

HBR.ORG

Harvard Business Review

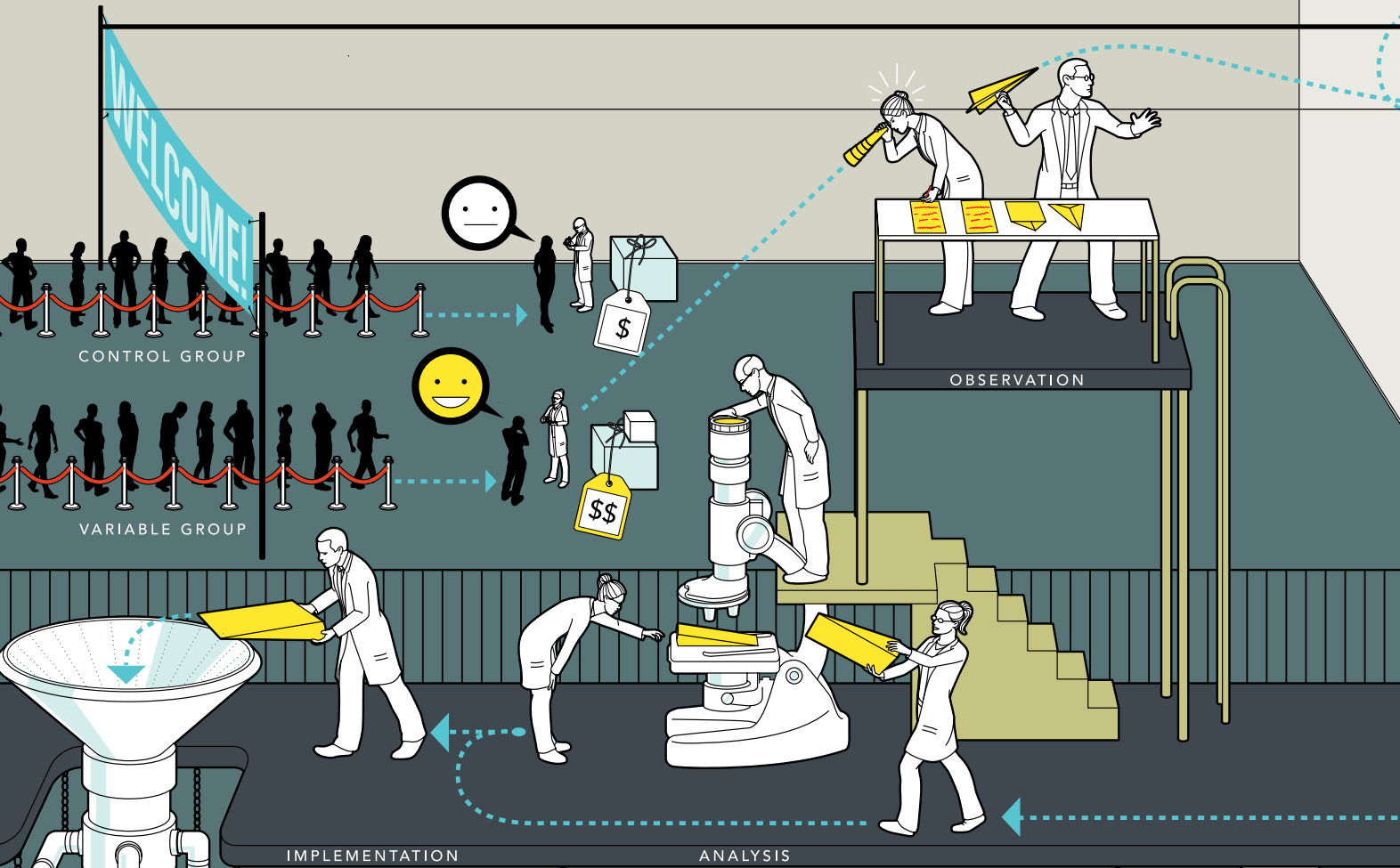


MARCH 2011
REPRINT R1103H

A Step-by-Step Guide to Smart Business Experiments

Every company can profit from testing customers' reactions to changes. Here's how to get started.

