# Michael Mauboussin and Daniel Kahneman
# In Conversation

A FARNAM STREET MEMBERSHIP EXCLUSIVE

The Santa Fe Institute Board of Trustees Chair Michael Mauboussin interviews Nobel Prize winner Daniel Kahneman. The wide-ranging conversation talks about disciplined intuition, causality, base rates, loss aversion and so much more. Don't want to read the transcript? You can watch the talk online.

❖

*Michael Mauboussin: Thank you, John. Good afternoon, everybody. There is nothing much to add, but I did -- Steven Pinker has a blurb on Danny's book, which I will read, and I think it's appropriate. It says, "Daniel Kahneman is among the most influential psychologists in history, and certainly the most important psychologist alive today." I would certainly echo that and I want to thank you for all you've done for our community and I certainly personally feel very honored to be here.*

*What I felt I would do is maybe ask a few questions, and then we'll open it up to broader discussion. Just to open, the topic of the day, of course, is big data, but we know that many decisions are made using what we call expert intuition. Can you share a few thoughts on when intuition is likely to work, when it's likely to fail, perhaps, talk a bit about the work you did with Gary Klein?*

*I also loved just a little bit of the history on the contributions of people like Paul Meehl. I suspect many people know a little bit about this work, and even Orley ... Speaking of Princeton, even Orley Ashenfelter and his work on wine, and related to that...why are we hostile to algorithms? Why are these things fundamentally uncomfortable for us?*

**Daniel Kahneman:** That's one question?

*Yeah, it's kind of three things in one, but we'll start with expert intuition. When does it work?*

Well first of all, it works less often than we think, and it works less often than experts think. We did a study of that a few years ago, I mean, there is no question that there is such a thing as professional expertise. All you have to do is look at chess masters, they have it, and it's not only professional. All of us have expertise in many domains ... We can recognize the mood of a spouse or a friend from one word on the telephone. That's high-level skill and those are intuitions and they're quite reliable, quite robust, quite valid.

How we know how those intuitions develop? In a sense, they develop from big data. That is, they develop with a lot of experience. This is how chess players become masters. They develop with a lot of experience and there is one additional essential requirement that applies much more to people and not to big data of the kind that we saw today, and this is the immediacy of feedback. One of the things that I was struck by today is that when you're dealing with big data, you can look four months ahead, but the way that the human learning machine works is not like that. We really depend on immediate data, especially for the kind of tacit learning that is the basis of expertise.

"We really depend on immediate data, especially for the kind of tacit learning that is the basis of expertise."

That's one aspect of it, and the conclusion that Gary Klein - I worked with Gary Klein for many years on a joint article on that - and I'm smiling because he is really an adversary of mine, intellectually. I don't know if you've seen his books. What is, The Power of...?

*The Sources of Power.*

The Sources of Power is a very eloquent book on expert intuition with magnificent examples, and so he is really quite hostile to my point of view, basically. We spent years working on that, on the question of when can intuitions be trusted? What's the boundary between trustworthy and untrustworthy intuitions? I would summarize the answer as saying there is one thing you should not do. People's confidence in their intuition is not a good guide to their validity. Confidence is something else entirely, and maybe we can talk about confidence separately later, but confidence is not it.

What there is, if you want to know whether you can trust intuition, it really is like deciding on a painting, whether it's genuine or not. You can look at the painting all you want, but asking about the provenance is usually the best guide about whether a painting is genuine or not.

Similarly for expertise and intuition, you have to ask not how happy the individual is with his or her own intuitions, but first of all, you have to ask about the domain. Is the domain one where there is enough regularity to support intuitions?

"Most of the time, we have to rely on intuition because it takes too long to do anything else."

That's true in some medical domains, it certainly is true in chess, it is probably not true in stock picking, and so there are domains in which intuition can develop and others in which it cannot.

