1

Dupers and False Positives

On September 14, 1986, First Lady Nancy Reagan appeared on national television to address the nation from the West Sitting Hall of the White House. She sat on a sofa next to her husband, President Ronald Reagan, and gazed into the camera. “Today there’s a drug and alcohol abuse epidemic in this country and no one is safe from it,” she said. “Not you, not me, and certainly not our children.”[1]

This broadcast was the culmination of all the traveling the First Lady had done over the preceding five years to raise awareness among American youth about the dangers of drug use. She had become the public face of the preventative side of President Reagan’s War on Drugs, and her message hinged on a catchphrase that millions of people still remember, which she employed once again that evening on television. “Not long ago, in Oakland, California,” Nancy Reagan told viewers, “I was asked by a group of children what to do if they were offered drugs. And I answered, ‘Just say no.’”

Although there are different accounts of where this infamous slogan originated—with an academic study, an advertising agency, or the First Lady herself—its “stickiness,” to use the parlance of marketing, was undeniable. The phrase appeared on billboards, in pop songs, and on television shows; school clubs took it as a name. And in the popular imagination it became inseparable from what government and law enforcement officials saw as the crown jewel of the Reagan-era drug prevention campaign: Drug Abuse Resistance Education, or D.A.R.E.

In 1983, Los Angeles chief of police Daryl Gates announced a shift in his department’s approach to the War on Drugs: instead of busting kids in possession of illegal substances, the new focus would be on preventing those drugs from getting into their hands in the first place.[2] This was how

D.A.R.E., with its iconic logo of red letters set against a black background, was born.

D.A.R.E. was an educational program built on a theory from psychology called *social inoculation,* which took from epidemiology the concept of vaccination—administering a small dose of an infectious agent to induce immunity—and applied it to human behavior. The approach of the program was to bring uniformed officers into schools, where they would use role-playing and other educational techniques to inoculate kids against the temptations of drugs. It certainly sounded like a great idea, and the early research on D.A.R.E. was encouraging. As a result, the government opened its taxpayer-funded faucet, and soon the program was scaled up in middle schools and high schools across the country. Over the next twenty-four years, 43 million children from over forty countries would graduate from D.A.R.E.[3]

There was only one problem: D.A.R.E. didn't actually work.

In the decades since Nancy Reagan urged the nation's youth to "just say no" to drugs, numerous studies have demonstrated that D.A.R.E. did not in fact persuade kids to just say no.[4, 5] It provided children with a great deal of information about drugs such as marijuana and alcohol, but it failed to produce statistically significant reductions in drug use when these same kids were presented with opportunities to use them. One study even found that the program *spurred* participants' curiosity about drugs and increased the likelihood of experimentation.[6]

It is hard to overstate the cost of D.A.R.E.'s voltage drop at scale. For years, the program consumed the time and effort of thousands of teachers and law enforcement officers who were deeply invested in the well-being of our greatest natural resource: future generations. Yet all of this hard work and time, never mind taxpayer dollars, was wasted on scaling D.A.R.E. because of a fundamentally erroneous premise. Worse, it diverted support and resources away from other initiatives that might have yielded real results. Why D.A.R.E. became the disaster that it did is a textbook example of the first pitfall everyone hoping to scale an idea or enterprise must avoid: a *false positive*.

The Truth About False Positives

A first truth about false positives is that they can be considered as "lies," or "false alarms." At the most basic level, a false positive occurs when you

interpret some piece of evidence or data as proof that something is true when in fact it isn't. For example, when I visited a high-tech plant in China that produced headsets, if a headset working properly got marked as defective due to human error, that was a false positive.[7] When I was called for jury duty, a false positive would have occurred had we determined that an innocent suspect was guilty. False positives also show up in medicine, a phenomenon that gained attention during the Covid-19 pandemic, when some test results for the virus turned out to be unreliable, showing people had contracted the virus when in reality they had not. Unfortunately, false positives are ubiquitous across contexts; consider a 2005 study that found that between 94 and 99 percent of burglar-alarm calls turn out to be false alarms, and that false alarms make up between 10 and 20 percent of all calls to police.[8]

In the case of D.A.R.E., the National Institute of Justice's 1985 assessment involving 1,777 children in Honolulu, Hawaii, found evidence "favoring the program's preventative potential," and a subsequent study conducted soon after in Los Angeles among nearly as many students also concluded that D.A.R.E. led to a reduction in drug experimentation.[9, 10] These purportedly strong results drove schools, police departments, and the federal government to just say yes to expanding D.A.R.E. nationwide. Yet numerous scientific analyses over the following decade examining all of the known studies and data on the program yielded incontrovertible proof that D.A.R.E. *didn't* actually have a meaningful impact. So what happened?[11]

The simple answer is this: it is not uncommon for data to "lie." In the Honolulu study, for example, the researchers had calculated that there was a 2 percent chance their data would yield a false positive. Unfortunately, subsequent research shows that either they underestimated that probability or they just, unfortunately, fell within that 2 percent. There was never any voltage in D.A.R.E.

How can something like this happen in the hallowed halls of science? First, I should clarify that when I say the data are "lying," what I'm actually referring to is "statistical error." For example, when you draw a sample of children from a certain population (i.e., children living in a single city in Hawaii), random differences among them might produce an "outlier group" that leads you to make a false conclusion. Had the researchers gone back to the original population of children in Honolulu and tested D.A.R.E. again with a new group of students, they would have likely found the program didn't work. (A related kind of inference problem is when the results from

one group don't generalize to another; we take up this issue in Chapter 2.) Unfortunately, statistical failures of this sort happen all the time.

As we saw with D.A.R.E., false positives can be very costly because they lead to misinformed decisions with downstream consequences—time and money that would have been better invested elsewhere. This is especially true when the "lie" or error is missed early on, causing enterprises that were never actually successful to begin with to suffer an inevitable voltage drop at scale. In other words, eventually the truth will come out, as it did for D.A.R.E. when its critics produced overwhelming empirical evidence that the program didn't work. I have witnessed this firsthand in my own work in the business world.

In 2006, the newly christened CEO of Chrysler, Thomas LaSorda, was working tirelessly to try to save his company from bankruptcy. He reached out to me and two of my University of Chicago colleagues, Steve Levitt and Chad Syverson, to see if we had any ideas for how to increase profits, so we invited him and four top Chrysler executives to Chicago to meet. One idea that came out of our sit-down was to implement a wellness program. This might seem like an odd recommendation for a struggling auto manufacturer striving for a bigger slice of the pie in the marketplace, but improving the bottom line isn't always about sales.

Chrysler had an absenteeism problem. As a result, they were hemorrhaging money maintaining a "bullpen" of available workers to take the place on the assembly line of employees who called in sick on a given day. This might seem like a trivial expense for one of the world's biggest automakers, but in fact, a roughly 10 percent absenteeism rate cost them millions of dollars annually in these bullpen wages. Moreover, when we examined defect rates from Chrysler's Sterling Heights plant, we found that over a three-year period a 5 percent reduction in absenteeism had decreased product defects by around five hundred per month! Beyond the absenteeism challenges, the company also struggled with high healthcare costs and "presenteeism" (employees underperforming because of health conditions). Studies suggested that employee wellness programs could address these intertwined problems, so LaSorda believed—and I hoped—that instituting one grounded in the principles of behavioral economics would help Chrysler, too.[12]

A company called Staywell Health Management, which handled programs and services relating to employee health and productivity at Chrysler, agreed to partner with us on a pilot study: a seven-month intervention we called ANewHealthyLife. Conducted at one of Chrysler's thirty-one factories, the initiative used financial incentives (meaning we paid people) to get employees

to engage in healthy activities. Initial results were promising. People in our wellness program engaged in a greater number of healthy behaviors, had lower medical expenditures, and were absent at work less compared to employees who did not participate in the program. In short, our experiment appeared to have saved Chrysler a lot of money in a fairly short amount of time. LaSorda was impressed enough to commit resources for an expansion of the program across the remaining thirty plants.

Although my team and I were pleased with our results, we were more cautious. During my many years of conducting fieldwork and reviewing the research of others, I had seen my share of false positives. We argued that, since the evidence was just from one sample of workers from one plant, we should run one more pilot study before fully rolling it out. The company agreed, and we tried the same program with a new sample of workers from that same plant. This time the results were less thrilling. The employees who participated in the health initiative had no better outcomes than those who didn't across *all* of the main outcome metrics important to Chrysler: absenteeism, presenteeism, healthcare costs, et cetera. Uh-oh. It seemed the initial results were a statistical blip—a false positive.