by Eric T. Anderson and Duncan Simester



A STEP-BY-STEP GUIDE TO

SMART BUSINESS

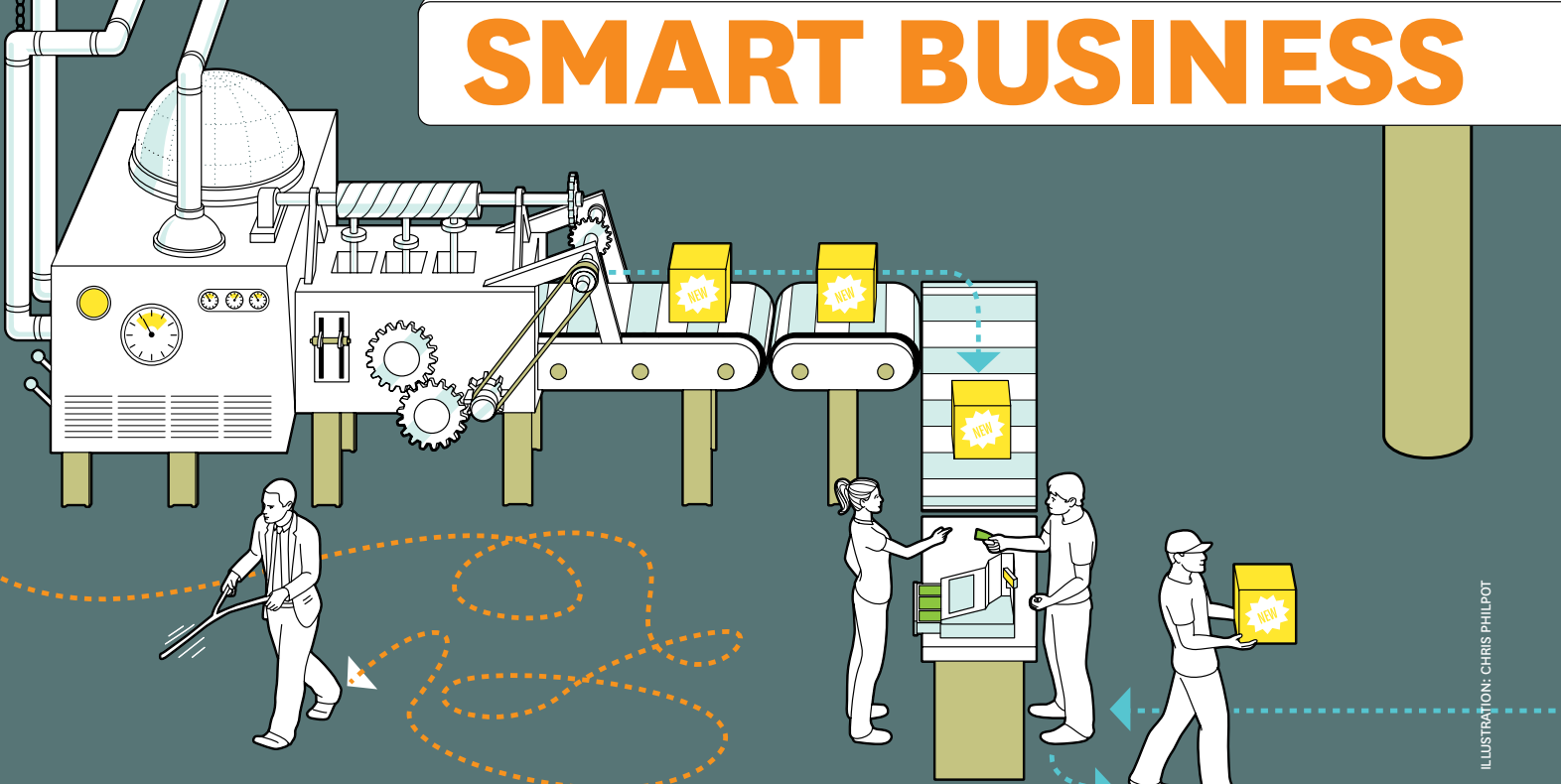
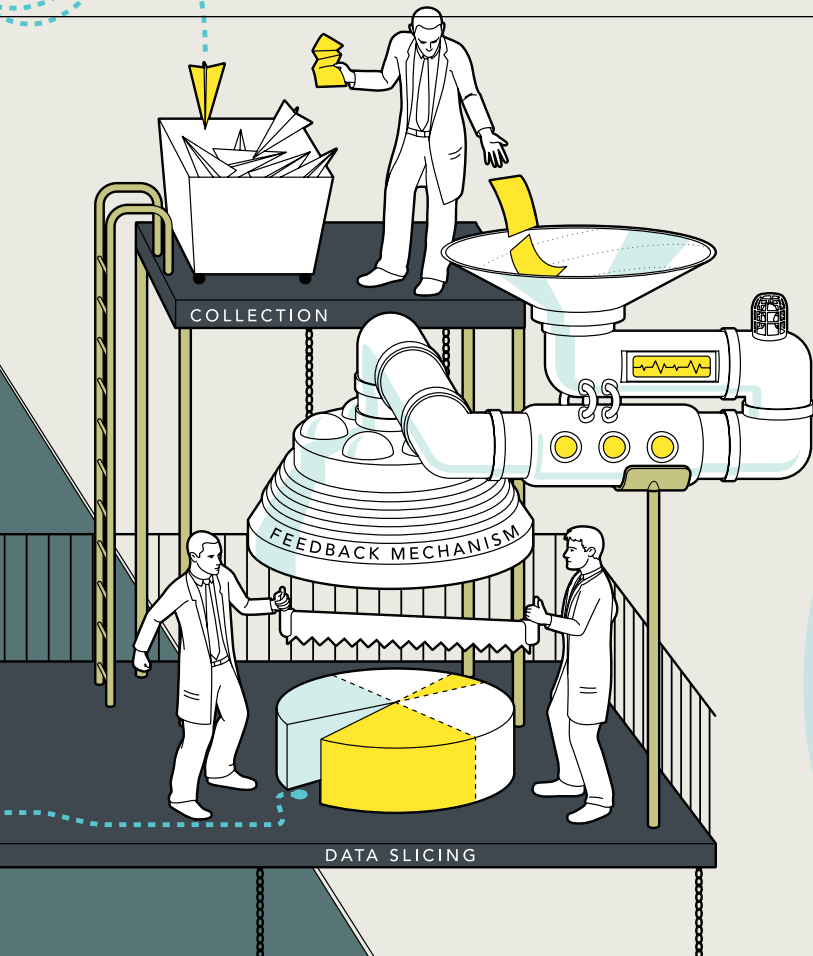


