Chapter 9

Marginal Gains

I

At around 9 a.m. the riders of Team Sky, the British professional cycling team, made their way out of a small hotel in Carcassonne, a beautiful town in the Languedoc-Roussillon region of southern France. It was a warm morning and the riders walked to the team bus in silence, contemplating the day to come.

They were about to start Stage 16 of the 2014 Tour de France, one of the sternest tests of endurance in the sporting world. They had already ridden 3,000 kilometers over the preceding fifteen stages and now faced a 237.5-kilometer ride culminating at the feared Port de Balès, a 19-kilometer climb into the Pyrenees. “Here we go again,” Bernhard Eisel, one of the team members, said with a grim smile.

On the Team Sky bus there was a sense of anticipation. The riders were getting into their sports gear. The coaches were reviewing race plans. With thirty minutes to go, Nicolas Portal, one of Team Sky’s sporting directors, began his pre-race briefing. He talked about the importance of the stage and alerted the riders to difficult sections along the route. As he did so photographs of tough corners and steep climbs were flashed onto a screen at the front of the bus.

As he finished his talk, a man toward the back, silent until that moment, started to speak. He had a shaved head, dark-rimmed glasses and an intense manner. He is the man who always has the final word before the race: the general manager of Team Sky, Sir David Brailsford.

“At the end of the day, success is about getting in the breakaway [where a group of cyclists ride away from the main pack],” he said. “Let’s not f*** about. Either we are in it or we are not. I know it is difficult. I know how

hard it is. But everyone needs to buy into this. All focus on that. That is our goal for today. The rest will look after itself. Don't let anyone else make it happen; make it happen for yourselves . . . OK, hit it!"

A quiet buzz reverberated around the bus. Brailsford had struck the right note. All eight riders stood up and exchanged glances. They then made their way down the steps to the starting line of the sixteenth stage.

• • •

The previous evening Brailsford had given me a tour of the Team Sky operation. We looked at the trucks, the design of the team bus, and the detailed algorithms that are used to track the performance of each cyclist. It was an opportunity to glimpse behind the curtains of one of the most admired and tightly policed operations in all sport.

The success of Brailsford is legendary. When he joined British track cycling as an adviser in 1997, the team was behind the curve. In 2000 Great Britain won a single Olympic gold medal in the time trial. In 2004, one year after Brailsford was appointed performance director, Britain won two Olympic gold medals. In 2008 they won an astonishing eight gold medals and, at the London Olympics in 2012, repeated the feat.

Meanwhile, something even more remarkable was happening. Track cycling is competitive, but the most prestigious form of the sport is professional road cycling. Britain had never had a winner of the Tour de France since the race was established in 1903. British riders had won individual stages, but nobody had come close to winning the general classification.

But in 2009, even as the British track cycling team was preparing for the London Olympics, Brailsford embarked upon a new challenge. He created a road cycling team, Team Sky, while continuing to oversee the track team. On the day the new outfit was announced to the world, Brailsford also announced that they would win the Tour de France within five years.

Most people laughed at this aspiration. One commentator said: "Brailsford has set himself up for an almighty fall." But in 2012, two years ahead of schedule, Bradley Wiggins became the first-ever British rider to win the event. The following year, Team Sky triumphed again when Chris

Froome, another Brit, won the general classification. It was widely acclaimed as one of the most extraordinary feats in British sporting history.

How did it happen? How did Brailsford conquer not one cycling discipline, but two? These were the questions I asked him over dinner at the team's small hotel after the tour of the facilities.

His answer was clear: "It is about marginal gains," he said. "The approach comes from the idea that if you break down a big goal into small parts, and then improve on each of them, you will deliver a huge increase when you put them all together."

It sounds simple, but as a philosophy, marginal gains has become one of the hottest concepts not just in sports, but beyond. It has formed the basis of business conferences, and seminars and has even been debated in the armed forces. Many British sports now employ a director of marginal gains.

But what does this philosophy actually mean in practice? How do you deliver a marginal gains approach, not just in sport, but in other organizations? Most significantly of all, why does breaking a big project into smaller parts help you to tackle really ambitious goals?

To glimpse an answer, let us leave cycling for a moment and look at a very different area of life. For it turns out that the best way to grasp the meaning of marginal gains is to examine one of the most pressing issues facing the world today: global poverty.