In other words, the initial data seemed to be lying.

Just to make sure, we ran the program at two additional plants; twice more, our intervention didn't have an impact. The wellness program was just not as effective as the trial run's data had suggested. LaSorda was understandably disappointed, though not as disappointed as he would have been had Chrysler paid Staywell to scale up our intervention across all thirty-one of the company's plants. Detecting the false positive early on allowed us to pivot to a different wellness program that did end up proving effective.

The whole episode was a powerful reminder that when you draw a sample of people for a study, you must understand that it is just that: a sample. And samples are sometimes not representative of an entire population, meaning that sometimes the results you get from your sample will not be true for the entire population. In this case, it turned out that the employees who participated in the first pilot at Chrysler were not representative of all the employees at that plant, much less at all of the other plants. So while the early data appeared solid, clearly they did not tell the full truth.

The fact that statistical errors—like false positives—can occur even in well-designed research studies can be unsettling, especially for those of us (myself included) who see science as a bastion of truth. Yet remember what Winston Churchill once said about democracy: it's the worst form of government, except for all the others.[13] Similarly, the scientific method is the

"least worst" method we have for testing and refining important ideas. And as we will see later in this chapter, there is a method for avoiding the kinds of landmines in your data that prevent an idea from scaling.

Statistical errors, however, are only one of the reasons false positives can be found across so many domains. Another main culprit is biases hidden in the human mind.

Confirmation Bias, the Bandwagon Effect, and the Winner's Curse

In 1974, the psychologists Daniel Kahneman and Amos Tversky published an academic paper titled "Judgment Under Uncertainty: Heuristics and Biases."[14] If you ever need a counterexample to the argument that good ideas will fail to catch on without good branding, this is it. In spite of the article's unsexy title—par for the course in academia—Kahneman and Tversky essentially launched a new field with its publication: the study of cognitive biases. With a series of ingenious experiments, they uncovered a constellation of hidden weaknesses in human judgment that steer us away from rational decision-making.

Cognitive biases are distinct from computational errors and other errors that result from misinformation. Mistakes that people make due to misinformation can be corrected by simply providing more accurate information, but cognitive biases are "hardwired" in the brain, which makes them difficult to change and impervious to correction, since the mind's faulty interpretation of accurate information is precisely the problem. Kahneman and Tversky's landmark collaboration has since been chronicled in multiple books—for instance, Kahneman's *Thinking, Fast and Slow,* Dan Ariely's *Predictably Irrational,* and Michael Lewis's *The Undoing Project*—and some of the cognitive biases they studied have even found their way into the cultural lexicon.[15, 16, 17] One of these is called *confirmation bias,* and it helps to explain why innocent yet avoidable false positives frequently occur.

In the most basic sense, confirmation bias prevents us from seeing possibilities that might challenge our assumptions, and it leads us to gather, interpret, and recall information in a way that conforms to our previously existing beliefs.[18] The reason we have this trapdoor in our thinking is that when individuals are presented with information, their brains are already filled with vast quantities of previously obtained information, social context, and history that project meaning onto the new information. Because we have limited brainpower to process all of this, we use mental shortcuts to make

quick, often gut-level decisions. One such mental shortcut is to essentially filter out or ignore the information that is inconsistent with our expectations or assumptions. This is because science has taught us that reconciling new, contradictory information requires more mental energy than processing new information that is consistent with what's already in our heads, and our brains prefer the easier route.[19]

This tendency might appear counter to our own interests, but in the context of our species's long-ago Darwinian history, confirmation bias makes perfect sense. Our brain evolved to reduce uncertainty and streamline our responses. For our ancestors, a shadow might have meant a predator, so if they assumed it was one and started running, this assumption could have saved their lives. If they stopped to gather more information and really think about it, they might have ended up as dinner.

While confirmation bias was useful for our species in the distant past and continues to be helpful in certain scenarios, for endeavors that require deep analysis and slow deliberation—like testing an innovative idea we hope to scale—it can be troublesome. It can hamper creativity and critical thinking, which are the pillars of both innovation and high-quality work. It can cause doctors to make lazy diagnoses and pursue the wrong treatment. It can drive policymakers, business leaders, administrators, and investors to pour massive amounts of resources into the wrong initiative or venture. And when it comes to interpreting information, whether in business or science, it can produce false positives.

The British psychologist Peter Wason's classic Rule Discovery Test from the 1960s illustrates confirmation bias in action.[20] He gave subjects three numbers and asked them to come up with the rule that applied to the selection of those numbers. Given the sequence 2, 4, 6, for example, they typically formed a hypothesis that the rule was even numbers. Then the participants would present other sequences of even numbers, and researchers would tell them whether or not those numbers conformed to the rule. Through this process, participants were tasked with determining whether their hypothesis was correct. After several correct tries the participants believed that they had discovered the rule. But in fact they hadn't, because the rule was much simpler: increasing numbers.

The most interesting aspect of this study (and the many others like it) is that almost all subjects only tested number sequences that conformed to their personal hypothesis, and very few tried a number sequence that might *disprove* their hypothesis. Wason's experiment demonstrated that most

people, regardless of intelligence, fail to examine their hypotheses critically. Instead, they try only to confirm them by "fast thinking," using quick heuristics, or mental shortcuts.

Another mental shortcut that has a knack for producing false positives is *bandwagon bias*. Also known as "herding" or "cascades," the bandwagon effect arises from social influences on our mental processes. Like confirmation biases, bandwagon bias interferes with our ability to accurately recall and evaluate information. But in this case, we are under the unconscious sway of the views and behaviors of others—the social side of decision-making. In 1951, the pioneering social psychologist Solomon Asch developed a now-famous laboratory experiment that can help us understand this kind of groupthink.[21] He recruited students to participate in what they thought was a vision experiment. These students joined several other supposed participants—who were in fact experimental confederates, or scientists masquerading as participants—in a classroom.

Everyone in the room was shown a picture with three lines of varying lengths, in which one line was very obviously longer than others. Each person in the room was asked to state out loud which was the longest line. After the early confederates all identified the wrong line, more than a third of the participants, on average, went along with the clearly incorrect answer; over the course of twelve trials a whopping 75 percent went along with the obviously wrong answer in at least one instance. In contrast, when no confederates were present to tempt them to hop on their bandwagon, virtually all subjects chose the correct answer—demonstrating just how easily our independent judgment can be subsumed by our desire to "fit in" or be "one of the pack." Not only is this a disturbing blow to one's self-image as a freethinking individual, it also has unsettling implications for the science of scaling.

If you look at the bandwagon effect from the perspective of marketers, whose mandate is to create demand for products at scale, this quirk of the human mind is a godsend: the desire to conform that drives so many of our thoughts and actions can be converted into dollar signs. Indeed, there are mountains of research showing how the bandwagon effect shapes consumer choices, such as the clothing we buy (ever wonder why different colors and styles come into fashion every year?), the toys children ask their parents for (remember Tickle Me Elmo? For your sake I hope you don't), and the sports teams we root for and whose apparel we purchase (the top-selling basketball jerseys in the United States historically correspond to the star players of

teams that make it to the NBA finals in any given year).[22] The bandwagon effect—or social contagion, as it is sometimes called—can even influence our political leanings, and thus electoral results. While this is all well and good for marketers and strategists hired to nudge people toward certain choices over others, for those creating and launching innovations to benefit society, it can create false positives and lead to the scaling of bad ideas.

Bandwagon bias can lead to voltage drops at scale because it has the potential to relegate the selection of ideas to a few individuals rather than an entire team of freethinkers. In my own experiences, whether at a PTA meeting, a meeting in the West Wing, or a meeting in a boardroom, the exact same song and dance play out nearly every time. The leader, who usually is the most passionate, strong-willed person present, tends to talk first and loudest, setting the agenda, dominating the subsequent conversation, and influencing everyone else's opinions and decisions, either implicitly or explicitly. After all, we tend to want to align ourselves with the person who makes promotion and salary-raise decisions, though there doesn't need to be a power dynamic for the bandwagon bias to shape the dialogue—it can kick in anytime anyone makes their preferences known early on. So when an idea or intervention is vocally championed by a trusted or influential source (who isn't necessarily an expert and may have their own agenda), it can result in others jumping on board. In cases like these, what seems like an honest consensus is in fact a false positive, in that any number of people in the room might have opposed the idea were it not for their desire to conform with views expressed by others. The next thing you know, the group has voted to scale a bad idea, or to scale too soon, before additional research can be done. All of which could have been avoided had the leader instead listened carefully to others, understood that implicitly forcing a consensus is not the same as building a consensus, and been wary of bandwagon bias.

