CREDIT SUISSE

GLOBAL FINANCIAL STRATEGIES
www.credit-suisse.com

Outcome Bias and the Interpreter

How Our Minds Confuse Skill and Luck

October 15, 2013

Authors

Michael J. Mauboussin
michael.mauboussin@credit-suisse.com

Dan Callahan, CFA
daniel.callahan@credit-suisse.com

- Outcome bias occurs when people judge outcomes without considering the true quality of the decision.
- This bias is rife in business, investing, sports, and politics.
- There is a module in the left hemisphere of the brain called "the interpreter" that takes any effect that it sees and effortlessly and rapidly creates a cause to explain it.
- The interpreter is a natural pattern seeker and knows nothing of luck. This leads to numerous poor decisions.
- To manage the interpreter it is useful to know ahead of time how much luck determines results in an activity and to focus on one's decision-making process in instances where luck is ample.

Introduction

Here is one quote from Monday morning's media:

"Each and every week we see bad coaching decisions in the NFL, but never, and I mean never, have I seen one as dumb as the decision Patriots coach Bill Belichick made Sunday night against the Colts.

His brain was more frozen than Ted Williams'."[1]

Numerous press accounts echoed this sentiment. Here was the situation that created the stir. On the evening of November 15, 2009, the New England Patriots were leading the home team, the Indianapolis Colts, by the score of 34-28. There were just over two minutes remaining in the game, the Patriots had the ball on their own 28 yard line, and it was fourth down with 2 yards to go.

Bill Belichick, the head coach of the Patriots, had two choices. He could instruct his team to punt, which would likely give the ball to the talented Colts around their own 35 yard line and force them to march downfield to score the game-winning touchdown in limited time.

Or Belichick could go for it on fourth down. A successful conversion would effectively win the game for the Patriots and hand the Colts their first loss of the season. But a failure to convert would turn the ball over to the Colts at a very attractive position on the field, substantially increasing the probability that the Colts would score and win.

The Patriots had faced a similar situation seven weeks earlier against the Atlanta Falcons. They went for it, got the first down, and clinched the game. On this night, the Patriots went for it again and were ruled down just shy of the marker. Moments later, the Colts scored a touchdown. Final score: Colts 35, Patriots 34.

Belichick was unapologetic as he addressed his team the next day: "My decision there at the end was based on what I felt was best for the football team and our best chance to win. I'm just telling you I did what I thought was best. And it didn't work out." Later, he added, "The sign on the door when you walk out says 'ignore the noise.' That's the most important thing. Ignore the noise. I hope you believe it because it's the truth. I'll always tell you the truth."[2]

Did Belichick make the correct decision? The sports analytics community quickly rendered its verdict, and the consensus was that he did the right thing. Brian Burke at *Advanced NFL Stats* summed it up by saying, "The better decision would be to go for it, and by a good amount." While Burke's analysis relied on numerous assumptions, it would have been hard to assign credible probabilities that would have flipped the odds in favor of punting.[3]

If Belichick indeed made the right decision from a statistical standpoint, why was he excoriated in the press? We argue that the reason is something called outcome bias, which says that "people take outcomes into account in a way that is irrelevant to the true quality of the decision."[4] Outcome bias is rife in the worlds of business, investing, sports, and politics. We will discuss the mental processes that underlie outcome bias and will finish with some ideas about how to cope with it.

That outcome bias is so widespread is not just an interesting psychological observation. It also encourages poor decisions. You might imagine the mental process of a coach who is less secure in his job than Belichick. That coach would find the "safe" but inferior choice to punt the ball preferable precisely because it would shield him from the wrath of the fans, the press, and perhaps even the team's owners.

Every day, people who make good decisions with bad short-term outcomes risk losing their jobs. This might include the head of a studio in Hollywood who failed to deliver a blockbuster, a chief executive officer who made a reasoned investment that soured, or a money manager with poor results for a quarter or two. The career risk in making better but bolder decisions can be too high for many professionals to handle. Most leaders punt.

Outcomes Dictate Assessments of Decision Quality

Outcome bias only arises when certain conditions are in place. Consider two parties, the person making the decision and the person judging the decision. If the judge knows nothing of the information that decision maker had, it is reasonable to equate poor outcomes with poor decisions. The judge has nothing else to go on.