II

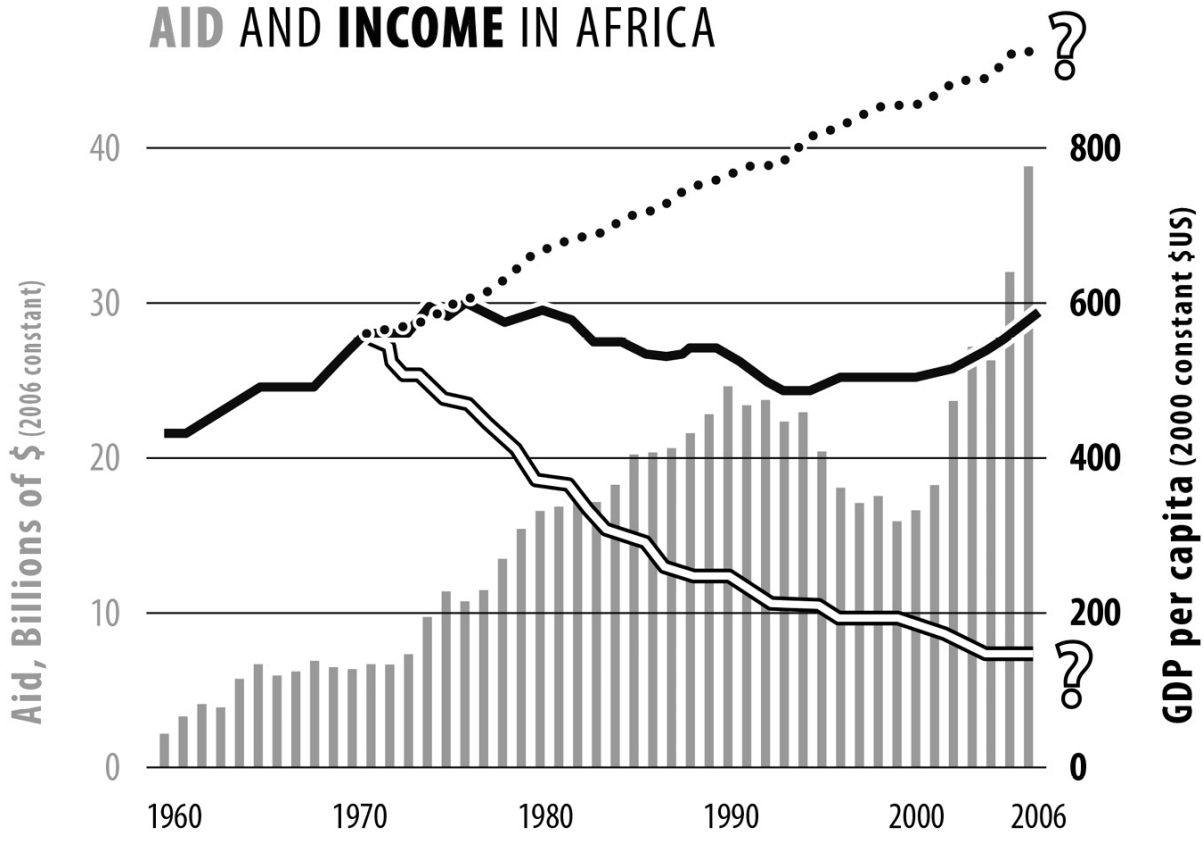
Take a look at the graph here.[1] It is reproduced from the work of Esther Duflo, one of the world's most respected economists, currently working out of MIT.

The vertical, light-gray bars show the amount of aid spending on Africa over the last thirty years. As you can see, the funding has gradually increased since the early 1960s, peaking at almost $800 million in 2006. The investment has a simple imperative: to improve the lives of the world's poorest. It is an important objective given that 25,000 children die of preventable causes every day.[2]

The key question here is, Did the investment make a difference? Did it improve the lives of the people it was designed to help?

A sensible place to start when answering that question is with African GDP. In the diagram African GDP is shown by the solid black line. As you can see, this has stayed roughly constant over the period. This might lead one to the conclusion that all the aid spending hasn't done much good. It hasn't boosted economic activity. It hasn't raised the living standards of those living in Africa. In fact it all seems like an expensive waste of time.

But the insights from the previous chapter should urge a little caution. Why? Because the data don't give us an insight into the counterfactual. Perhaps the aid spending was incredibly successful. Perhaps, without it, GDP in Africa would have been far lower—the white line in the graph.



Of course, there is another possibility. Perhaps aid spending was even more detrimental than the solid black line might lead you to believe. Perhaps it was a disaster, destroying incentives, boosting corruption, and

lowering growth below what it would otherwise have been. Perhaps without it Africa would have actually surged ahead: as per the dotted line in the graph. How can we know either way?

Each of these two alternatives has high-profile supporters. Jeffrey Sachs, director of the Earth Institute at Columbia University, for example, is a vocal advocate of development spending. He argues that aid has benefited the lives of Africans and claims that more money could eradicate poverty altogether. *The End of Poverty*, his best-selling book, is based in part upon this premise.[3]

Conversely, William Easterly, an economist at New York University, profoundly disagrees. He argues that aid spending has had all sorts of negative side effects, and that Africa would have been better off without it. His book *The White Man's Burden* presents this case with as much intellectual force as that of Sachs.[4]

The best way to adjudicate between these stances would be to conduct a randomized control trial. This would enable us to isolate the effect of development spending from all the other influences on African GDP. But there is a rather obvious problem. There is only one Africa. You cannot find lots of different Africas, randomly divide them into groups, give aid to some and not to others, and then measure the outcomes.