Recall how, in the case of D.A.R.E., the National Institute of Justice's favorable assessment in 1986 of the program's efficacy turned out to be a false positive, but this wasn't widely accepted until long after the program had been scaled nationwide. By then, D.A.R.E. already had a network of influential figures on its bandwagon: everyone from police chiefs to educators to community leaders to Nancy Reagan (certainly no expert, but trusted by millions of people), whose "Just say no" campaign had essentially gone viral. Moreover, by this point D.A.R.E. had also received generous funding, another form of social signaling that can help a dud of an idea be perceived as a seemingly scalable one.

The point here is that when bad ideas are endorsed by influential people and institutions, they can be contagious. And once a critical mass has climbed aboard that bandwagon, confirmation bias makes it much harder to convince people to change their now entrenched beliefs. This is why it took many years, and many misspent millions, for D.A.R.E. to abandon its flawed approach.

This unfortunate pattern can play out everywhere from education (e.g., the adoption of trendy but ineffective new curricular programs) to medicine (e.g., the widespread acceptance of medical procedures with dubious efficacy) and beyond.[23] In all of these cases, bandwagon bias likely not only resulted in social and economic harm but also hijacked funding from programs or ideas that would have been more beneficial. Generations of people were robbed of ideas that might have worked.

There is one last set of hidden behavioral tics that often sabotage scaling: the *winner's curse* combined with the *sunk cost fallacy* (the latter of which we'll explore in depth in Chapter 7). As background, think of yourself as a private equity firm considering a purchase of a new company. You and several other companies enter into a bidding war; the highest bidder will win and be required to buy the company for that amount. Would you be tempted to bid more than the company is worth to ensure you come out on top?

I have simulated this scenario in a classroom experiment, only instead of a company, students bid on a jar full of coins. After everyone had written down their bids, I announced the winner—let's call him Eli—who had emerged as the high bidder at $25. I said, "Congratulations, Eli, you've just won all the coins in the jar for twenty-five bucks! How do you feel?"

"I feel pretty good," Eli said. Yet after we counted the change and learned that the jar contained just under $10 in coins, Eli no longer felt quite so good.

I have conducted this exercise for over a decade, and every single time students dramatically overpay, just like Eli did. This is because when everyone is guessing the value, some will inevitably guess too low and some too high. Since the highest bidder wins and has to pay up, the "winner" is guaranteed to lose money.

This is a classic example of the winner's curse. It happens all the time in just about every scenario that involves competitive bidding—venture capitalists vying for the chance to invest in a hot new app, Hollywood producers aggressively going after a buzzed-about screenplay, art collectors bidding for a Basquiat painting, you bidding on eBay, and so on. These are just a few examples illustrating a broader phenomenon: anytime there is competition over an asset of uncertain value, the party who wins the auction

(or buys the idea, hires the worker, et cetera) oftentimes pays more than the asset is actually worth.[24]

Seemingly scalable ideas are like that jar of coins. We overpay to get something off the ground, and later, when we find out that the idea really isn't very good, we ignore that truth and continue to scale because of cost bias: we don't want to admit we have just invested in a loser. Throwing good money at a bad endeavor, however, doesn't make a problem go away. It just compounds your losses. When you find yourself in such situations, the only time you should be comfortable being the highest bidder is if you have the secret sauce for making that idea great, or what economists call a "comparative advantage." This could come in the form of a patented technology that is necessary to scale an idea, ownership of a key resource necessary to operate at scale, or even personal expertise that will allow you to scale up that idea more quickly than competitors. Only then can you skirt the winner's curse, because you are winning by winning.

Whether it's due to statistical errors or errors in human judgment, the belief that an idea is more scalable than it actually is will almost always result in overspending and sunk costs. Fortunately, there is a method that can protect against these kinds of false positives. And we owe this method to, of all things, a cup of tea that changed the history of science.

The Revolution of Replication

In the early 1920s, a young, brilliant, and headstrong statistician named Ronald Fisher was working at the Rothamsted Experimental Station, an agricultural research center thirty miles north of London. One of his coworkers there was a brilliant biologist specializing in algae named Muriel Bristol. During afternoon teatime one day, Fisher made Bristol a cup of tea that she promptly refused. When he asked why, she explained that he had poured milk into the cup before the tea, and she liked the flavor better when the milk came second.[25]

Fisher scoffed at the idea that the order in which the tea and milk were poured could make a difference in the flavor. It didn't make sense to him scientifically, since the molecular makeup of the final liquid would end up the same regardless. But Bristol doubled down, claiming that she could tell the difference between a cup of tea with the milk poured first versus one with the milk poured second. One of their colleagues, a chemist named William

Roach, suggested that the three of them conduct an experiment to see who was right. With that, the stage was set for this epic showdown.

They ran the experiment in a very simple way: Bristol tasted eight cups of tea, four with milk poured first and four with tea poured first, in an order unknown to her. Then came the moment of truth. What happened?

She correctly guessed all eight.

Fisher was stunned. Bristol was vindicated. If they had been gunslingers in the Wild West instead of prim scientists in England, she would have blown the smoke from her gun's barrel.

The important scientific breakthrough that came out of this encounter had nothing to do with milk and tea, however (though it would later be shown that milk globules *do* respond differently and produce subtly different flavors depending on whether tea is poured onto milk or the other way around). Its lasting legacy was this: in designing the eight-cups-of-tea test, Fisher realized that there was a real science to scientific experiments, and soon after, he set out to study it. (Well, that's not entirely true. As with any tale worth telling, there was also a love story here—William Roach and Muriel Bristol would end up getting married!)

In 1925, Fisher published the book *Statistical Methods for Research Workers,* and a decade later he published *The Design of Experiments,* both of which were regarded as groundbreaking and both of which became foundational texts. One of the key pillars of experimental design he established was *replication:* the idea that the act of repeating a test would increase the reliability of its results (which is exactly what Bristol did in the "lady tasting tea" experiment, as Fisher named it, by trying not two cups of tea but eight). By creating a larger data set, a researcher reduces the likelihood of variability, or statistical error, in the results. Indeed, galled by losing the tea argument, Fisher determined that they should have had Bristol taste *more* cups of tea to ensure that her correct guesses weren't a fluke. Put another way, he believed that the more the experiment was repeated, the lower the likelihood of ending up with a false positive.

But replication cannot in and of itself protect against false positives. You have to take the concept of replication one step further and seek an *independent* replication of your result. That is, a person or team with no vested interest in your success must test your idea to see if it replicates. This is precisely what *didn't* happen until much too late with the National Institute of Justice's report on D.A.R.E. Yet it is similar to what my team did do with the wellness program we created for Chrysler. If instead of replicating we had scaled a full rollout based on the hidden false positive, the company would

have lost a lot of money, not to mention I would have lost my credibility. In this case, we caught the error without an independent replication, since we were motivated to doubt ourselves—if Chrysler had scaled the program and it failed, it would have hurt our reputations.

In order to replicate effectively, you must conduct the same successful study, program, or product test again with the same type of population, and ideally you must do so three or four times before you can be confident you have the truth. In some cases this might not be feasible, but my overarching point is that in those instances you should tread with caution rather than "move fast and break things." Otherwise you will soon be broke yourself.

This principle holds true beyond the realms of scientific research. We actually use the same basic approach all the time in our daily lives. Consider dating. At a party you briefly meet someone and feel a connection, and you consider going on a date. But you wonder: was the great rapport real, or—come to think of it—did the drinks you had at the party make the conversation seem better than it really was? Maybe the other person's jokes and stories (or your own) will lose sparkle after a few dates. There's one way to find out: only by spending more time with them will you know if you're truly compatible. This goes for trying out new restaurants, new apps, new hobbies, and so on; we know instinctively that one good experience could be a false positive, but three or four good experiences provide reliable data. In other words, replication is baked into human behavior.

Consider another real-life scenario in which a lot more is at stake. You go to the doctor and get a chest X-ray that reveals what appears to be a cancerous lesion in the upper lobe of your left lung. Uh-oh, big trouble. At this point, you can have your lung removed or you can see a pulmonologist, who will offer a second opinion, take a biopsy, and collect additional data to determine whether such a dramatic surgery should be performed. This "replication" might find the lesion to be just a harmless inflammation and the cancer diagnosis a false positive.