ILLUSTRATION: CHRIS PHILPOT



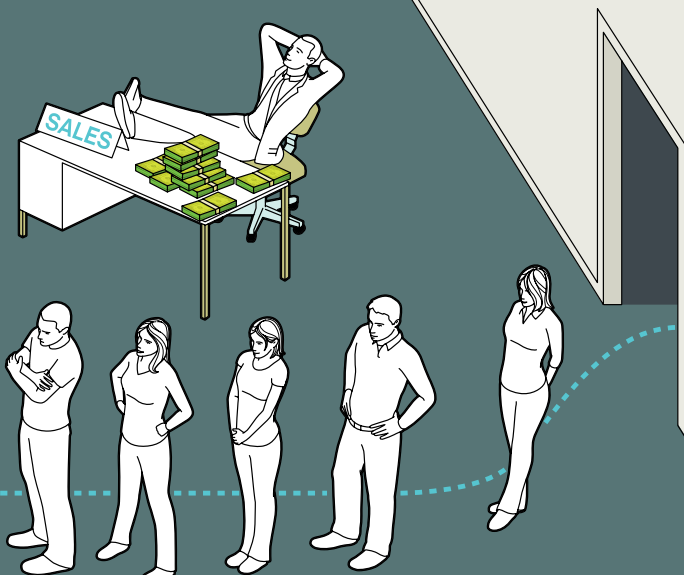
Every company can profit from testing customers' reactions to changes. Here's how to get started.
by Eric T. Anderson and Duncan Simester

OVER THE PAST DECADE, managers have awakened to the power of analytics. Sophisticated computers and software have given companies access to immense troves of data: According to one estimate, businesses collected more customer information in 2010 than in all prior years combined. This avalanche of data presents companies with big opportunities to increase profits—if they can find a way to use it effectively.