This may sound like a trivial point, but it has wider implications. When it comes to really big issues, it is very difficult to conduct controlled experiments. To run an RCT you need a control group, which is not easy when the unit of analysis is very large. This applies to many things beyond development aid, such as climate change (there is only one world), issues of war and peace, and the like.

This brings us directly to the concept of marginal gains. If the answer to a big question is difficult to establish, why not break it down into lots of smaller questions? After all, aid spending has many subcomponents. There are programs on malaria, literacy, road-building, education, and infrastructure, each of them constructed in different ways, with different kinds of incentives, and delivered by different organizations.

At this level of magnification, by looking at one program at a time, it is perfectly possible to run controlled experiments. You try out the program with some people or communities, but not with others, and then compare

the two groups to see if it is working or not. Instead of debating whether aid is working *as a whole* (a debate that is very difficult to settle on the basis of observational data), you can find definitive answers at the smaller level and build back up from there.

To examine a concrete example, suppose you were trying to improve educational outcomes in Africa. One way to see if aid spending is working would be to look at the correlation between the quantity of spending and the average grade score across the continent. The problem is that this wouldn't give you any information about the counterfactual (what would have happened to scores without the funding).

But now suppose that instead of looking at the big picture, you examine an individual program. That is precisely what a group of pioneering economists did in the impoverished Busia and Teso regions in the west of Kenya. As the author Tim Harford points out in his book *Adapt*, these economists wanted to know whether handing out free textbooks to schools would boost grades. Intuitively, they were pretty sure it would. In the past the observational data had been good. Schools that received books tended to improve their test scores.

But the economists wanted to be sure, so they performed an RCT. Instead of giving the textbooks to the most deserving schools, which is the common approach, they randomly divided a number of eligible schools into two groups: one group received free textbooks and the other group did not. Now, the charity had a treatment group and a control group. They had a chance to examine whether the books were making a real difference.

The results, when they came in, were both emphatic and surprising. The students in the schools that received free textbooks didn't perform any better than those who did not. The test results in the two groups of schools were almost identical. This outcome contradicted intuition and the observational data. But then randomized trials often do.

The problem, it turned out, was not the books, but the language they were written in. English is the third language of most of the poor children living in remote Busia and Teso. They were struggling to grasp the material as it was presented. Researchers might not have realized this had they not run a trial. It pierced through to one of the untested assumptions in their approach.

Confronted with failure, the economists tried another approach. They conducted another randomized trial but instead of using textbooks they used visual aids. These were flipcharts with bold graphics that covered geography, math, etc. Again, the economists expected them to boost test scores. And again, when they compared the test scores in the treatment group with those of the control group, the flipcharts were a failure. They led to no significant improvement in learning.

Undeterred, the economists started to think about the problem in a fresh way. They tried something completely new: a de-worming medication. This may seem like a curious way to improve education, but researchers were aware that these parasites stunt growth, cause children to feel lethargic, and lead to absenteeism. They disproportionately affect children in remote communities, just like those in Busia and Teso.

This time the results were excellent. They vastly exceeded the expectations of the researchers. As Tim Harford put it: “The program was a huge success, boosting children’s height, reducing re-infection rates, and also reducing absenteeism from school by a quarter. And it was cheap.”[5]

This was a marginal gain. It was just one program in one small region. But by looking at education at this level of magnification, it was possible to see what really works, and what doesn’t. The economists had tested, failed, and learned. They could now roll it out in other areas, while continuing to test, and iterate, and create yet more marginal gains.

This may sound like a gradual way to improve, but look at the alternative. Consider what would have happened if the economists had relied on intuition and observational data. They might have continued with free textbooks forever, deluding themselves that they were making a difference, when they were doing virtually nothing at all.

This approach is now the focus of a crusading group of economists who have transformed international development over the last decade. They do not come up with grand designs; rather, they look for small advantages. As Esther Duflo, the French-born economist who is at the forefront of this approach, put it: “If we don’t know if we are doing any good, we are not any better than the medieval doctors and their leeches. Sometimes the patient gets better; sometimes the patient dies. Is it the leeches or something else? We don’t know.”[6]

Critics of randomized trials often worry about the morality of "experimenting on people." Why should one group get X while another is getting Y? Shouldn't everyone have access to the best possible treatment? Put like this, RCTs may seem unethical. But now think about it in a different way. If you are genuinely unsure which policy is the most effective, it is only by running a trial that you can find out. The alternative is not morally neutral, it simply means that you never learn. In the long run this helps nobody.

Duflo, who is petite and dynamic, doesn't regard her work as lacking in ambition; rather, she regards these incremental improvements as pioneering. She told me:

> It is very easy to sit back and come up with grand theories about how to change the world. But often our intuitions are wrong. The world is too complex to figure everything out from your armchair. The only way to be sure is to go out and test your ideas and programs, and to realize that you will often be wrong. But that is not a bad thing. It leads to progress.

This links back to the work of Toby Ord, whom we met in chapter 7. He uses the data discovered by the likes of Duflo to advise private individuals on where to donate their money. He realized that relying on hunch and narrative can mean that millions of pounds are squandered on ineffective programs. And this is why hundreds of controlled experiments are now being conducted across the developing world. Each test demonstrates whether a policy or program works, or if it doesn't.