When it comes to matters of health, one could argue that getting a third or even fourth opinion might be warranted. If that suggestion sounds outrageous, consider an explosive study conducted by researchers at Johns Hopkins Medicine in 2016, which estimated that more than 250,000 Americans die each year from medical errors, making such errors the third-leading cause of death behind heart disease and cancer![26]

For businesses, false positives are usually easier to detect than in the world of medicine, simply by testing a product or feature out on a sample set of customers. For instance, when Lyft wants to test a new feature, they simply

build it into the app and make it available to users, usually in just two or three markets to start. This gives my team ample data to analyze in order to determine if the new feature is worth scaling or not.

At the intersection of scientific research and public policy, however, garnering such data is trickier due to a lack of funding for independent replications and slow-moving assessments. Several years ago, a small pilot study of an educational enhancement program for children in Florida, Collaborative Strategic Reading, appeared to increase uptake of reading.[27] It was scaled quickly elsewhere, only to fail miserably: in a series of tests in five different districts in Oklahoma and Texas, the program showed no discernible effect on reading and comprehension.[28] A great deal of money would have been saved had they simply tested the program on a few groups of students in Oklahoma and Texas *before* scaling. Sure, it will take longer to get to the point where such programs can be widely implemented, but this might not be a bad thing.

In science, the practice of replication has become a hot-button issue, leading to what the media outlets are fond of calling the "replication crisis." Since the 2010s, researchers have set out to replicate several high-profile experiments—especially in psychology—that made splashes in major newspapers and on television, and found that the results failed to replicate. This emerging pattern even spurred one psychologist to organize an effort to replicate a hundred experiments whose results had appeared in prestigious academic journals.[29] Shockingly—or perhaps not, considering what we have learned about false positives—only thirty-nine of the one hundred experiments produced the results reported in the published study. The replication crisis has, unsurprisingly, led to a credibility crisis, which scientists are working hard to address. I myself am involved in an effort to rectify what is known as the "file drawer problem"—the phenomenon of studies that failed to produce the desired results being relegated to the proverbial file drawer rather than being published.[30] This is a problem, of course, because even such failures—or *especially* such failures—contribute to the ever-growing body of scientific knowledge, from which everyone benefits (that is, everyone except the researchers whose hypotheses fail, which explains why the results stay in the drawer). Replication is the key to making sure we learn from these failures, rather than waste time and resources pursuing ideas on the basis of faulty science.

As we have seen throughout this chapter, false positives are in most cases perfectly innocent, stemming from statistical errors or from the numerous

cognitive biases that shape our perceptions and behaviors. But there is also a darker side.

This is where false positives cross into the realm of duplicity—another enemy of scaling and a problem that blights the noble intentions of science, while also wiping out billions of dollars for businesses and investors.

The Duper Effect

Brian Wansink was a rock star.

To the extent that a behavioral scientist ever attains mainstream cultural relevance, he was like the Mick Jagger of the food psychology world. As the director of Cornell University's prestigious Food and Brand Lab, he commanded attention both inside and outside academia thanks to his groundbreaking research on the intersection of environment, food consumption, and buying patterns. Over the years, he published a string of stunning findings demonstrating, for instance, that grocery shopping while hungry causes you to buy high-calorie items, that eating from a bigger bowl leads you to eat more, and that the classic cookbook *The Joy of Cooking* offered progressively less healthy recipes over the course of its many editions.[31, 32, 33] Discoveries such as these, which garnered giddy coverage in the press, endowed him with authority and influence that stretched from the federal government to the corporate world. He helped the United States government issue new dietary guidelines, he consulted for Google and the American military, and he published several popular books. People in positions of power trusted Wansink, so they invested resources to scale his ideas.

Then came the twist. It turned out that for years, Wansink had been falsifying results. His seemingly supernova career was built on corrupt science.

As of this writing, nineteen of Wansink's studies have been retracted, while dozens more are under review.[34] In 2018, the *Journal of the American Medical Association* retracted no fewer than six of his studies in a single day.[35] Even as Wansink defended his work, more and more evidence disproving the validity of his research continued to mount. Cornell launched an investigation, culminating in a damning press release stating that Wansink had engaged in academic misconduct, including "data falsification, a failure to assure data accuracy and integrity … inappropriate research methods,

[and] failure to obtain necessary research approvals."[36] In 2019, Wansink left Cornell and academia. His star had imploded.

Unfortunately, such behavior is more common than you might think.[37] I wrote about this several years ago in an article titled "Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior." Using a randomized response survey technique (which is used when trying to obtain the truth pertaining to a sensitive issue), colleagues and I asked a thousand academic economists various questions about ethical behaviors. Shockingly, nearly 5 percent of respondents answered affirmatively to the simple question "Have you ever falsified research data?" In other words, 5 percent of respondents were in effect admitting to being what my friend Antonio Gracias, former chairman of the board of Tesla and founder of the billion-dollar investment firm Valor Equity Partners, calls a "duper"—a person who deliberately lies or misrepresents information to produce a false positive or to get what they want.

As an economist who studies incentives, I was interested in the question of what motivates people like Wansink to engage in such behavior, which was at best bad-faith science and at worst outright deceit. What could compel someone to risk their reputation and professional future in this way? The answer: the way the incentives are structured.

In academia, you only climb the ranks, get tapped to run a lab, and secure large grants (which usually increase your salary, and potentially your speaking fee) if you publish your research in marquee journals. And what gets you into those top journals? Fresh, exciting findings, and all the better if they're media-friendly. So when Wansink took shortcuts and even engaged in outright deception, he was in effect responding to the incentives built into academia. Of course, most top-tier scientists who reach Wansink's level of success do so through legitimate research, innovation, and elbow grease, but they're motivated by the same rewards. When you think about it this way, it's easy to see how someone could be tempted to try to give themselves a leg up by employing an approach that shows little respect for science and its ethical code.

The incentives to lie or cheat are even greater in business, especially in our world of inflated IPOs that turn struggling founders into millionaires or even billionaires overnight. A dramatic example is Elizabeth Holmes, one of the most notorious dupers of the twenty-first century. Her company, Theranos, raised over $700 million from investors and was valued at $9 billion thanks to its breakthrough blood-testing technology that was supposedly poised to

revolutionize the practice of medicine around the planet. Of course, we now know that Theranos was in fact a ticking time bomb: its much-ballyhooed technology didn't actually exist.[38] Once it became clear that the portable blood analyzer the company had raised billions of dollars to build was a dud, Holmes began faking it, by using machines made by other companies. Naturally, it was impossible for Theranos to scale its nonexistent technology, leading to one of the most severe and infamous voltage drops in business history!

Holmes's story is fascinating for a variety of reasons. First, it highlights a vulnerability at the heart of many high-risk investment ventures: people are betting on the person with the idea as much as on the idea itself. This was especially true in the case of Theranos's "moonshot" technology that "required" years of "research and development." In the absence of a market-ready product whose success could be measured, Theranos investors relied on *social proof* rather than hard data.

Holmes was smart and charismatic, allowing her to win over investors, who then vouched for her to other investors, leading to a bandwagon effect. Moreover, thanks to confirmation bias, investors with a financial stake in the company likely overlooked signs that ran counter to their expectation that the company would be a massive success. And these weren't people born yesterday—they included business titans like Rupert Murdoch, Carlos Slim, and Larry Ellison.

Could this debacle have been avoided? Or are some people simply bad actors whose duplicitous behavior can't be prevented? The answer comes down to incentives. Thanks to her stock options, Elizabeth Holmes had the most to gain from Theranos's success and would do anything for success. At one point those options were valued at over $4 billion![39] In other words, she had 4 billion incentives to pull off a win with the technology. But when it became increasingly clear that the promise of the company was unattainable—that the basic premise of this revolutionary blood-testing technology remained fictitious—the long-term incentive of financial success gave way to the short-term one of hiding the truth to put off the inevitable disaster. She was highly motivated to keep the plane in the air with new fuel (money) rather than making sure it could actually stay aloft with a real product and business plan. But fuel was not infinite, making a crash unavoidable.

But could it perhaps have been avoided if there had been incentives in place that encouraged other people in the company to notice—and speak up about—signs that something was amiss? Will employees possibly face

repercussions if they surface information that runs counter to the desires of their superiors, as was the case at Theranos for employees who were attacked for raising red flags about problems inside the company, then threatened with lawsuits for speaking out publicly?[40] If so, they may consciously or unconsciously alter their reports to protect themselves. But what if there were rewards, or at least a contractual guarantee barring punishment, for sharing data that challenged the company's assumptions or hopes? Then they would be incentivized to report the truth, which will eventually come to light one way or another. Theranos would have collapsed sooner and less disastrously had the company operated in this fashion.