Then you have to ask whether, if it's a good domain, one in which there are regularities that can be picked up by the limited human learning machine. If there are regularities, did the individual have an opportunity to learn those regularities? That primarily has to do with the quality of the feedback. Those are the questions that I think should be asked, so there is a wide domain where intuitions can be trusted, and they should be trusted, and in a way, we have no option but to trust them because most of the time, we have to rely on intuition because it takes too long to do anything else.

Then there is a wide domain where people have equal confidence but are not to be trusted, and that may be another essential point about expertise.

People typically do not know the limits of their expertise, and that certainly is true in the domain of finances, of financial analysis and financial knowledge. There is no question that people who advise others about finances have expertise about finance that their advisees do not have.

They know how to look at balance sheets, they understand what happens in conversations with analysts. There is a great deal that they know, but they do not really know what is going to happen to a particular stock next year.

They don't know that, that is one of the typical things about expert intuition in that we know domains where we have it, there are domains where we don't, but we feel the same confidence and we do not know the limits of our expertise, and that sometimes is quite dangerous.

*Related to that, in the world of business, certainly in the world of investing, I think most people do try to blend some quantitative and qualitative aspects. I think there's a phrase you've used called disciplined intuition, which is a phrase I love, by the way. I wonder if you can talk a bit about disciplined intuition and I think that you have this wonderful story in your book about setting up interview processes for the Israeli Defense Forces many years ago, but it also suggested there might be some structure about* *how we make our decisions. Can you share a little bit about that? This idea of disciplined intuition, even in thinking about people's practical interview processes, and how one might think about that.*

Well, yeah. It's a story I tell in my book and oddly enough, I'm in the process these days, I'm involved in trying to write an article with colleagues for the Harvard Business Review, and I discovered that, actually, what I want to write there is the story of something I did in the Israeli Army, actually 60 years ago, 59 years ago.

I was charged, this was 1955, and I was charged with setting up an interviewing system for the Israeli combat units to actually select who is fit for combat and who was not, and then to allocate people among different combat units, which it turned out we couldn't do.

The experience that I had there, it turns out, was very formative for me. I had read a book, and you mentioned that name, a very famous book by Paul Meehl, which appeared in 1954, which showed that when you pit against each other, human judgment versus very simple models, and he talked of regression models, but today we know that it can be even simpler than regression models. When you pit them against each other so that the human has access to all the data that are used in the model and then some, the model wins.

By now, there are more than 250 studies of this type, which compare the performance of professional experts to the performance of very simple models, and about half of them, the model wins outright, and then the other half it's a draw, which means, again, that the model wins because it's a lot less expensive. There are ... at least, there were not a few years ago when I last looked, there are no confirmed counter-examples. That's a massively important result.

I knew that result when I was doing that work in the Israeli Army, I was 21 years old, and so I instructed the interviewers who had to interview recruits, that they had to break down their judgment. They had to break down their interview into chapters. Each chapter was dedicated to figuring out one score for individuals, and there were six scores that they were to figure out. By asking many objective questions, but doing this one at a time, and independently of each other, so one at a time, one topic at a time, they were not to mix topics. Those were the instructions and we set up a list of questions.

The story that Michael refers to is that these interviewers have had experience with a system, which was much looser. They were basically, they were trying to form an overall global impression of the individual, and assess how good a soldier that individual would be. We knew that that interview had a validity of essentially zero. They knew nothing, although they really liked that way of interviewing.

When I instructed them on this idea, which now I would call disciplined, not intuition. First of all, disciplined interviewing, break it down by topics, they were furious with me, because they liked using their intuition. People do like using their intuition, and one of them, I still vividly remember, he accused me of, “You’re making us into robots,” he claimed.

Then I made a concession, a very grand concession to them. I was 21, they were 19, I mean, we were all kids. The concession was that after you finish interviewing, and generating your six scores, you can close your eyes and give a rating. How good a soldier would that individual be on a scale from one to five. Then we later, a few months later, we had validity data that, as we knew how well those recruits had turned out as soldiers in units, and, in the first place, we were much better than the previous interview. By much better, I mean, we were only poor. It’s not that we were good, we had a correlation of, say, 0.3 roughly with the criteria, which is as good as psychology typically gets, and it’s better than other disciplines, but I won’t elaborate on that.