Each test provides a small gain of one kind or another (remember that failure is not inherently bad: it sets the stage for new ideas). By breaking a big problem into smaller parts, it is easier to cut through narrative fallacies. You fail more, but you learn more.

As Duflo puts it: "It is possible to make significant progress against the biggest problem in the world through the accumulation of a set of small steps, each well thought out, carefully tested, and judiciously implemented."[7]

III

And this takes us back to David Brailsford and British cycling. Note the similarity of the final quote of Duflo with that of Brailsford earlier in this chapter. “The whole approach comes from the idea that if you break down a big goal into small parts, and then improve on each of them, you will gain a huge increase when you put them all together.”

Cycling is very different from international development, but the success of its most pioneering coach is based on the same conceptual insight. As Brailsford puts it: “I realized early on that having a grand strategy was futile on its own. You also have to look at a smaller level, figure out what is working and what isn’t. Each step may be small, but the aggregation can be huge.”

Running controlled trials in cycling is significantly easier than in development aid, not least because the aim of the sport is relatively simple: getting from A to B as quickly as possible. To obtain the most efficient bicycle design, for example, British cycling created a wind tunnel. This enabled them to isolate the aerodynamic effect, by varying the design of the bike and testing it in identical conditions. To discover the most efficient training methods, Brailsford created new data sets that enabled him to track every subcomponent of physiological performance.

“Each gain on its own was small,” Brailsford said. “But that doesn’t really matter. We were getting a deeper understanding of each aspect of performance. It was the difference between trailing behind the rest of the world and coming first.”

In *Corporate Creativity*, the authors Alan Robinson and Sam Stern write of how Bob Crandall, the former chairman of American Airlines, removed a single olive from every salad, and in doing so saved $500,000 annually.[8] Many seized on this as a marginal gain. But was it? After all, if removing an olive is a good idea, why not the lettuce too? At what point does an exercise in incremental cost-cutting start to impact on the bottom line?

Now we can see a clear answer. Marginal gains is not about making small changes and hoping they fly. Rather, it is about breaking down a big problem into small parts in order to rigorously establish what works and

what doesn't. Ultimately the approach emerges from a basic property of empirical evidence: to find out if something is working, you must isolate its effect. Controlled experimentation is inherently "marginal" in character.

Brailsford puts it this way: "If you break a performance into its component parts, you can build back up with confidence. Clear feedback is the cornerstone of improvement. Marginal gains, as an approach, is about having the intellectual honesty to see where you are going wrong, and delivering improvements as a result."

The marginal gains mentality has pervaded the entire Team Sky mindset. They make sure that the cyclists sleep on the same mattress each night to deliver a marginal gain in sleep quality; that the rooms are vacuumed before they arrive at each new hotel, to deliver a marginal gain in reduced infection; that the clothes are washed with skin-friendly detergent, a marginal gain in comfort.

"People think it is exhausting to think about success at such a high level of detail," Brailsford says. "But it would be far more exhausting, for me anyway, to neglect doing the analysis. I would much rather have clear answers than to delude myself that I have the 'right' answers."

• • •

Perhaps the most astonishing application of marginal gains is to be found not in cycling but in Formula One. In the closing weeks of the 2014 season I visited the Mercedes headquarters in Brackley, a few miles north of Oxford. It is a series of gray buildings on an industrial estate, with a stream running through it. It is populated with bright people, passionate about their sport—and whose attention to detail is staggering.

"When I first started in F1, we recorded eight channels of data. Now we have 16,000 from every single parameter on the car. And we derive another 50,000 channels from that data," said Paddy Lowe, a Cambridge-educated engineer, who is currently the technical leader of Mercedes F1. "Each channel provides information on a small aspect of performance. It takes us into the detail, but it also enables us to isolate key metrics that help us to improve."

The most intuitive way to glimpse the relationship between marginal gains and big achievements is to examine the pit stop. This is one of

thousands of different components that, collectively, determine whether an F1 team is successful or not. It is a marginal aspect of performance, but a crucial one. In order to gain a deeper insight I went out to the season-ending Grand Prix in Abu Dhabi and immersed myself within the Mercedes operation.

At the team's motor home, a small, three-story house within the Yas Marina Circuit, I talked to James Vowles, chief strategist for Mercedes F1. I asked him how the team went about developing the optimum pit-stop procedure. Vowles says:

> We use the same method for everything, not just pit stops. First of all, you need a decent understanding of the engineering problem. So, with the pit stops we came up with a strategy based on our blue-sky ideas. But this strategy was always going to be less than optimal, because the problem is complex. So we created sensors so we could measure what was happening and test our assumptions.
>
> But the crucial thing is what happened next. Once you have gone through a practice cycle with the initial strategy, you immediately realize that there are miscellaneous items that you are not measuring. Just doing a pit-stop practice-run opens your eyes to data points that are relevant to the task, but that were absent from the initial blueprint. So the second stage of the cycle is about improving your measurement statistics, even before you start to improve the pit-stop process.