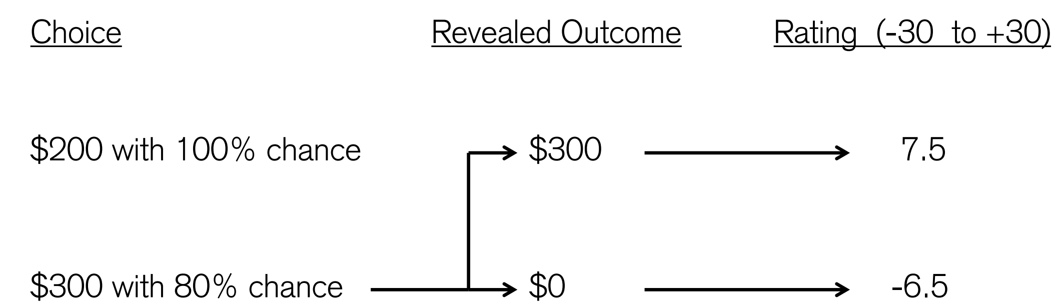
But when the judge has access to the same information as the decision maker and still equates a good outcome with good skill and a bad outcome with bad skill, outcome bias is at work. The bias is especially pronounced for decisions with outcomes that include a healthy dose of luck.

Jonathan Baron and John Hershey, scholars of decision science, ran an experiment to show this bias. In one version, the subjects of the experiment were the judges. They were told that a 25-year-old man won a prize and that the man was unmarried and held a steady job. The prize was a choice between winning $200 for certain or an 80 percent chance of winning $300 and a 20 percent chance of winning nothing. The researchers told the subjects that the man selected the gamble.

The researchers then showed the subjects two different outcomes. In one the man won $300 and in the other he won nothing. They then asked the subjects to rate the quality of the man's decision on a scale from 30 (clearly correct, the opposite decision would be unacceptable) to -30 (incorrect and inexcusable). Since the judges had the same information as the man making the decision, they should have rated the quality based solely on the decision itself, irrespective of the outcome. But that's not what happened.

When the subjects were told that the man had won the money, they rated the quality of his decision a 7.5. When the researchers told the subjects that the man had earned nothing, they rated his decision a -6.5. (See Exhibit 1.) These ratings are clear evidence that the outcomes deeply influenced how the subjects assessed the decision. Somehow, the subjects didn't distinguish between two independent issues: the quality of the decision and the outcome from the decision.[5]

Exhibit 1: Outcome Bias Revealed in an Experiment



Source: Jonathan Baron and John C. Hershey, "Outcome Bias in Decision Evaluation," Journal of Personality and Social Psychology*, Vol. 54, No. 4, April 1988, 569-579.*

We can examine the furor around Belichick's decision in this light. Even though the statistics on "going for it" were widely available before the decision, most people assessing the quality of the decision did not have the numbers at their fingertips. Had there been no information from any source, it would have been very difficult to judge the quality of the choice objectively. So while the criticism of Belichick's decision appears misplaced, it is at least understandable given that few knew what he did at the moment he decided.

Baron and Hershey's experiment is even more powerful, as it suggests that even those who have that information struggle to disentangle it from the outcome. How could judges place a different rating on the identical decision given only a different outcome? We'll now explore why it's so hard to distinguish between the quality of the process of making decisions and the outcomes that appear.

A Look Under the Hood: How the Interpreter Rules the Day

Michael Gazzaniga is a neuroscientist at the University of California, Santa Barbara, who is best known for his fascinating study of split-brain patients. These are people who at one point suffered from debilitating epileptic seizures. A doctor observed that one of his patients got relief after he developed a tumor on his corpus callosum, the bundle of nerves connecting the left and right hemispheres of the brain.

So for the first time in 1940, doctors started treating these patients by surgically severing the corpus callosum, effectively stunting communication between the brain's hemispheres. The treatment was very successful as the number of seizures decreased sharply and the patients all reported feeling fine after the operation. But for our purpose, the fascinating part of this procedure is that it created an extraordinary experimental condition that allowed neuroscientists to determine what happens in each of the hemispheres.[6]

One of the most remarkable findings from this research is that there is a module in the left hemisphere that Gazzaniga calls "the interpreter." This part of the brain takes any effect that it sees and creates, effortlessly and rapidly, a cause to explain it. The interpreter "continually explains the world using inputs that it has from the current cognitive state and cues from the surroundings" and doesn't consider the source or completeness of the information. Further, the interpreter is "driven to infer cause and effect." It knows nothing of probability, randomness, or luck.[7]

Gazzaniga figured out the location and function of the interpreter through experiments with split-brain patients. Here's an example. The scientists showed the left hemisphere a picture of a bird's foot and the right hemisphere a picture of a snowy scene. Because the corpus callosum is cut, neither hemisphere has access to the image flashed to the other one.