The reality is that most firms can't. Analytics, which focuses on dissecting past data, is a complicated task. Few firms have the technical skills to implement a full-scale analytics program. Even companies that make big investments in analytics often find the results difficult to interpret, subject to limitations, or difficult to use to immediately improve the bottom line.

Most companies will get more value from simple business experiments. That's because it's easier to draw the right conclusions using data generated through experiments than by studying historical transactions. Managers need to become adept at using basic research techniques. Specifically, they need to embrace the "test and learn" approach: Take one action with one group of customers, take a different action (or often no action at all) with a control group, and then compare the results. The outcomes are simple to analyze, the data are easily interpreted, and causality is usually clear. The test-and-learn approach is also remarkably powerful. Feedback from even a handful of experiments can yield immediate and dramatic improvements in profits. (See the sidebar "How One Retailer Tested Its Discount Strategy.") And unlike analytics, experimentation is a skill that nearly any manager can acquire.

EXPERIMENTS



Admittedly, it can be hard to know where to start. In this article, we provide a step-by-step guide to conducting smart business experiments.

It's All About Testing Customers' Responses

In some industries, experimentation is already a way of life. The J. Crew or Pottery Barn catalog that arrives in your mailbox is almost certainly part of an experiment—testing products, prices, or even the weight of the paper. Charitable solicitations and credit card offers are usually part of marketing tests, too. Capital One conducts tens of thousands of experiments each year to improve the way it acquires customers, maximizes their lifetime value, and even terminates unprofitable ones. In doing so, Capital One has grown from a small division of Signet Bank to an independent company with a market capitalization of \$19 billion.

The ease with which companies can experiment depends on how easily they can observe outcomes. Direct-mail houses, catalog companies, and online retailers can accurately target individuals with different actions and gauge the responses. But many companies engage in activities or reach customers through channels that make it impossible to obtain reliable feedback. The classic example is television advertising. Coke can only guess at how viewers responded to its advertising during the last Olympics, a limitation recognized by John Wanamaker's famous axiom, "Half the money I spend on advertising is wasted; the trouble is, I don't know which half." Without an effective feedback mech-

anism, the basis for decision making reverts to intuition.

In practice, most companies fall somewhere between these two extremes. Many are capable of conducting tests only at an aggregate level, and they're forced to compare nonequivalent treatment and control groups to evaluate the response. If Apple wants to experiment with the prices of a new iPhone, it may be limited to charging different prices in different countries and observing the response. In general, it's easier to experiment with pricing and product decisions than with channel management or advertising decisions. It's also easier to experiment in consumer settings than in business-to-business settings, because B2C markets typically have far more potential customers to serve as subjects.

Think Like a Scientist

Running a business experiment requires two things: a control group and a feedback mechanism.

Though most managers understand the purpose of control groups in experimentation, many companies neglect to use them, rolling out tests of new offerings across their entire customer base. A company that wants to evaluate the effect of exclusivity on its dealer network, for instance, is missing an opportunity if it offers all its dealers exclusivity. It should maintain nonexclusivity in certain regions to make it easier to evaluate how exclusivity affects outcomes.

Ideally, control groups are selected through randomization. When Capital One wanted to test the effectiveness of free transfers of balances from other

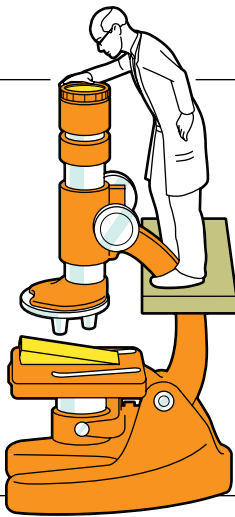
Overcoming Reluctance to Experiment

A large bank we worked with decided to use experiments to improve the way it advertised its certificates of deposit, a core product. In the past, decisions on ads had been made largely by a single manager, whose extensive experience endowed him with power and status within the organization—and a big salary.

The possibility that the bank would use experiments to supplant his intuitive decision making was a threat to the manager. Not surprisingly, he obstructed the process, arguing that planning lead times were too long and decisions had already been made. A senior leader whose P&L was directly affected by the advertising decisions had to intervene.

