more quickly to experience by laying out testable assumptions and goals. The biggest challenge tends to be making sure that those assumptions are specific enough to test—and not just broad requirements for the possibility of hitting goals. But there are a few firms that test their assumptions rigorously. GE appears to be one.

There are too many great resources on GE's Six Sigma version of continuous process improvement to review it in detail here.[40] Nevertheless, one aspect of it is central to this chapter's overview of how and why organizations need testable strategies to learn from experience. GE uses expectations about outcomes to guide every performance improvement effort. The firm is a monument to the idea that what you learn depends on what you expect.

It derives those expectations in part from customer requirements. I wish I could also report that GE has found a way to test the customer perspective on which it relies in setting goals such

as quality targets. Since GE's growth no doubt broadens that perspective, I once thought its aggressive acquisition programs were a deliberate learning strategy. The firm won't generally justify a bid by what it expects to learn from the customers it's acquiring, however. GE may grow to learn, but it won't pay to learn.

The name "Six Sigma" reveals the most powerful way GE sets performance expectations across the myriad processes its businesses manage. It refers to the low probability of any result that is as far as six standard deviations above or below normal.[41] GE uses it as a goal for the rarity of defects in any measure of quality established to be critical to customer satisfaction. For example, an uptick in the frequency of defects in power turbine blades at a GE plant might push a quality measure from six to four sigma. The frequency of a defective result just four standard deviations above or below normal is a lot higher than that of a result six standard deviations away from normal.

One advantage of the goal is its utter lack of ambiguity. No matter what aspect of quality or what kind of process you're considering, it makes sense to measure the frequency of defects in the output of the process. The larger advantage, however, is that the measure is comparable across widely differing processes. Better still, the existence of a common quality standard or goal ensures a coordinated approach to quality in the ultimate product.

To take a simplistic example, suppose GE measured some aspect of the quality of raw material for its power turbine blades, as well as some aspect of quality of the finished product. It would make little sense to hold the material to a six sigma standard while holding the final product to a lower three sigma standard. Comparability in the levels of quality attained by the various parts of a production process makes sure no effort is wasted on quality at one stage only to be undermined by a quality breakdown at another.

The existence of a common quality standard promotes learning at GE because it sets an explicit expectation across complex production and service processes. Moreover, GE can raise or lower that expectation as it learns about any particular new product

or service. If GE set its quality expectations for an entirely new product—let's say fuel cells—as high as its expectations for power turbines, process improvement for the new product would quickly bog down under an avalanche of missed goals. A new fuel cell division could aim instead at a relatively low quality bar such as three sigma, and then raise all aspects of its production processes gradually to the level of older products.

In one sense, it's obvious that the expectations embedded in a six sigma quality standard promote learning. When quality defects in a process rise above the expected level, the process owner must launch an effort to control them, and that effort requires learning about the root causes of the defects.

There's also a deeper sense in which GE's quality expectations drive continuous learning. The highest expectations are a very sensitive indication of how well the firm understands a business process. You need to know a lot about a process to think you can hold defects below one in a million. If your theory about ensuring quality is rich enough, you'll be in a good position to probe for weak assumptions when you miss a quality goal. It's a little bit like throwing darts at a dartboard. If the target has no sections marked off except the center, it's hard to make corrections. On a board with a lot of well-marked sections, however, you can make precise adjustments.

A stringent and consistent quality standard forces you to separate what you know about a process into a lot of well-marked compartments. And that lets you keep track of which precise changes in the process improve the quality of its outcomes. By forcing clarity in your expectations about what affects quality, the standard accelerates learning.

There's just one fly in the ointment. How do you define a defect? How do you tell which aspects of quality matter and what level of quality is acceptable for each? Like Motorola before it, GE starts with the voice of the customer. Through satisfaction surveys, after-sales calls, service interactions, and even interactive Web sites, quality teams determine customers' critical-to-quality

(CTQ) requirements. For example, on-time delivery might be a CTQ requirement. If so, GE finds out in what time interval delivery would be acceptable. Deliveries outside that interval are defects.

Enlisting customers in goal setting has a long and successful history going back to the earliest days of quality control. Customer-defined quality goals lead process owners to formulate expectations about how to meet them that experience can quickly test. Even so, there's always a chance that the customer may not be right. If you're serving a cadre of loyal customers who don't care about your product's color, for example, you're unlikely to brighten it even if the unserved larger market might care a lot about it.

Clay Christensen has explored the idea that the customer is not always right in *The Innovator's Dilemma*.[42] There he provides the memorable example of makers of cable-actuated construction equipment missing the needs of an emerging class of owners of tract homes after World War II. They needed agile backhoes and steam shovels that only the inferior technology of hydraulics, then in its infancy, could provide. Of course, hydraulics manufacturers came to dominate construction equipment. Customers of the cable-actuated equipment manufacturers weren't right for their future.

Before Christensen published his book in 1997, however, and even before GE adopted Six Sigma in 1997, I thought the firm had an answer to the dilemma posed by the customer who is not always right. I thought the answer was to buy more customers.

As a young Lehman Brothers banker in the early 1990s, I helped AT&T Capital build up its book of equipment leases and other forms of equipment finance by acquiring smaller finance companies. The other big buyer in every deal—often victorious—was GE Capital. GE Capital's due diligence team would often be heading out the door of a finance company that had put itself up for bid as the AT&T Capital team came in, or vice versa. When I asked GE Capital staff why they were so aggressive in acquiring finance companies, they would say, "We're buying customers."

In retrospect, what they meant was simply that they were buying business opportunities they could develop through careful management. The four current and former GE executives with whom I talked while writing this book insisted they wouldn't let a learning opportunity become an excuse for an aggressive bid. They all agreed that GE benefits from growth that broadens its customer perspective. But hard-to-measure considerations like learning can erode the discipline of an acquisition program. Perhaps this book can change that by providing more concrete ways to measure learning.

Happily, there are easier ways to draw up a testable strategy than embarking on an aggressive acquisition program. Chapter Two proposes a method for distilling the strategy relevant to a manager at any level of a firm to a short list of testable assumptions. Setting out a testable strategy is the first step in the guess-test system of performance management. Starting with guesses in the form of testable strategies, it lets you pick out facts that can improve decisions from large amounts of conflicting performance data and uses performance results to sharpen those strategies continuously.