> “What I mean by disciplined is delayed inution. The problem with our intuitions ... is they come too fast.”

The thing that was a real surprise and an education for me, was that the close your eyes exercise was better than any of the independent grades. In fact, it was as good as the sum of the six, and it added content, there was something independent. The conclusion I draw from that, which is really quite general, is that--that’s the one that I’m trying to write an article about--the conclusion I draw from that is that there is a need for disciplined intuition. What I mean by disciplined is delayed intuition.

The problem with our intuitions, one of the many problems with our intuitions, is they come too fast. We form impressions very, very quickly, and then we tend to confirm them. If you do what we did in that interviewing system without really understanding the theory, because I certainly did not, which is to look at one thing at a time independently and reserve judgment until you have it all, and then you can close your eyes.

What comes to your mind when you close your eyes after that exercise is going to be much more valid than the intuition you might form if you don't go through the disciplined process.

This, I think, is a big deal, and it's not something that is very common. It leads you to a kind of thinking or to a kind of analysis that is fairly systematic, and with a stress on independence, that is, with a stress on assessing the various dimensions of the problems independently of each other to resist and to overcome a problem that otherwise defeats intuition repeatedly, which is called, I call it associative coherence, so it's the halo effect, it's a tendency to form a global impression and to derive the specifics from your global impression instead of going bottom up from the specifics to the global.

"Delaying intution is ... a good idea."

Delaying intuition is, I think, a good idea.

*I did want to pick up, this is actually one of the questions I had written down about associative coherence and then this idea of confidence. Do you have any optimism that the introduction of big data will help alleviate this problem or will this make the problem worse? Compound the problem? Because there are more things I can do to confirm my views, or things will be what they've always been?*

I think that there were many things that impressed me about what went on today. I don't see how it could compound the problem because, as I understand the use of big data they would describe today, you have a huge mass of data, and then you search, and then the data sort of speak for themselves. Now occasionally, you might have an absurd result that you can pick out and reject, but otherwise, you have to be open to surprises, and there will be surprises, and so I don't think that...I don't see that as a major problem.

*Another thread I'd like to pick up on is something that's come up in one of John's questions, as well, about causality and specifically I want to frame this in the context of regression toward the mean, which is obviously an extraordinarily important, yet, I think poorly understood concept. I wonder if you could just give a fairly formal definition of regression to the mean from a statistical point of view, and maybe offer a few common mistakes associated with it, maybe causality being one or you also have another lovely story in your book about the misinterpretation of feedback, might be another concept. Regression to the mean as a broader concept.*

I'm sure how many people here ... Never mind. I suppose most people here know what regression to the mean is, it's a very familiar concept, and yet, it is not fully understood.

One of my favorite examples is that if you raise the question of why do smart women tend to marry husbands who are less smart than they are, it sounds like a good question. It sounds like a topic that is worthy of discussion, but it really isn't because I can reformulate that as being algebraically equivalent to the statement that the distribution of intelligence is essentially the same for men and for women, and the correlation of intelligence of husbands and their wives is less than perfect.

Those two descriptions are identical, I mean, they're exchangeable, they are algebraically equivalent. They don't sound equivalent. When I raised the question of why do smart women et cetera, you were looking for a cause, but actually, the effect, regression to the mean, that husbands of women who are selected for being very smart, are going to be less smart than the women they're married to. That is an effect without a cause. There is no causal explanation. You've got to rid yourself of the idea of causation to understand that result.

This is extraordinarily difficult to do, and by the way, it is not the only example of an effect without a cause. There is an example that I described in my book of a study that was done on the incidence of kidney cancer in counties across the United States. The counties where the incidence of cancer was high during the year that the study was conducted, they were characterized as being mostly rural and mostly Republican. I mean, a few characteristics of that kind, you have located them geographically. They would be in the center and south of the country.