The more general question of incentives for employees is a crucial issue for scaling up any product and initiative. When I was at Uber, if a manager had an idea that he or she thought had the potential to benefit the company, guess who was in charge of testing the idea? That same manager! Guess how people were promoted at Uber? By putting forth product ideas that were shipped!

Of course, showing data in a way that highlights the strengths of an idea or downplays the weaknesses is much different from knowingly falsifying data, but the solution to both is the same: independent replication, which is just as crucial for businesses as it is for scientists. If someone has an idea, it should be someone *else* with no stake in the idea and who doesn't stand to benefit financially who should test it out, or at least replicate it before it is shipped. Otherwise, incentives potentially conflict with full honesty.

So the best way to protect yourself from dupers is by keeping incentives top of mind. For example, someone thinking about buying a company should consider if the seller wants to stay involved or not. If the seller is planning to keep a stake in the company, they are incentivized to act in the company's best interests. If they don't want to keep a stake when they sell, this may be because they know negative things about the research, product, or market that you don't. Only they know how good the data truly are. I once conducted an experiment simulating this scenario with a group of MBAs playing the role of entrepreneurs with an idea to sell, but only they had the data indicating whether or not the idea was scalable. They were then presented with various deals from which they could choose. In one case the entrepreneur received immediate payment and went on their way. In others the entrepreneur's payment would depend on how well the product actually performed when it went to market (this is a form of "sharing profit flow," in the parlance of economics). As one would expect, the entrepreneurs who knew they had a losing proposition on their hands tended not to choose this strategy. Instead,

they tended to opt for immediate payment, offloaded the idea, and washed their hands of it. They sold their (unscalable) idea at a handsome profit.

A general lesson is that beyond selling an idea, companies should engage in more profit-sharing with their employees, especially in those cases where their employees make key scaling and purchase decisions. Outside of the start-up world, this is rare, but there are many cases where linking compensation to some metric of future performance (for example, buyers at a retail chain or acquisitions editors at publishing houses having their compensation linked to sales of the products or books they choose) more accurately aligns incentives within organizations and will lead to fewer scaling errors.

Second, if you're in a leadership position, you should also incentivize your staff to be "bias disruptors." As we have seen, organizations don't always incentivize employees to speak the truth, and in many cases the person who comes up with an idea is also the sole tester of the idea. This speaks more broadly to the need for every business and organization to have a devil's advocate deputy, team, and/or function built into its structure—in other words, a force that is always pushing for more data, more proof. A truly good, scalable idea will hold up even under the closest scrutiny.

* * *

BY NOW I hope it's clear that the most hazardous obstacle to successful scaling is not ignorance. It is the illusion of knowledge, arising from either misleading data, hidden biases, or outright deception. Luckily, as we've seen, these hurdles are all surmountable. Yet even solid data that have been independently replicated, trustworthy people, and the right incentives do not guarantee that an enterprise or idea will scale. Even if your idea has proven itself among certain groups, it doesn't necessarily mean those results will hold up for the general public. Are those who gain from your idea or product representative of a large enough group to merit scaling? The answer to this question can make or break your enterprise.

3

Is It the Chef or the Ingredients?

When the boyish British celebrity chef Jamie Oliver opened the flagship location of his restaurant Jamie's Italian in the United Kingdom in 2008, it appeared he had the perfect recipe—not just for healthy and scrumptious Italian food at a reasonable price but also for a chain that had the potential to scale up fast.

By this time, Oliver was already well known around the world. After being discovered by a BBC film crew in 1997, the telegenic young sous-chef soon had a hit show of his own, *The Naked Chef,* followed by a bestselling cookbook of the same name ("naked" referred to the stripped-down style of cooking that Oliver preached, not an ill-advised choice regarding his wardrobe). He prized fresh yet inexpensive ingredients and simple recipes over rare delicacies and sophisticated techniques, and he showed people that you didn't have to have extensive training or aspire to culinary greatness in order to cook tasty, nourishing food at home. This approach had mass appeal, winning over viewers of all ages, men and women alike.

More Jamie Oliver TV shows and cookbooks followed, along with the admirable efforts of his foundation, which sought to reduce obesity and food-related diseases by promoting cooking and nutrition education in British schools. He also created a nonprofit training program to turn young people from disadvantaged backgrounds into skilled chefs, which not only put them on a path to a promising career but also injected new diversity into the restaurant industry. Oliver had scaled not only his approach to cooking but also a set of values and a hopeful message—that food could be a pathway toward positive social change. It seemed that the only goal this apple-cheeked kid who had grown up chopping onions and pulling pints in the kitchen of his parents' village pub had left was to conquer the highly competitive restaurant world. And who was to say he couldn't pull it off? If anything, Oliver's career

to this point had demonstrated that he was the kind of person who could build an empire and make it look easy.

So no one was surprised when the first Jamie's Italian, in Oxford, drew large crowds of hungry diners upon opening its doors. It received an admiring review from *The Guardian*.[1] Most importantly, it served high-quality food at prices that didn't break the bank for the average person. It seemed that Oliver's Midas touch would keep on turning food to gold. Yet he had an ambitious expansion plan for Jamie's Italian that would put his business acumen to the test. As the critic for *The Guardian* put it, "If Jamie can replicate this early Oxford form, he will soon be driving a prize herd of recession-proof cash cows across the land." In other words, if he could scale this initial proof of concept, Jamie's Italian could be unstoppable. This was Oliver's ambition, yet as we have already seen, scaling up often reveals the hidden fault lines in an enterprise. So would Jamie's Italian be able to crank up the voltage and live up to the hype? Or would it suffer a crushing voltage drop?

Oliver's new chain was so successful that within a few years it expanded to seventy locations in twenty-seven different markets, many of them abroad.[2] That he was able to scale the enterprise at lightning speed was especially impressive because, as a general rule, a chef's special touch in the kitchen is unscalable. A person's unique magic at any given specialized skill isn't something that can be mass-produced, for one simple reason: that person can't be cloned. To put it in classical economic terms, supply (of singular cooking talent) is limited no matter how high the demand, which makes scaling a losing bet unless a chef is able to sufficiently pass on her magic to protégés or employees—something that is very hard to do. This explains why so few truly exquisite restaurants ever open in more than one location, or if they do, the quality tends to plunge and soon after so does the number of customers. Human talent is difficult, if not impossible, to scale. The restaurant industry has learned this lesson the hard way, which is why so many of the best chefs in the world are wise to measure the success of their enterprises on the basis of reputation and quality rather than scalability.

The chef Ferran Adrià's one-of-a-kind El Bulli in Spain epitomized this kind of success; this bold experiment in molecular gastronomy relied entirely on his unique expertise and gifts.[3] Adrià's creative process was so intense that for many years he closed the restaurant for up to six months at a time to invent groundbreaking new dishes. Not exactly efficient, but he knew his limitations and played them to his advantage. In the final years before the

restaurant closed its doors (by choice) in 2010, it was reported that 1 million people annually tried to make reservations at El Bulli for only 8,000 spots (showing that scarcity and exclusivity sell). Adrià was well aware that if he had tried to scale his innovative menu—with its pioneering white bean foam and deconstructed liquid olives—in kitchens he didn't oversee, he almost surely would have seen a huge drop in voltage and a huge drop in quality. In other words, he understood that he, as the chef, was the secret ingredient in El Bulli's success.

So how did Oliver bypass this obstacle and manage to scale the unscalable?

To begin with, the formula of Jamie's Italian scaled so well from the get-go because Oliver's name and face served as a brand that got people in the door. Fame and brand recognition are extremely scalable if people like and trust you, and Oliver wasn't just a celebrity but a celebrity chef with earned credibility. However, this alone wouldn't have guaranteed that the chain would thrive at scale. Countless restaurants owned by celebrity chefs—and just plain celebrities—have failed even before scaling up. What kept customers coming back and new ones coming into Jamie's Italian was that, like the recipes in his cookbooks, the dishes relied on fresh ingredients simply prepared, rather than on extravagant ingredients or a laborious creative vision. Put simply, Oliver wasn't a Ferran Adrià type of chef. His brilliance in the kitchen wasn't an innovative or technical style; it was his everyman approach. This meant that other chefs could easily replicate his cooking, while also allowing a price point for dishes that kept demand high. In this way he had circumvented the Catch-22 that makes most great chefs unscalable. While Adrià's secret ingredient was the man himself, Oliver's secret ingredient *was* his ingredients. And unlike humans with special talents, ingredients are eminently scalable.