He allowed the experiments to go forward—and reassured his team that any missteps resulting from the experiments would not affect their year-end bonuses.

Organizational recalcitrance is one of the key hurdles companies encounter when trying to create a culture of experimentation. The main obstacle to establishing the new usual



Idea in Brief

Companies today understand the power of analytics, but dissecting past data is a complicated task that few firms have the technical skills to master. Most companies will get more value from simple business experiments.

To grow profits, managers need to become adept at techniques used by lab scientists and medical

researchers: They should establish control and treatment groups to test the effects of changes in price, promotion, or product variation. They should also grasp the opportunities provided by general changes in the business—like store openings—that constitute natural experiments in consumer behavior.

Creating a culture of experimentation requires companies to overcome internal political and organizational obstacles. And not every experiment will succeed. But over time, companies that embrace a test-and-learn approach are more apt to find the golden tickets that will drive growth.

credit cards (the innovation that initially launched its success), it offered the promotion to a random sample of prospective customers, while a different random sample (the control group) received a standard offer. Often it makes sense for a company to set up a treatment group and then use the remainder of the customer base as a control group, as one bank did when it wanted to experiment with its online retail trading platform. That approach gave bank managers a very large sample of equivalent customers against which to evaluate the response to the new platform.

The key to success with treatment and control groups is to ensure separation between them so that the actions taken with one group do not spill over to the other. That can be difficult to achieve in an online setting where customers may visit your website repeatedly, making it challenging to track which versions of the site they were exposed to. Separation can also be hard to achieve in traditional settings, where varying treatments across stores may lead to

spillovers for customers who visit multiple stores. If you cannot achieve geographic separation, one solution may be to vary your actions over time. If there is concern that changes in underlying demand may confound the comparisons across time, consider repeating the different actions in multiple short time periods.

The second requirement is a feedback mechanism that allows you to observe how customers respond to different treatments. There are two types of feedback metrics: behavioral and perceptual. Behavioral metrics measure actions—ideally, actual purchases. However, even intermediate steps in the purchasing process provide useful data, as Google's success illustrates. One reason Google is so valuable to advertisers is that it enables them to observe behavioral expressions of interest—such as clicking on ads. If Google could measure purchases rather than mere clicks, it would be even more valuable. Of course, Google and its competitors realize this and are actively exploring ways to measure the effects

is the old usual. Organizations have their ways of making decisions, and changing them can be a formidable challenge.

One mistake some firms make is to delegate experimentation to a customer intelligence group. This group has to lobby each business unit for the authority to conduct experiments. That's the wrong approach: Experiments are

designed to improve decision making, and so responsibility for them must occur where those decisions are made—in the business units themselves.

It is also important to set the right expectations. It's a mistake to expect every experiment to discover a more profitable approach—perhaps only 5% of them will do that. Those odds mean

that taking eight months to implement a single large-scale experiment is a bad strategy. Productive experimentation requires an infrastructure to support dozens of small-scale experiments. Of perhaps 100 experiments, only five to 10 will look promising and can be replicated, yielding one to two actions that are almost certain to be profitable. Focus your or-

ganization on these and scale them hard. Your goal, at least initially, is to find the golden ticket—you're not looking for lots of small wins.

Golden tickets can be hard to find, but that's largely because most organizations lack the perseverance to overcome the institutional resistance that stands in the way of discovering them.

Much of what companies learn from mammoth experiments can be gleaned from smaller tests that involve fewer variables, saving resources for follow-up tests.

of advertising on purchasing decisions in online and traditional channels.

Perceptual measures indicate how customers think they will respond to your actions. This speculative form of feedback is most often obtained via surveys, focus groups, conjoint studies, and other traditional forms of market research. These measures are useful in diagnosing intermediate changes in customers' decision processes.

Given that the goal of most firms is to influence customers' behavior rather than just their perceptions, experiments that measure behavior provide a more direct link to profit, particularly when they measure purchasing behavior.

Seven Rules for Running Experiments

As with many endeavors, the best experimentation programs start with the low-hanging fruit—experiments that are easy to implement and yield quick, clear insights. A company takes an action—such as raising or lowering a price or sending out a direct-mail offer—and observes customers' reactions.