The researchers then asked the patient to point to an object that was consistent with the picture they saw. The hand affiliated with the left hemisphere selected a chicken, a match with the bird's foot, and the hand affiliated with the right hemisphere picked a shovel—again, the correct choice given the image of the snowy scene. The patients made these accurate choices with ease. But the left hemisphere, where the capacity for language resides, now confronted discordant information. It could observe the hand pointing to the shovel, but had no access to the image of the snowy scene.

The researcher asked the patient why he was pointing at the shovel. The correct answer would have been, "I don't know." But the interpreter didn't miss a beat, informing the scientist that he needed a shovel to clean out the chicken shed.[8] Note what's going on here. The interpreter sees an effect (pointing to the shovel) and creates a story (a shovel to clean the shed) to explain the cause. More often than not, the story the interpreter tells is true and follows the facts closely. But the essential point is that facts do not constrain the interpreter. It operates even if it only gathers the gist of a situation.

Gazzaniga and his colleagues ran another test, called the probability guessing experiment, to understand how the left and right hemispheres operate. In this experiment, the subject sees lights that flash either at the top or the bottom of the computer screen. The subject's objective is to guess whether the light will flash above or below the midpoint. The scientists set the flashes so that they appeared at the top 80 percent of the time.

Researchers have done a version of this experiment with pigeons, rats, and children under the age of four. The scientists give the animal and young human subjects rewards for correct guesses. Each of them behaves as a "maximizer." Once they figure out that the flashes are predominantly on the top, they select the top every time and are correct 80 percent of the time. They maximize their reward.

But humans older than four are pattern seekers, and try to determine the order in which the random lights flash. As a result, they answer correctly only about two-thirds of the time. Scientists call those who try to guess the pattern "frequency matchers." While frequency matchers recognize that 80 percent of the flashes are above the midpoint, they are not content to guess the top every time but rather attempt to discern a pattern.

With split-brain patients, Gazzaniga could see where this tendency to search for patterns resides. When the researchers presented the probability guessing experiment to the right hemisphere, they found that it was a maximizer just as the pigeons, rats, and little kids were. But with the same experiment the left hemisphere acted as a frequency matcher. It is the part of your brain that is responsible for finding patterns.

Where does all of this leave us? We know that when we see an outcome and don't know what information the decision maker had, our minds assume that good outcomes are associated with good decisions and bad outcomes are linked to poor skill. Further, we know that there is a part of our brain that does this fluently and automatically. The interpreter will come up with a correct story if it has the proper information. But it is willing to make up a story, imposing order on the world, even if it doesn't have all of the proper information. Researchers have isolated the function of the interpreter in split-brain patients, but it runs in all of our heads every day. And this leads us to make poor decisions ourselves and to craft faulty assessments of the judgments of others.

Luck Throws the Interpreter for a Loop

The interpreter tends to be particularly poorly calibrated when randomness and luck play a large role in determining results. The interpreter sees the effect but has a difficult time seeing a clear-cut cause, so it has to infer some skill or other causal mechanism.

A recent experiment shows just how hard it is for us to accept the role of randomness.[9] The researchers separated their subjects into two groups. Group A had 20 players and group B had 28. In phase I of the experiment, members of group A were asked to call five coin tosses and received €2 for every correct call and nothing for a wrong call. Those in group B simply observed the action.

The researchers then asked six members of group A, representing a range of success rates from all correct (5-of-5) to all incorrect (0-of-5), to leave the room.

In Phase II, the researchers told the members of group B that two players from group A, the individual who got the most calls right and the individual who got the most wrong in Phase I, would engage in an additional round of calling five coin tosses. Group B members were given €10 and were told that the calls of one of the players would determine their payoff—they would receive €4 for a correct call and lose €2 for an incorrect call.