Yet such rapid, large-scale expansion would catch up with Oliver. What he failed to understand was that there were other, less visible ingredients so essential to his empire's recipe for success that if they changed, the whole franchise would fall apart. Which is precisely what happened.

The first was managing director Simon Blagden. A seasoned executive, Blagden had been running business operations for Jamie's Italian since the chain's inception and was almost eerily adept at knowing where, when, and how to open new locations. It was a unique and hard-to-scale skill, one that involved leveraging his years of experience in the restaurant industry to determine exactly whom to choose as franchise partners—ideally, business partners whose values aligned with Oliver's and who would prioritize sourcing good organic ingredients and maintaining high food quality, even if

it meant lower profits. Blagden also oversaw a positive company culture that kept the retention rate of the kitchen staff high, no easy task in an industry in which fast and frequent turnover is the norm. Blagden's business instincts had been honed over time, particularly when it came to hiring and working with the right people. He was indispensable to Oliver's enterprise as it began to scale.

Then there was Oliver himself. Even as the chain grew, he had tried to stay as present as possible in overseeing its operations. His spirit and mission were part of his culinary brand, and this part of him was scalable. Even though Oliver wasn't the one cooking the food that was served under the banner of his name around the world, his impact on each one of his restaurants could be felt. Unfortunately, Oliver underestimated his own importance to his chain.

Things began to unravel in 2017, when Blagden left Jamie's Italian along with two other top executives. In replacing them, Oliver made the mistake of elevating someone new to his company who was woefully unqualified to keep a restaurant chain functioning well at scale: his brother-in-law, Paul Hunt. A former stockbroker, Hunt had been fined for insider trading in 1999, suggesting he wasn't the best fit for the values-oriented culture of Jamie's Italian. More importantly, Hunt hadn't logged the years of experience in the restaurant industry that the job required. Hunt was good at taking an axe to those branches of Oliver's business empire that had run into debt and were hemorrhaging money, but he lacked Blagden's magic touch when it came to the critical task of selecting new sites and partners for expansion. As the future CEO of Jamie's Italian would say in a public postmortem, "We were opening too many restaurants, too quickly, in the wrong places."[4]

Hunt wasn't very good at retaining employees, either, which made him unprepared to maintain the positive company culture at scale. Former employees of Oliver's company would later describe him as a bully and a sexist, which torpedoed company morale. (For the record, Oliver publicly defended his brother-in-law and sought to diminish Hunt's blame.)

To make matters worse, by this time Oliver had overstretched himself and was only able to devote a limited amount of time to Jamie's Italian. His absence, combined with Hunt's limitations, resulted in a leadership void that prevented the company from adapting agilely to a changing market being undercut by app-driven food-delivery services. This is a useful reminder that being successful at scale isn't just about maintaining productivity, distribution, and demand. It's also about being nimble enough to adapt when the landscape starts to change. And if key people are spread too thin—remember, humans don't scale!—an enterprise can become sluggish.

But then the worst thing that can happen to any restaurant—even worse than losing ground to new competitors—happened to Jamie's Italian: the quality of the food tanked. Abominable reviews piled up online ("Absolutely terrible food and service," wrote one customer on Tripadvisor, echoing the sentiments of countless others) and in the press. By the time the influential *Sunday Times* critic complained about Jamie's Italian, writing that the experience made her want to scream and kick things, Jamie Oliver himself probably felt like kicking and screaming, too.[5] At the start of 2019, the chain was $100 million in the red, forcing Oliver to shutter twenty-five restaurants in the United Kingdom and lay off a thousand employees.[6] In 2020, Oliver had no choice but to close still more locations, including ones in Taiwan and Hong Kong.

For a few years, Oliver pulled off what few chefs ever can: success at scale. But in the end, the voltage gains proved unsustainable, and the enterprise ultimately failed.

Negotiables Versus Non-negotiables

The collapse of Jamie's Italian didn't stem from a false positive (taste buds don't lie) or from the company not knowing its audience (both the menu offerings and prices were designed to appeal to middle-class diners). Rather, it occurred because Oliver didn't seem to fully understand the reasons for its ascendancy to begin with. He didn't appreciate the relevant conditions that had led to its success at scale—and which needed to *stay in place* in order to sustain high voltage.

In Chapter 2, we saw that loss of voltage can result from misjudging the degree to which one's audience or customers are representative of the larger population. But anyone wishing to scale an idea or enterprise must also take into account the *representativeness of the situation,* or the circumstances needed to scale. Were Jamie's Italian's situational conditions at scale similar to those that were present at the chain's inception? The answer was yes … until certain indispensable pieces shifted.

For any idea or enterprise—not just restaurant chains—to hold strong at scale, you need to know what the drivers of high performance are and do everything in your power to keep them in place. To achieve this, before anything else you must determine if your secret sauce is the "chef" or the "ingredients." In other words, does your success at small scale rest largely on the people indispensable to your idea or product—say, the engineer who built

the platform your business runs on, or the celebrity spokesperson who fundraises for your nonprofit—or is it the idea or product itself? If it involves people, a key piece to understand is whether those responsible for implementing the idea will be faithful to its ingredients (this is called "political will" in policy circles). Knowing this isn't half the battle. It's the whole battle.

If your answer is the "chef" (people), there will likely be a limit to how big you can get, since, as we've seen, people with unique skills are inherently unscalable. However, that doesn't mean you can't be profitable, provided that you recognize your scaling limitations from the beginning. List Trucking is one of countless examples. My gramps started our family business, and he was a one-man show: one guy, one truck, one good living. My dad took it over and expanded a little, but he still had just a few trucks, with him running things solo. Then my brother took it over, and now it is basically him and one other guy. My gramps, dad, and brother are all great men and hard workers. But guess what? They don't scale, because they are in the chef business rather than the ingredients business. Of course, there are trucking companies that have scaled to hundreds or even thousands of trucks and employees, but for many small companies the personal touch is the key, if an unscalable one. And that's okay. The men in my family accepted that they would run a small firm their whole lives, and they have been happy and successful doing so.

The point is simply that you need to know if your idea is of the mom-and-pop variety or the Uber variety. If yours hinges on a combination of both chef *and* ingredients, like Jamie's Italian, you need to determine the relative importance of each. When it comes to ingredients, you must know your *negotiables* and *non-negotiables,* then figure out whether your non-negotiable ingredients—the ones your enterprise can't survive without—are in fact scalable.

Consider the publisher of this book, Currency, which is an imprint of Penguin Random House. Their distribution network is a key ingredient of their business, and it's scalable, because there are systems and infrastructure in place such that if two hundred new bookstores were to open all across the country, the publisher would be able to ship books to them almost overnight. But another key ingredient is great content, which doesn't scale because there are a limited number of great ideas and writers in the world (just as there are a limited number of world-class chefs). So if they plan to scale further, they first need to identify which of the key ingredients are scalable and which aren't. In this case, if they were to double in size, they wouldn't be able to maintain fidelity for content, yet they could for distribution.

If you slow down and take a look around, you'll see that the idea of negotiables and non-negotiables is everywhere. When you break down the component parts of a car, it's fairly easy to see what's negotiable and what's not. For an automobile to serve its purpose, four tires and an engine are non-negotiables, while a state-of-the-art navigation system and TV screens for the backseat passengers are not. In economics, the practice of valuing such component parts is known as *hedonics.* Making similar hedonic assessments of your idea is a central feature of the science of scaling.

Non-negotiables have infinite value, since your enterprise will not function at scale without them. Negotiables, meanwhile, have finite value. Scaling can succeed only when the non-negotiables remain in place. In general, a growing enterprise will begin to lose voltage once it reaches the point where its non-negotiable ingredients no longer scale.

Keeping the Faith

As the preceding examples showcase, to achieve high voltage as you grow you need to maintain *fidelity* to your scalable, non-negotiable sources of success. In the case of Airbnb, for example, its non-negotiables when it started out were its digital platform and network of hosts, and the company has been able to sustain these elements at scale (though this hasn't prevented them from also launching experiments scaling other services). As you grow, you may also establish new non-negotiables, either in place of or in addition to others. Take Netflix. In its early years, its non-negotiable ingredients included the shipping infrastructure needed to operate its DVD-by-mail service. Today, that infrastructure remains relevant but arguably negotiable, while new non-negotiables include its library of streaming content and the online platform through which that content is delivered.