Think about that for a moment. We have talked about the concept of an open loop. This is where a strategy is put in action, then tested to see if it is working. By seeing what is going wrong, you can then improve the strategy. Mercedes takes this one step further. They use the first test not to improve the strategy, but to create richer feedback. Only when they have a deeper understanding of all the relevant data do they start to iterate.

Vowles says:

> We have placed eight sensors on every single one of the wheel-nut guns in order to access the most systematic data. Just by looking at

> this data, without speaking to the human involved, I can ascertain exactly what has happened on each pit stop. When the gun operator initially connected to the wheel nut, I can tell that they, say, connected 20 degrees off the optimum angle. When they start rotating the gun, I can tell how long it has taken for the nut to physically loosen all its preloaded torque and for the wheel to start moving off the axle.
>
> I can tell how quickly the gun man has moved away; how quickly he has reconnected, how long it has taken for the tire to be removed, the second tire to be refitted to the axle, how clean the second connection was to it, and how long he was gunning on for. The precision of this information helps us to create an optimization loop. It shows us how to improve every time-sensitive aspect.

This is marginal gains on turbocharge. “You improve your data set before you begin to improve your final function; what you are doing is ensuring that you have understood what you didn’t initially understand,” Vowles says. “This is important because you must have the right information at the right time in order to deliver the right optimization, which can further improve and guide the cycle.”

Later that evening I went to the pit-lane to watch the team practice. It was an astonishing feat of collective endeavor. The car of Lewis Hamilton, the top driver for Mercedes, was pushed into position by three runners, and then instantly pounced upon by a team of around sixteen people, all with clearly defined tasks and exquisitely coordinated procedures. Again and again they practiced, dealing with every contingency that might arise in the race the next day. Every practice run was measured with the eight sensors, and videotaped, so it could pass through another optimization loop. One of the pit stops I witnessed was completed in an astonishing 1.95 seconds.*

Vowles said:

> The secret to modern F1 is not really to do with big ticket items; it is about hundreds of thousands of small items, optimized to the nth degree. People think that things like engines are based upon high-level strategic decisions, but they are not. What is an engine except

> many iterations of small components? You start with a sensible design, but it is the iterative process that guides you to the best solution. Success is about creating the most effective optimization loop.

I also spoke to Andy Cowell, the leader of the team that devised the engine. His attitude was a carbon copy of that of Vowles.

> We got our development engine up and running in late December [2012]. We didn't design it to be car friendly. We didn't try and figure out the perfect weight and aerodynamic design. Rather, we got a working model out there early, so that we could test it, and improve. It was the process of learning in the test cell that enabled us to create the most thermally efficient engine in the world.

The marginal gains approach is not just about mechanistic iteration. You need judgment and creativity to determine how to find solutions to what the data is telling you, but those judgments, in turn, are tested as part of the next optimization loop. Creativity not guided by a feedback mechanism is little more than white noise. Success is a complex interplay between creativity and measurement, the two operating together, the two sides of the optimization loop.

We will examine the creative process in more detail in the next chapter, but Vowles and Cowell have described a compelling model. It is the model used by Brailsford and the latest generation of development economists. Mercedes clocks up literally thousands of tiny failures. As Toto Wolff, the charismatic executive director of the team, put it: "We make sure we know where we are going wrong, so we can get things right."

The basic proposition of this book is that we have an allergic attitude to failure. We try to avoid it, cover it up, and airbrush it from our lives. We have looked at cognitive dissonance, the careful use of euphemisms, anything to divorce us from the pain we feel when we are confronted with the realization that we have underperformed.

Brailsford, Duflo and Vowles see weaknesses with a different set of eyes. Every error, every flaw, every failure, however small, is a marginal

gain in disguise. This information is regarded not as a threat but as an opportunity. They are, in a sense, like aviation safety experts, who regard every near-miss event as a precious chance to avert an accident before it happens.*

On the eve of the Grand Prix at the Yas Marina Circuit, qualifying took place. This is where the drivers compete to see who can post the fastest lap, with the winner taking pole position (the most advantageous place on the starting grid) for the Grand Prix. Nico Rosberg, a German driver for Mercedes, took first place on the grid and Lewis Hamilton, his British teammate, took second place.

Afterward, I was given access to the highly secretive debriefing meeting. At a table in a room in the Mercedes garage, a few meters from the track, Hamilton and Rosberg sat facing each other. They were flanked by their respective race engineers. On the left was Paddy Lowe, the technical boss, and on other tables were experts in different aspects of performance.

Everybody wore headsets with microphones and scrutinized data on computer screens. On a big screen in the corner of the room was the team back in the UK, all hooked into the conversation. Much of the meeting was confidential. But the process was fascinating. Hamilton and Rosberg were taken through each dimension of performance: tires, engine, the helmet, whether the drinks provided during qualifying were at the right temperature.