You can identify opportunities for quick-hit experiments at your company using these criteria.

1 Focus on individuals and think short term. The most accurate experiments involve actions to individual customers, rather than segments or geographies, and observations of their responses. The tests measure purchasing behavior (rather than perceptions) and reveal whether changes lead to higher profits. Focus your experiments on settings in which customers respond immediately. When UBS was considering how to use experiments to improve its wealth management business, it recognized that the place to start was customer acquisition, not improving lifetime customer value. The effects of experiments on customer acquisitions can be measured immediately, while the impact on customer lifetime value could take 25 years to assess.

2 Keep it simple. Look for experiments that are easy to execute using existing resources and staff. When a bank wanted to run a customer experiment, it didn't start with actions that required retraining of retail tellers. Instead, it focused on ac-

tions that could be automated through the bank's information systems. Experiments that require extensive manipulation of store layout, product offerings, or employee responsibilities may be prohibitively costly. We know one retailer that ran a pricing experiment involving thousands of items across a large number of stores—a labor-intensive action that cost more than \$1 million. Much of what the retailer learned from that mammoth experiment could have been gleaned from a smaller test that used fewer stores and fewer products and preserved resources for follow-up tests.

3 Start with a proof-of-concept test. In academic experiments, researchers change one variable at a time so that they know what caused the outcome. In a business setting, it's important to first establish proof of concept. Change as many variables in whatever combination you believe is most likely to get the result you want. When a chain of convenience stores wanted to test the best way to shift demand from national brands to its private-label brands, it increased the prices of the national brands and decreased the private-label brand prices. Once it established that shifting demand was feasible, the retailer then refined its strategy by varying each of the prices individually.

4 When the results come in, slice the data. When customers are randomly assigned to treatment and control groups, and there are many customers in each group, then you may effectively have multiple experiments to analyze. For example, if your sample includes both men and women, you can evaluate the outcome with men and women separately. Most actions affect some customers more than others. So when the data arrive, look for subgroups within your control and treatment groups. If you examine only aggregate data, you may incorrectly conclude that there are no effects on any customers. (See the exhibit "Slicing an Experiment.")

The characteristics that you use to group customers, such as gender or historical purchasing patterns, must be independent of the action itself. For example, if you want to analyze how a store opening affects catalog demand, you cannot simply compare



How One Retailer Tested Its Discount Strategy

customers who made a purchase at the store with customers who did not. The results will reflect existing customer differences rather than the impact of opening the store. Consider instead comparing purchases by customers who live close to the new store versus customers who live far away. As long as the two groups are roughly equivalent, the differences in their behavior can be attributed to the store opening.

5 Try out-of-the-box thinking. A common mistake companies make is running experiments that only incrementally adjust current policies. For example, IBM may experiment with sales revenues by varying the wholesale prices that it offers to resellers. However, it may be more profitable to experiment with completely different sales approaches—perhaps involving exclusive territories or cooperative advertising programs. If you never engage in “what-if” thinking, your experiments are unlikely to yield breakthrough improvements. A good illustration is provided by Tesco, the UK supermarket chain. It reportedly discovered that it was profitable to send coupons for organic food to customers who bought wild birdseed. This is out-of-the-box thinking. Tesco allows relatively junior analysts at its corporate headquarters to conduct experiments on small numbers of customers. These employees deliver something that the senior managers generally don’t: a steady stream of creative new ideas that are relevant to younger customers.

6 Measure everything that matters. A caution about feedback measures: They must capture all the relevant effects. A large national apparel retailer recently conducted a large-scale test to decide how often to mail catalogs and other promotions to different groups of customers. Some customers received 17 catalogs over nine months, whereas another randomly selected group received 12 catalogs over the same time period. The retailer discovered that for its best customers the additional catalogs increased sales during the test period, but lowered sales in subsequent months. When the retailer compared sales across its channels, it found that its best customers purchased more often through the catalog channel (via mail and telephone) but less from its online stores. When the firm aggregated sales across the different time periods and across its retail channels, it concluded that it could mail a lot less

Walk into any large retail store, and you’ll find price promotions being offered on big national brands—discounts that are funded by the manufacturer. For retailers, these promotions can be a mixed bag: Although the lower prices may increase sales of the item, the promotion may hurt sales of competing private-label products, which offer higher margins. We worked with one national retailer that decided to conduct experiments to determine how it might shield its private-label market share by promoting these products while the national brands were on sale.