Staying faithful to your non-negotiables, however, often presents difficulties as you scale. One of the clearest illustrations of this issue is the long-standing challenge for medical practitioners of getting patients to consistently take medications—what is known as medication compliance (or "adherence"). Prescribed medication is a non-negotiable for the effective practice of medicine. Patients need it in order to heal. That fungal rash isn't going away if you don't apply the ointment, and your strep throat won't clear up quickly without antibiotics. Naturally, then, a patient who doesn't consistently take medication as prescribed (with the exception of experimental medicines without established efficacy or drugs with a high risk

of interactions or intolerable side effects) won't receive the needed or even lifesaving benefits.

Yet noncompliance with treatments plagues medical professionals around the world. In fact, doctors have been trying to understand this phenomenon for over a century. You can't scale medical knowledge and advances if the patients they are meant to benefit don't properly administer them. So recalcitrant patients are a consistent obstacle to scaling.

Lack of compliance can present a significant challenge to scaling policies and public goods as well. Think of it this way: a community doesn't benefit from building a new park if no one uses it. Similarly, there are zero benefits to a back-to-work training program that no one enrolls in. These individual examples may seem inconsequential, but imagine five hundred parks that no one uses or five hundred back-to-work training programs with no enrollment. These represent shattering failures at scale.

Compliance—another way of simply saying "use"—is a non-negotiable ingredient for just about every policy, program, or enterprise at scale. Just as focus groups need to be wary of misrepresenting who the population will be at scale, so designers of pilot studies have to look out for whether compliance levels might drop with an expansive rollout. This hidden trap crosses over to the business world as well.

There is a reason there is so much research on compliance in the field of medicine. Sure, the medical establishment cares about the well-being of patients, but because the profit motives for big pharmaceutical companies are immense, those companies have a strong vested interest in getting people to take their meds. It shouldn't be a surprise, then, that many studies in this area are funded by Big Pharma. Insurance companies, too, have a clear profit motive: they want you to take your meds and get healthier so you have fewer claims! The same goes for life insurance: the longer you keep living, the more premium payments accrue to the insurer, and the further out they can push the final settlement payment.

Any business that is built on recurring purchases and consumption runs into this compliance conundrum—that to scale you need people to not only come to what you've built but also actually use it and receive enough of a benefit that they'll use it again and again. Trader Joe's can't get you hooked on one of their dips if you don't try it and like it first. (The same goes for products that are actually physically addictive, like cigarettes.) Similarly, if you don't show up on the first day of the back-to-work training program and find it useful, you're not likely to stick with it, and if you don't take the medication for a few days but still start feeling better, you're less likely to

complete the full course of treatment. So for an idea or product to succeed at scale, it's not enough to simply find your audience; you also have to make sure your audience actually engages in the desired fashion with what you have to offer. This isn't up for negotiation.

Naturally, the question of why people don't take medication that will benefit them is more baffling and psychologically complex than why people don't buy this or that dip at Trader Joe's. The difference is that with most medications, the cost is immediate, while the benefit—feeling better—comes in the future. But with the dip from Trader Joe's, you experience the tasty benefit the moment you ingest the product. Since humans tend to have present bias (we place higher values on immediate things), it's not easy to get people to consistently take a medication if the benefits won't kick in for weeks or months. Yet ultimately the challenge is identical because in both cases (and everything in between) success all comes down to getting the incentives right. With the Chicago Heights Early Childhood Center educational curriculum, I saw firsthand what happens if you fail to do this.

By the third year of our experiment at CHECC, we had designed a program that produced impressive gains for the children. The curriculum itself had two non-negotiable pillars: explicit instruction time devoted to developing both cognitive and noncognitive skills, and the guided involvement of parents. We knew we faced a dual challenge in getting parents first to sign up for Parent Academy and then to comply with its protocols, as a variety of life circumstances made it difficult for many parents in the community to be involved in the way our program called for. So we decided we would offer cash incentives to ensure their participation.

While many of us might have romantic notions about how money shouldn't be necessary to motivate parents to get more involved in their children's schooling, my research had shown financial incentives like these to be a powerful behavioral nudge, especially in underserved communities in which parents are frequently overworked and struggle to afford basic necessities. In Chicago Heights, the payments were essential and we knew this. But we didn't know exactly how essential until a school district in London contacted us to say they had read about the success of our CHECC curriculum and wanted to implement it themselves.

As my colleagues Robert Metcalfe, Sally Sadoff, and I helped the London school officials prepare the curriculum, we discovered a catch: the school district had a policy that prohibited giving parents financial incentives for involvement in their kids' education.[7] The administrators had their reasons for instituting this policy, all of which made sense to me. But this didn't

change the fact that payment to parents was a non-negotiable for scaling our program. We explained this to the school officials, but they insisted on implementing the Chicago Heights curriculum anyway.

Just as I'd feared, parents were loath to sign up—mostly because they couldn't afford to devote the time—and those who did sign up had even worse attendance than what we observe among the most truant students in the lowest-performing school districts in the United States. Unsurprisingly, their kids didn't benefit from the curriculum. How could they? They never received a non-negotiable component of the program. This would be akin to not taking your medication at the height of hay fever season but still expecting your allergies to get better.

The point here is that you don't have full control over how people will engage with your idea or enterprise when it is expanded; you can sometimes incentivize people to comply with the non-negotiable components, but you can't force them. Sometimes this failure to meet a requirement is dramatic, as was the case in London, where removing the financial incentives for parents was like taking the wheels off a car—a fatal hedonic mistake.

In other cases, the lack of fidelity to non-negotiables is more subtle or nuanced. This leads to *program drift* (also called *mission drift*), which is starkly different from noncompliance in that the lack of fidelity stems from the behavior of the idea's implementors rather than the behavior of its users. With program drift, the non-negotiables are not fulfilled at scale either because of organizational constraints that weren't present in the small scale or because implementors either won't or can't replicate the program faithfully. This causes an entirely different program to be provided at scale. Program drift is akin to a restaurant chain putting a lobster dish on the menu at its first few locations, then continuing to offer the item at new locations nationwide—only instead of lobster the dish is made with crab. For a more consequential example of program drift, we can look at the U.S. government's national Head Start program.

Head Start was launched in 1965 as part of Lyndon B. Johnson's Great Society, a sweeping social initiative that included legislation for the War on Poverty. A key component of the program, Head Start sought to mitigate a host of adversities facing low-income communities, everything from early childhood education to health and nutrition. Thirty years later, a new, more narrowly focused program came into being under the Head Start umbrella. Early Head Start grew out of a recent leap in the scientific understanding of child development—specifically, how much development occurs between birth and age three. As a result, its programs were tailored for the unique

needs of infants and toddlers and focused on promoting their physical, cognitive, social, and emotional development through safe and enriching caregiving. It has since grown to be one of the largest federally funded early childhood interventions in the world. Starting with sixty-eight programs in 1995, it has grown to more than 1,200 programs, and as of 2019, 3 million children and families have gone through Early Head Start.[8]

Following the Head Start model already established, Early Head Start aimed to support parents in taking a more active role in their children's educations. One of the linchpins (read: non-negotiables) of Early Head Start is its home visit services. Twice a month, a worker visits each family for about an hour and a half and helps parents find ways to stimulate their child's development. In the years after Early Hard Start launched, the home-visiting program yielded significant gains in school-readiness and strengthened parenting practices. A win all around—at least when done on a small scale. However, when the program scaled, it experienced considerable voltage drops, specifically with home visits for at-risk families.[9] Upon closer scrutiny, it turned out that there was significant variation in the quality of home visits across the country. More people served at scale meant more opportunities for the quality to slip.

The problem was, the more at-risk families Head Start workers visited, the more they were greeted by parents facing many more distractions than the typical parent has to deal with. These parents understandably had much less time to devote to their kids' schooling; they were too busy keeping food on the table and making sure the bills got paid. Basically, life was getting in the way. As a result, the home visits "drifted" away from Early Head Start's method and mission. It was no longer Early Head Start as originally designed and tested. It had morphed into something that sounded the same but was actually quite different and less effective.