Each observation from the two drivers was then double-checked against the hard data, and possible improvements noted. After the meeting, the next stage of the optimization loop was already underway, with analysts creating new marginal gains. I couldn’t help contemplating the contrast between the spirit of this approach and that of other areas of our world.

The following day I observed the race from the Mercedes garage. Hamilton made a blistering start from second position on the grid and went on to win the race. The points from his victory propelled him to the overall driver’s championship. Rosberg came in second in the overall classification. Mercedes won the constructors championship: the most successful team in F1.

Afterward, champagne bottles were uncorked in the garage as mechanics, engineers, pit-stop operators, and the two drivers finally let their hair down. “I drive the car, but I have an incredible operation behind me,”

Hamilton said. Vowles added: “We will enjoy tonight, but tomorrow we will feed what we learned today into the next stage of the optimization loop.”

Paddy Lowe, the man responsible for the technical operation, looked on from the back of the garage. “F1 is an unusual environment because you have incredibly intelligent people driven by the desire to win,” he said. “The ambition spurs rapid innovation. Things from just two years ago seem antique. Standing still is tantamount to extinction.”

IV

Google had a decision to make. Jamie Divine, then one of the company’s top designers, had come up with a new shade of blue to use on the Google toolbar. He reckoned it would boost the number of click-throughs.

The narrative surrounding the new shade sounded very good. The color was enticing; it meshed with what was known about consumer psychology. Divine, after all, was one of the top designers at the company. But how could Google be sure that he was right?

The conventional way would have been to change the color on the Google toolbar and see what happened. The obvious problem with this approach should, by now, be obvious. Even if clicks increased, Google could not be certain if the increase was caused by the color change or by something else. Perhaps the number of clicks would have gone up *even more* if the color had stayed the same.

And this is why, even as executives were debating Divine’s shade, a product manager decided to conduct a test. He picked a slightly different shade of blue (one with a hint of green) and put it into a contest with the shade selected by Divine. In effect, users clicking on the Google website were randomly assigned to one of the two shades and their behavior monitored. It was an RCT. The result of the experiment was clear: more people clicked through on the blue with a hint of green.

There was no room for spin or bluster of the kind that often accompanies business decisions. There was just a flip of a coin, a random assignment, and a precise measurement.* The fact that Divine’s shade lost

out in this trial didn't mean he was a poor designer. Rather, it showed that his considerable knowledge was insufficient to predict how a tiny alteration in shade would impact consumer behavior. But then nobody could have known that for sure. The world is too complex.

But this was just the start. Google executives realized that the success of the greeny-blue shade was not conclusive. After all, who's to say that this particular shade is better than all other possible shades? Marissa Mayer, of Yahoo!, then a vice president at Google, came up with a more systematic trial. She divided the relevant part of the color spectrum into forty constituent shades and then ran another test.

Users of Google Mail were randomly grouped into forty populations of 2.5 percent and, as they visited the site at different times, were confronted with different shades, and tracked. Google was thus able to determine the optimal shade, not through blue-sky thinking or slick narratives, but through testing. They determined the optimum shade through trial and error.

This approach is now a key part of Google's operation. As of 2010, the company was carrying out 12,000 RCTs every year. This is an astonishing amount of experimentation and it means that Google clocks up thousands of little failures. Each RCT may seem like nitpicking, but the cumulative effect starts to look very different. According to Google UK's managing director, Dan Cobley, the color-switch generated $200 million in additional annual revenue.[*]

Perhaps the company most associated with randomized trials, however, is Capital One, the credit card provider. The business was created by Richard Fairbank and Nigel Morris, two consultants with backgrounds in evidence-based research. They created the company with one objective in mind: to test as widely and as intelligently as possible.

When sending out letters to solicit new clients, for example, they could have gone to a number of different experts who would doubtless have come up with different templates and colors. Should the color be red or blue? Should the font be Times New Roman or Calibri?

Instead of debating the questions, however, Fairbank and Morris tested them. They sent out 50,000 letters to randomly selected households with one color and 50,000 with another color, and then measured the relative

profitability from the resulting groups. Then they tested different fonts, and different wording, and different scripts at their call centers.[9]

Every year since it was founded Capital One has run thousands of similar tests. They have turned the company into a "scientific laboratory where every decision about product design, marketing, channels of communication, credit lines, customer selection, collection policies, and cross-selling decisions could be subjected to systematic testing and using thousands of experiments."[10]

As of 2015, Capital One was valued at around £45 billion.

Jim Manzi, an American entrepreneur and author who helps companies to run randomized trials, estimates that 20 percent of all retail data is now put through his software platform. This hints, more than anything else, at how far the marginal gains approach has traveled in the corporate world. "Businesses now execute more RCTs than all other kinds of institutions combined," he told me. "It is one of the biggest changes in corporate practice for a generation."[11]

Harrah's Casino Group is symbolic of the quiet revolution that has been taking place. The brand, which operates casinos and resorts across America, reportedly has three golden rules for staff: "Don't harass women, don't steal, and you've got to have a control group."