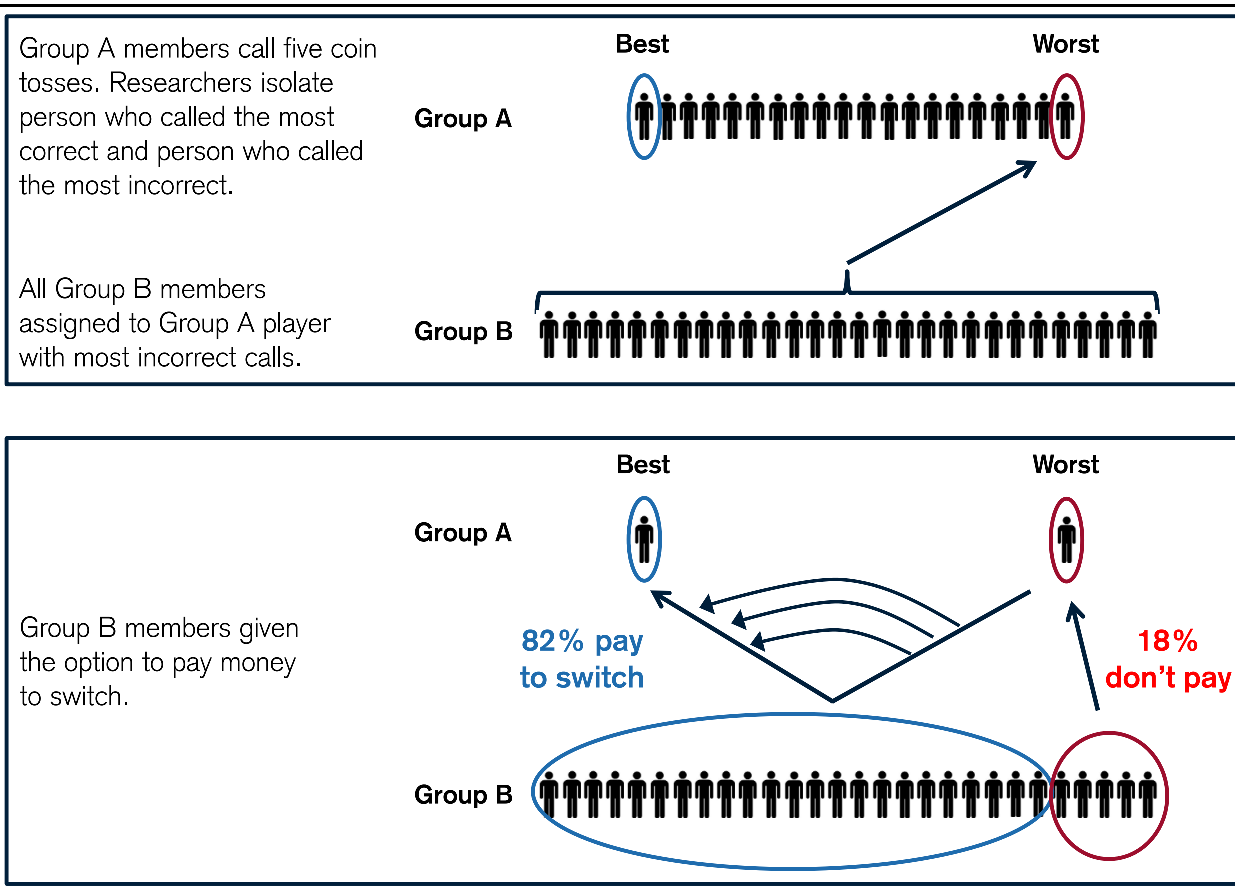
Here was the catch: All players in group B were assigned by default to the player who had the fewest correct calls in Phase I.

The researchers then presented the members of group B with a choice. They could either stick with the default assignment, the bad coin toss caller, or offer to pay some amount of their €10 to switch to the student who called the most right. At this point, the group B players indicated how much they would be willing to pay to switch.

The subjects in this experiment were students of finance who were explicitly taught about randomness. Given that the coin tosses were truly left to chance, the subjects had no basis to switch. This was especially true if they had to pay a price to do so. Yet 82 percent of group B (23 of 28) provided a price representing their willingness to pay to switch. The players in group B who switched stated an average price of almost €3 to switch from a person who called none correct to one who called all five (see Exhibit 2).[10]

We can interpret the result of this experiment as outcome bias. Our minds see an outcome and infer that the person making the decision has good or bad skill. You might be inclined to believe that such behavior wouldn't persist in the business world. But we see versions of it all the time.