The home visit services ceased to maintain fidelity to their non-negotiables at scale because Early Head Start didn't fully understand those non-negotiables—parents having the time and focus to engage with their kids' learning—to begin with. The consequence was a drop in participation, which only further isolated many children from services that might have benefited their development and future lives.[10]

With programs where academic research intersects with governmental/philanthropic funding and nonprofit implementation, program drift is quite common. Often this is due to funding coming from multiple sources, each with their own priorities and agendas—for example, a

foundation whose donors want improved test scores from everyone, a school district that requires no child to be left behind, and university researchers who want publishable results. Each party might impose demands that dilute the non-negotiables. Voltage drops in such situations are especially insidious because properly evaluating programs that are scaled in the public system takes a long time, which allows failure to go unchecked for years. Thus lots of money and human effort are misspent before the problems are identified, understood, and rectified.

The concept of drift isn't exclusive to the slow-moving world of policy interventions and academic research. It plagues businesses as well, especially when the quality of a product fails to maintain the standards of customers' satisfaction. At Jamie's Italian, for instance, the combination of poor management and overstretched resources caused the quality of its food to deteriorate—what the author Paul Midler calls "quality fade."[11] Interestingly, the place where mission drift has drawn the greatest scrutiny is at the intersection of nonprofit and for-profit enterprises.

Over the last several decades, businesses have increasingly reoriented to include social impact in their strategic priorities, while many nonprofit entities have simultaneously had to develop for-profit arms to sustain their operations. What this means in practice is that we now see venture capital firms trying not just to make truckloads of money but also to improve the world while doing so, while nonprofit groups try to support their programs with for-profit businesses—for example, the American Association for Retired Persons (AARP) funds its advocacy with an insurance-based business that generates more than $1 billion annually.[12] Researchers have found that in both social and commercial enterprises, these dual pulls of "serving two masters" often lead to mission drift. When generating both profit *and* social good is non-negotiable, resources are often spread too thin to provide the necessary support to each.

So how can an organization combat noncompliance and drift? Addressing the economic and psychological incentives in each is a meaningful place to start. For better compliance, we need to find ways to make benefits more immediate and visible, while making compliance less costly. For drift, we have found that having people with a personal interest in maintaining the fidelity of an idea—say, the company's founder, or the original scientist who discovered the medical breakthrough—on the implementation team minimizes program drift. Short of that, if the implementors understand the *whys* behind the mission, they will be much more faithful to it.

In the world of twenty-first-century business, however, the loss of fidelity to non-negotiables that leads to the steepest voltage drops at scale frequently has nothing to do with drift. Instead, it often results from the introduction of a wonderful new ingredient that appears to scale beautifully: an innovative new technology.

Smart Technologies, Dumb People

Most digital technologies are inherently scalable. Strings of code are infinitely and instantaneously replicable, and people are going to "comply" with your product because it is precisely what they paid good money for. So if a new technology is the foundation of your enterprise, you might think that your non-negotiables are secure at scale because each copy is identical.

Not so fast.

Recall the utility engagement platform Opower from Chapter 2. In 2010, two years after launching their energy-saving initiative, the company had a new data set that my brilliant postdoctoral student Rob Metcalfe offered to analyze for them with the help of our team. Working with Honeywell, Opower had recently created what they hailed as a game-changing smart thermostat. To conserve energy, it could do things like modify home temperatures when the occupants were at work or asleep and reduce costs by purchasing more power during off-peak hours and less during peak times. It was also linked to an app so that customers could easily adjust the thermostat wherever they found themselves in the house, as long as they had their phone on them.

It had all the makings of a winning product. After the smart thermostat passed every engineering test and performed exceptionally well as a prototype, Opower rolled it out widely in areas of central California. Inexplicably, the anticipated energy savings never materialized.[13] That was when our team (Chris Clapp, Rob Metcalfe, Michael Price, and me) came on board to figure out why.

Looking at the data from nearly two hundred thousand households, we determined that the explanation for this voltage drop was simple: adoption doesn't always equal compliance. Sure, the smart thermostat had been installed in customers' homes and its app downloaded to their phones, but this didn't mean that customers were using it properly. The default settings produced savings gains, but in practice customers gradually undid the default

settings and went back to their habitual usage patterns, which undercut the promised advantages.

The problem was that the engineers hadn't modeled how actual humans behave. In essence, the promising conservation gains they predicted assumed the "perfect" customer who used the technology in optimal ways. In reality, however, their customers were humans, and humans are impulsive, error-prone, and heuristic-driven, have a proclivity for short-term gratification, and aren't always very good at understanding or following instructions. Maybe customers didn't want their dogs to be stuck in a cold house when they were away at work, or perhaps they liked coming home to a warm house instead of one still warming up. Human incompetence, laziness, and wastefulness should not be underestimated—especially at scale! The engineers should have built and tested the thermostat precisely with these tendencies in mind. Without properly testing the product with real users, it would have been impossible to guess the myriad ways in which customers misused the innovation and entirely unraveled the intended gains.

In short, they may have built a smart thermostat, but the humans who used it were certainly not so smart. In fact, Opower's problem wasn't all that different from patients not taking their prescribed medications, since pills and innovative technologies are both perfectly scalable ingredients in theory, but only with proper compliance. My team has since designed fieldwork to show how to incentivize stronger customer engagement with the smart thermostat, and initial results are promising.

Of course, with 20/20 hindsight, the better move for Opower would have been to avoid disappointing performance altogether by anticipating and addressing compliance challenges during the design phase of the device. This takes both imagination and beta-testing with people who aren't tech-savvy engineers. Top tech companies like Apple often do this. Continuing the legacy of Steve Jobs, Apple has a beta program that lets users try out prerelease software to identify glitches. Registered users get early access to new software Apple is about to release and can give their feedback.

In contrast to innovations like the smart thermostat, which had a steeper learning curve, user-friendly technologies are generally much easier to scale right out of the gate. Look at Instagram. It's incredibly straightforward to use. You take a picture and post it, and then others see it. No instructions needed, and even people who don't post pictures can use the app as voyeurs. This simplicity and ease of use are non-negotiables for Instagram, and the app works just as well with the more than one billion people globally who use it today as it did with its first one hundred users.

One last example of a recent technology that scaled exceptionally well comes not from business but from social activism. In the days following the murder of George Floyd by Minneapolis police officer Derek Chauvin in the spring of 2020, before anyone had been charged with a crime, the Grassroots Law Project organized a call drive to flood the phone lines and answering machines of public officials in Minnesota to demand justice. All volunteers had to do was call the number the Grassroots Law Project provided, which played a message coaching them on what to say, then transferred them to the office of a public official. After they spoke to someone or left a message, volunteers simply pressed the star key and would be automatically rerouted to the next official. Volunteers hit the star key again and again and left message after message, making sure their voices were heard. Great innovations like this one add new efficiency to an already existing need. Within days, Derek Chauvin was arrested and charged, with the other officers who had been involved being arrested days later. While numerous factors played into this win for justice, the automated phone drive technology clearly helped. And it will continue to be a valuable tool for activists wishing to reach large numbers of people with their message, because easy-to-use technology scales.

The lesson here is to make compliance easy. It might take more time, money, legwork, and creative energy early on to both design and test a new technology with a wide range of average people. But given that—to invoke Murphy's law—what can go wrong almost always does go wrong at scale, it will be worth it.

To avoid voltage drops that come from circumstances shifting when an enterprise scales up, I advocate flipping the traditional research and product development models. This means you should *start* by imagining what success would look like at scale, applied to the *entire* population with their varying situations, over a long period of time. To accomplish this goal, you need to identify your non-negotiables from the start. For example, if an education project relies on having fifty thousand teachers at scale, you should not cherry-pick the ten best teachers for the pilot study; top-tier teachers will not be representative of the 49,990 additional second- and third-tier teachers you will ultimately have to hire. Similarly, if the likes of Albert Einstein are necessary to make the curriculum work, then develop a technology that can scale Einstein-caliber teaching, because the curriculum won't scale otherwise. For policymakers, this means flipping the script from evidence-based policy to policy-based evidence.

If there are certain organizational constraints at scale—ranging from building size to available technologies to safety concerns to intrusiveness—

test whether those are negotiable or not in the original study. If they are non-negotiables but aren't available at scale, then you don't have a scalable idea. This general approach is called "backward induction," and we will take it up again in greater detail in Chapter 5.

* * *

ONCE WE ACCEPT the realities of what our ideas have to offer, we can more realistically scale them for success. We need to understand our constraints in ecologically valid settings before we can begin to work inside of those constraints. This doesn't mean slowing down your pace. It means figuring out which way to run before sprinting ahead. I have worked hand in hand with a lot of fast runners in the business world, and even the fastest among them never win when running in the wrong direction.

But what do you do when you uphold your non-negotiables, only to discover that scaling up successfully creates unintended consequences that threaten to undo all your hard work?