• • •

RCTs, whether in business or beyond, are often very dependent on context. A trial that improves, say, educational outcomes in Kenya has no claim to improve outcomes in London.* This is both the beauty of the social world, and its challenge. We need to run lots of trials, lots of replications, to tease out how far conclusions can be extended from one trial to other contexts. To do this we need to create the capacity for running experiments at scale and at a lower unit cost.

But this doesn't mean that we cannot draw big conclusions from RCTs. Perhaps the most ambitious use of randomized trials in public policy took place in regard to employment policy. In America in the 1980s, how to get people off welfare and into work was one of the most pressing issues of the day. Policy would conventionally have been decided by the top-down

deliberations of presidents and congressmen in collaboration with advisers and pressure groups.

Instead, it was determined by experimentation. As Jim Manzi details in his excellent book *Uncontrolled*, states were given waivers to depart from federal policy on the proviso they used randomized trials to evaluate the changes. The results were dramatic. The trials revealed that financial incentives don't work. Time limits don't work.

The only thing that worked? Mandatory work requirements. This paved the way for Bill Clinton's highly successful workfare program, secured with the backing of a Republican Congress.

V

Marginal gains may seem like an approach that only big corporations, governments, and sports franchises can hope to adopt. After all, running controlled experiments requires expertise and, often, sizable budgets. But a willingness to test assumptions is ultimately about a mindset. It is about intellectual honesty and a readiness to learn when one fails. Seen in this way, it is relevant to any business; in fact to almost any problem.

Take Takeru Kobayashi. At one time, he was an impoverished economics student, struggling to pay the electric bill of the apartment he shared with his girlfriend in Yokkaichi, on the eastern coast of Japan. Then he heard about a televised speed-eating contest in the area that had a first prize of $5,000. He entered the competition, did a bit of serious practice, and won.[12]

Intrigued, he discovered that speed-eating is a globally competitive sport, with serious rewards. This was a possible route out of poverty. So, as documented in the excellent book *Think Like a Freak*, Kobayashi targeted the world's biggest competition—Nathan's Hot Dog Eating Contest, which takes place every July Fourth in Coney Island, New York.

The rules are straightforward: eat as many hot dogs and buns as you can in twelve minutes. You are allowed to drink anything you like, but you are not allowed to vomit significantly (a problem known in the sport as a "reversal of fortune").

Kobayashi approached the contest with a marginal gains mindset. First, instead of eating the hot dog as a whole (as all speed-eating champions had done until that point), he tried breaking it in half. He found that it gave him more options for chewing, and freed his hands to improve loading. It was a marginal gain. Then he experimented with eating the dog and the bread separately rather than at once. He found that the dogs went down super fast, but he still struggled with the chewy, doughy buns.

So he experimented by dipping the buns in water, then in water at different temperatures, then with water sprinkled with vegetable oil, then he videotaped his training sessions, recorded the data on spreadsheets, tracked slightly different strategies (flat out, pacing himself, sprint finishing), tested different ways of chewing, swallowing, and various "wriggles" that manipulated the space in his stomach in order to avoid vomiting. He tested each small assumption.

When he arrived at Coney Island he was a rank outsider. Nobody gave him a chance. He was slight and short, unlike many of his super-sized competitors. The world record was 25.125 hot dogs in twelve minutes, an astonishing total. Most observers thought this was close to the upper limit for humans. Kobayashi had other ideas. The student smashed the competition to pieces. He ate an eye-watering 50 hot dogs, almost doubling the record. "People think that if you have a huge appetite, then you'll be better at it," he said. "But, actually, it's how you confront the food that is brought to you."