The retailer designed six experimental conditions—a control and five discount levels—that ranged from zero to 35% for the private-label items. The retailer divided its stores into six groups, and the treatments were randomized across the groups. This meant each store had a mixture of the experimental conditions distributed across the different products in the study. For example, in Store A private-label sugar was discounted 20%, and private-label mascara was full price, whereas in Store B mascara was discounted, but sugar was not. This experimental design allowed the retailer to control for variations in sales that occurred because the store groups were not identical.

The test revealed that matching the national brand promotions with moderate discounts on the private-label products generated 10% more profits than not promoting the private-label items. As a result, the retailer now automatically discounts private-label items when the competing national brands are under promotion. After establishing proof of concept, it refined its shielding policies by testing responses to various types of national brand promotions. For example, it discovered that a “Buy

One, Get One for 50% Off” promotion on a national brand should also be matched with the same offer on the private label, rather than just a straight discount.

These experiments were successful for two main reasons: The actions were easy to implement, and the results were easy to measure. Each set of experiments lasted just one week. There were few enough products involved that stores did not require any additional labor. The experiments piggybacked on the standard procedures for promoting an item—indeed, the store employees were unaware that they were helping to implement an experiment.

In previous experiments the retailer had learned that if it changed too many things at once, the stores could not handle the implementation without long delays and a lot of additional cost. In some cases temporary labor had to be trained to go into the stores, find the products, and change the prices and shelf signage. Moreover, if the experiment extended beyond a week, problems arose as shelves were constantly rearranged and new signs applied. A maintenance program was required to monitor store compliance. In effect the retailer experimented on experimentation itself—it learned how to design studies that it could analyze more quickly and implement more easily.



Slicing an Experiment

When you're conducting an experiment, it's important to remember that initial results may be deceiving. Consider a publishing company that tried to assess how discounts affect customers' future shopping behavior. It mailed a control group of customers a catalog containing a shallow discount—its standard practice. The treatment group of customers received catalogs with deep discounts on certain items. For two years, the company tracked purchases at an aggregate level, and the difference between the two groups was negligible:

ALL CUSTOMERS

Customers who received the catalog with shallow discounts spent:

PER CUSTOMER
\$159
 CONTROL

Customers who received the catalog with deep discounts spent:

\$157
 TEST

.....
\$(2)
 CHANGE

But that view of the data did not tell the whole story. Further analysis revealed a disturbing outcome among customers who had recently purchased a high-priced item and then received a catalog offering the same item at a 70% discount. Apparently upset by this perceived overcharge, these customers (some of them long-standing ones) cut future spending by 18%:

CUSTOMERS WHO RECENTLY BOUGHT A HIGH-PRICED ITEM

Customers who received the catalog with shallow discounts spent:

PER CUSTOMER
\$506
 CONTROL

Customers who received the catalog with deep discounts spent:

\$415
 TEST

.....
\$(91)
 CHANGE

Upon learning these results, the publishing firm modified its direct-mail approach to avoid inadvertently antagonizing its best customers.

frequently to its best customers without sacrificing sales. Viewing results in context is critical whenever actions in one channel affect sales in other channels or when short-term actions can lead to long-run outcomes. This is the reason that we recommend starting with actions that have only short-run outcomes, such as actions that drive customer acquisition.

7 Look for natural experiments. The Norwegian economist Trygve Haavelmo, who won the 1989 Nobel Prize, observed that there are two types of experiments: “those we should like to make” and “the stream of experiments that nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers.” If firms can recognize when natural experiments occur, they can learn from them at little or no additional expense. For example, when an apparel retailer opened its first store in a state, it was required by law to start charging sales tax on online and catalog orders shipped to that state, whereas previously those purchases had been tax-free. This provided an opportunity to discover how sales taxes affected online and catalog demand. The retailer compared online and catalog sales before and after the store opening for customers who lived on either side of the state’s southern border, which was a long way from the new store. None of the customers were likely to shop in the new store, so its opening would have no effect on demand—the only change was the taxation of online and catalog purchases, which affected consumers only on one side of the border. The comparison revealed that the introduction of sales taxes led to a large drop in online sales but had essentially no impact on catalog demand.