Exhibit 2: Subjects Are Willing to Pay for Past Good Luck in a Random Game



Source: Silvia Bou, Jordi Brandts, Magda Cayón, and Pablo Guillén, "The Price of Luck," The University of Sydney Economics Working Paper Series 2013-10, *June 2013.*

Take, for instance, the evidence that individual and institutional investors have a clear tendency to shed funds or asset classes that have done poorly and buy those that have done well. Individuals have such a proclivity to buy what's hot that academics have a name for it: "the dumb money effect." This research shows that individual investors, on average, would have earned one percentage point more in returns if they had simply stayed put with their prior investments versus switching to new ones. Similar to the subjects who were willing to pay to have access to the good coin toss caller, investors are willing to incur costs to switch to hot funds.[11]

Institutional investors, despite their ostensibly greater sophistication, are not immune to the bias. Analysis of asset allocation decisions, both from one asset class to another and from one manager to another within an asset class, show that institutions generally fail to add value. The institutions buy products that underperform the ones they sell. One estimate places the foregone value at $170 billion, a sizeable sum even considering the trillions of assets under management.[12]

Another example is the decision to hire a star from another company. Boris Groysberg, a professor of organizational behavior at Harvard Business School, has examined this topic in detail. One of his studies was of top-rated sell-side analysts, as determined by *Institutional Investor*, who switched from one firm to another. Outcome bias would suggest that the excellent performance of the analysts was perceived to be solely a reflection of their skill. That skill, as a result, should be portable to another company.

Groysberg found that the performance of analysts who switched firms "plunged sharply." Successful analysts benefit from the organization in which they work and likely a large dose of good luck. Neither of those transfers from one firm to the next. This is the hiring version of the "dumb money effect."[13]

How to Keep the Interpreter in Check

Given the prevalence of outcome bias and the poor decisions that it spawns, the question is what to do about it. Here are two steps you can take to help keep the interpreter in check:

- **Measure the amount of luck in the activity.** There are few sure things in life. Most events only happen with some probability. The first step is determining how much room there is for the interpreter to run. Where there is randomness, there is ample opportunity for the interpreter to get into mischief. For realms where causality is hard to pinpoint and experts are poor at predicting outcomes, the interpreter requires close monitoring. In certain fields, there are reasonably precise methods for estimating the relative contributions of skill and luck.[14]

- **Where there is luck, focus on the process.** When skill exclusively determines results, outcomes alone are an accurate measure of skill. Think of the sprinter in a 100-meter dash or a concert pianist on stage. The link from cause to effect is clear, and there is no reason to worry about outcome bias.

  When luck contributes to results, you must consider the process by which the decision was made. While a good process leads to the highest probability of good results over time, the link between process and outcome is loose in the short run. This is the realm where you must diligently fend off outcome bias. The results should not color your assessment of the quality of the decision. Note that if you don't have full access to the decision maker's process, your mind will associate poor results with poor skill.

  A related point is that you should not assume that good results are a reflection of a good process. This is an important topic in ethics. For instance, research shows that "people blame others too harshly for making sensible decisions that have unlucky outcomes," and "we let ethically-questionable decisions slide for a long time until they result in negative outcomes, even in cases in which such outcomes are easily predictable."[15] An example of the latter is auditor independence. A substantial amount of research shows that it is difficult for auditors to remain objective if they are also engaged as a firm's consultant. A handful of high-profile failures, including Enron, WorldCom, and Tyco, revealed the downside of these conflicts. Yet while these companies generated what appeared to be satisfactory outcomes, their processes were rarely questioned.

Our minds are wondrous. The interpreter, in particular, is amazing at associating cause and effect. Indeed, some scientists argue that the ability to infer cause and effect is essential to what distinguishes humans from other species.[16] The interpreter effortlessly takes information that it has and creates a narrative to explain it. In areas where cause and effect are closely linked, there is no problem. But in areas where results are a combination of skill and luck or where complexity hides causality, the interpreter blunders badly. This can lead to poor decisions.

Michael Gazzaniga's pathbreaking research on split-brain patients allowed for crucial insight into how the mind works. A key lesson is that the interpreter tirelessly spins its tales in all of our heads. Knowing when the interpreter is likely to stumble is very valuable in improving decision making.