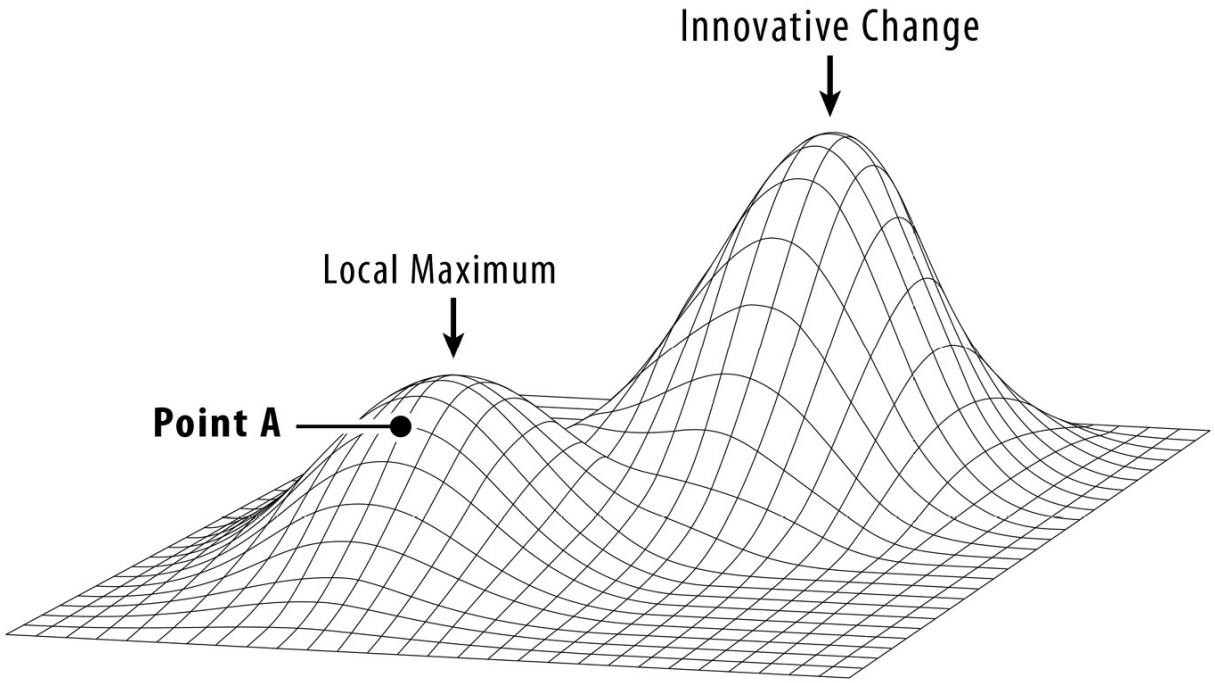
Kobayashi had eaten more than any competitor in history not because he had a surgically enlarged stomach or an extra esophagus (as some competitors alleged); rather, he triumphed via the aggregation of marginal gains. By failing in all sorts of small, well-measured, rigorously tested ways, he iterated his way to success. It was bottom-up rather than top-down, if you'll forgive the expression.

And if this approach can be applied to eating salty tubes of sandwich meat, it can be applied to almost anything.

VI

To conclude this chapter, let's examine the concept of marginal gains in visual form. The process of optimization can be compared to trying to get to the top of a summit. Suppose you start from a position below the summit of the smaller of two hills, Point A, and take a tiny step in a particular direction. You then test to see if you have gone up and, if you have, you take another small step, and test again.

In this way, by taking lots of small steps, each rigorously examined to see if it is taking you in the right direction, you will eventually end up at the smaller summit. Indeed, this method is so powerful that it will work even if you are wearing a blindfold, as the business expert Eric Ries has written in an excellent essay on the art of optimization.[13]



This is the potency of marginal gains. By dividing a big challenge into small parts, you are able to create rigorous tests, and thus deliver incremental improvements. Each may seem small or, as Brailsford often says, "virtually negligible," but over time, and with discipline, they accumulate. You eventually reach the optimum point, the summit of the smaller hill. This is the Local Maximum.[14] It is often the difference

between winning and losing, whether in sports, business, or speed-eating hot dogs.

But this visualization also reveals the inherent limitations of marginal gains. Often in business, technology, and life, progress is not about small, well-delivered steps, but creative leaps. It is about acts of imagination that can transform the entire landscape of a problem. Indeed, these are sometimes the most important drivers of change in the modern world.

To see this difference, take Blockbuster. This was a business based around the renting of videos and later DVDs. As a concept it fared well for more than two decades, delivering an impressive rate of return. You can imagine a manager at the company using a marginal gains approach: altering the company's logo, tweaking the design of the shelving at the stores, trialing different discount approaches like two-for-one, and so on.

Each of these tests would have been useful. Over time they would have accumulated, taking the company toward the top of the local optimization summit. But the problem is also obvious: the business model was eventually superseded by Netflix and the like, rendering videos and DVDs, to a large extent, obsolete.* The entire landscape fundamentally changed. And no amount of marginal gains (at least within a realistic time frame) would have helped Blockbuster to survive. The company was liquidated in 2013.*

In the diagram, the new landscape is represented by the taller hill. Marginal gains is a strategy of local optimization: it takes you to the summit of the first hill. But once you are there, taking little steps, however well tested, runs out of traction. To have stayed ahead of the competition, Blockbuster would have needed to move into an entirely new space, leveraging new technology and fresh insights.

There is an ongoing debate in the political, scientific, and business worlds about whether to focus on the bold leaps that lead to new conceptual terrain, or on the marginal gains that help to optimize one's existing fundamental assumptions. Is it about testing small assumptions or big ones; is it about transforming the world or tweaking it; is it about considering the big picture (the so-called gestalt) or the fine detail (the margins)?

The simple answer, however, is that it has to be both. At the level of the system and, increasingly, at the level of the organization, success is about

developing the capacity to think big and small, to be both imaginative and disciplined, to immerse oneself in the minutiae of a problem and to stand beyond it in order to glimpse the wider vista.

In this chapter we have looked at small steps and found that they are driven by discovering little failures. Marginal gains, as a philosophy, absolutely depends on the ability to detect and learn from small, often latent weaknesses. Now we are going to look at giant leaps, the audacious changes in technology, design, and science that transform our world.

And we will see that beneath the inspirational stories told about these shifts, the deepest and most overlooked truth is that innovation cannot happen without failure. Indeed, the aversion to failure is the single largest obstacle to creative change, not just in business but beyond.