The key to identifying and analyzing natural experiments is to find treatment and control groups that were created by some outside factor, not specifically gathered for an experiment. Geographic segmentation is one common approach for natural experiments, but it will not always be a distinguishing characteristic. For example, when GM, Ford, and Chrysler offered the public the opportunity to purchase new cars at employee discount levels, there was no natural geographic separation—all customers were offered the deal. Instead, to evaluate the outcome of these promotions, researchers compared transactions in the weeks immediately before and after the promotions were introduced. Interestingly, they discovered that the jump in sales levels was accompanied by a sharp increase in prices. Customers thought they were getting a good deal, but in reality

prices on many models were actually lower before the promotion than with the employee discount prices. Customers responded to the promotion itself rather than to the actual prices, with the result that many customers were happy with the deal, even though they were paying higher prices.

Avoid Obstacles

Companies that want to tap into the power of experimentation need to be aware of the obstacles—both external and internal ones. In some cases, there are legal obstacles: Firms must be careful when charging different prices to distributors and retailers, particularly firms competing with one another. Although there are fewer legal ramifications when charging consumers different prices (the person sitting next to you on your airline flight has usually paid more or less than you), the threat of an adverse consumer reaction is a sufficient deterrent for some firms. No one likes to be treated less favorably than others. This is particularly true when it comes to prices, and the widespread availability of price information online means that variations are often easily discovered.

The internal obstacles to experimentation are often larger than the external barriers. In an organization with a culture of decision making by intuition, shifting to an experimentation culture requires a fundamental change in management outlook. Management-by-intuition is often rooted in an individual's desire to make decisions quickly and a culture that frowns upon failure. In contrast, experimentation requires a more measured decision-making style and a willingness to try many approaches, some of which will not succeed.


Some companies mistakenly believe that the only useful experiments are successful ones. But the goal is not to conduct perfect experiments; rather, the goal is to learn and make better decisions than you are making right now. Without experimentation, managers generally base decisions on gut instinct. What's surprising is not just how bad those decisions typically are, but how good managers feel about them. They shouldn't—there's usually a lot of room for improvement. Organizations that cultivate a culture of experimentation are often led by senior managers who have a clear understanding of the opportunities and include experimentation as a strategic goal of the firm. This is true of Gary Loveman, the CEO of Harrah's, now called Caesars Entertainment, who transformed the culture of a 35,000-employee

organization to eventually enshrine experimentation as a core value. He invested in the people and infrastructure required to support experimentation and also enforced a governance mechanism that rewarded this approach. Decisions based solely upon intuition were censured, even if the hunch was subsequently proved correct.


There is generally a practical limit on the number of experiments managers can run. Because of that, analytics can play an important role, even at companies in which experiments drive decision making. When Capital One solicits new cardholders by

The goal is not to conduct perfect experiments; it is to make better decisions.

mail, it can run thousands of experiments; there's no need to pretest the experiments by analyzing historical data. But other companies' business models may allow for only a few experiments; in such cases, managers should carefully plan and pretest experiments using analytics. For example, conducting experiments in channel settings is difficult because changes involve confrontation and disruption of existing relationships. This means that most firms will be limited in how many channel experiments they can run. In these situations, analyzing historic data, including competitors' actions and outcomes in related industries, can offer valuable initial insights that help focus your experiments.

Whether the experiments are small or large, natural or created, your goal as a manager is the same: to shift your organization from a culture of decision making by intuition to one of experimentation. Intuition will continue to serve an important role in innovation. However, it must be validated through experimentation before ideas see widespread implementation. In the long run, companies that truly embrace this data-driven approach will be able to delegate authority to run small-scale experiments to even low levels of management. This will encourage the out-of-the-box innovations that lead to real transformation. 

[HBR Reprint R1103H](#)

 **Eric T. Anderson** is the Hartmarx Professor of Marketing at Northwestern's Kellogg School of Management. **Duncan Simester** is the NTU Professor of Management Science at MIT's Sloan School of Management.