Endnotes:

[1] Pete Prisco, "Colts Make Pats Pay for Bill's Unusually Dumb Decision," *CBSSports.com*, November 16, 2009. For those unfamiliar with the story behind the final line, Ted Williams was a famous baseball player whose head was cryogenically frozen after he died in the hope that a technology in the future would be able to bring him back to life.
[2] "A Football Life: Bill Belichick," *NFL Films*, (NFL Network: August 29, 2011).
[3] Brian Burke, "Belichick's 4th Down Decision vs the Colts," *Advanced NFL Stats*, November 16, 2009. See http://www.advancednflstats.com/2009/11/belichicks-4th-down-decision-vs-colts.html.

Burke's calculation is as follows. He assumed that a fourth down conversion in that situation would occur 60 percent of the time and that if the Patriots were to turn the ball over, the opposing team would score 53 percent of the time. So the win probability would be:

(0.60 * 1) + (0.40 * (1-0.53)) = 0.79 WP

A punt from that spot on the field typically places the opponent on their own 34 yard line, and teams score on 30 percent of their chances within that amount of time. So the Patriots win probability for punting would be:

(1-0.30) *1 = 0.70 WP

These underlying probabilities assume league averages. Arguably, the Patriots offense had a probability of converting that was higher than 60 percent—the team had 477 yards in total offense that game—and the Colts offense perhaps had a higher probability of scoring after taking over on downs. But even if you stretch the assumptions in favor of the Colts, it's hard to make an analytical case for punting versus going for it.
[4] Jonathan Baron and John C. Hershey, "Outcome Bias in Decision Evaluation," *Journal of Personality and Social Psychology*, Vol. 54, No. 4, April 1988, 569-579.
[5] Ibid.
[6] Michael S. Gazzaniga, *Who's In Charge?: Free Will and the Science of the Brain* (New York: HarperCollins, 2011), 53-54.
[7] Ibid., 86.
[8] Michael S. Gazzaniga, "The Split Brain Revisited," *Scientific American,* July 1998, 50-55.
[9] Silvia Bou, Jordi Brandts, Magda Cayón, and Pablo Guillén, "The Price of Luck," *The University of Sydney Economics Working Paper Series 2013-10*, June 2013.
[10] In Phase III, the researchers determined which members of group B did indeed switch and established payoffs for the additional five rounds. For example, group B subjects assigned to the poor caller bet €2 per round, winning €4 if the caller got a hit and receiving zero for a miss. So no group B player could lose more than the €10 they started with.
[11] Andrea Frazzini and Owen A. Lamont, "Dumb Money: Mutual Fund Flows and the Cross-Section of Stock Returns," *Journal of Financial Economics*, Vol. 88, No. 2, May 2008, 299-322.
[12] Scott D. Stewart, CFA, John J. Neumann, Christopher R. Knittel, and Jeffrey Heisler, CFA, "Absence of Value: An Analysis of Investment Allocation Decisions by Institutional Plan Sponsors," *Financial Analysts Journal*, Vol. 65, No. 6, November/December 2009, 34-51. Also, Amit Goyal and Sunil Wahal, "The Selection and Termination of Investment Management Firms by Plan Sponsors," *Journal of Finance*, Vol. 63, No. 4, August 2008, 1805-1847; and Jeffrey Heisler, Christopher R. Kittel, John J. Neuman, and Scott D. Stewart, "Why Do Institutional Plan Sponsors Hire and Fire Their Investment Managers?" *Journal of Business and Economic Studies*, Vol. 13, No. 1, Spring 2007, 88-118.
[13] Boris Groysberg, *Chasing Stars: The Myth of Talent and the Portability of Performance* (Princeton, NJ: Princeton University Press, 2010); and Boris Groysberg, Lex Sant, and Robin Abrahams, "When 'Stars' Migrate, Do They Still Perform Like Stars?" *MIT Sloan Management Review*, Vol. 50, No. 1, Fall 2008, 41-46.

[14] Michael J. Mauboussin, *The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing* (Boston, MA: Harvard Business Review Press, 2012), 67-90.
[15] Francesca Gino, Don A. Moore, and Max H. Bazerman, "No Harm, No Foul: The Outcome Bias in Ethical Judgments," *Harvard Business School Working Paper 08-080*, April 2009.
[16] Lewis Wolpert, *Six Impossible Things Before Breakfast: The Evolutionary Origins of Belief* (New York: W.W. Norton, 2007).

Important information