Why? Everybody has a theory, but why? It turns out if I ask the question of what are the counties where the incidence of kidney cancer is particularly small, it's the same answer. It's counties, they're Republican and rural and so on. The reason is that the population is not the same across counties, and rural counties tend to have a smaller population, the samples are smaller, and it's because the samples are smaller, the incidents of high cancer rates and low cancer rates, is higher in small samples. It's an effect without a cause. Thinking causally will give you trouble when you encounter purely statistical irregularities.

“We tend to impose a causal interpretation on anything that we hear.”

The difficulty that we have, that the human mind has with these kinds of things. We tend to impose a causal interpretation on anything that we hear. We are built that way. Sometimes, it causes us to fall into gross misunderstandings of what is actually going on.

*Also following up on that, just to help us to also sharpen our thinking, one of the potential uses of data is to better inform, like what you've called the outside view, you might call it base rates or reference classes. I wonder if you could just spend a moment talking, and you've already talked about these automatic processes. Maybe spend a moment talking explicitly about the inside versus the outside view, and I have a follow-up to that, but maybe that basic description might be very helpful and I'll follow up.*

I will count on the fact that people, most people here have not read what I wrote about this because it's in a chapter that frankly many people who have bought the book never get to.

The chapter is called The Outside View, and it begins with a true story, and it's one of my favorite stories so I'll tell it.

When I was living in Israel, 40-45 years ago, shortly after I began working with Amos Tversky on our joint project on judgment and decision making, I became involved in writing a textbook for high schools on judgment and decision making without mathematics. The idea was to develop sort of critical thinking curriculum.

We worked on that for quite a while, a little over a year, and I had assembled a team of teachers and one of the members of the team or the dean of the school of education, Seymour. One day, I don't know what possessed me, I had been at it for about a year, and we were really doing quite well. I asked a question of the team, when do you think we're going to finish the book? Please, everybody write it on a slip of paper. When do you think that, and I gave it a formal definition so that it would pass what's called a clairvoyance test. When will we hand in a draft for review to the Ministry of Education?

We all wrote it down and then we tabulated the answers, and they were all between 18 months and 30 months, a year and a half to two and a half years, including mine, and including Seymour's. But Seymour was an expert on curriculum development and I had an idea, which ...

And I asked him, “Seymour, do you know about other teams that have tried to do what we are trying to do, not in that topic, but have tried to develop a curriculum where no curriculum existed before?” He said, “Yes.” He could think of several. So that was good.

I said, “Could you visualize those teams that you know about, when they were at approximately the same level of progress that we have achieved.” He said, “Yes.” He could do that. I said, “Well what happened to them?” The first thing he said, and he was really quite shaken when he said that. He said, “Not all of them actually wrote a book.” 40% of them gave up. Then we asked, “And those who did write a book,” he said, “I can’t think of any that took less than seven years. I can’t think of any that went on much beyond ten, so somewhere between seven and ten.”

> “There are two different ways of looking at the problem.”

There are many things that are interesting in this story, I think, but what I drew from it was that there are two very different ways of looking at the problem. One is the kind of thinking that we had done in estimating a year and a half to two and a half years, and that is basically it’s not quite the best case scenario, but it is anchored on a plan, and it’s anchored on a reasonable extrapolation of what you have done so far, and that’s the way we normally would go about answering this problem.

I call that the inside view. It’s not necessarily the insider view, but it’s a view looking at the problem, at the specifics of the problem. Another way of looking at forecasting problems is to abstract completely from the case at hand and to look at the category to what the case belongs, and to ask what are the statistics of the category. Intuition tends to prefer the inside view. The outside view, by and large, will get you at least in the ballpark, which the inside view quite often doesn’t.

Nobody in our team thought that we had a 40% chance of failing, and certainly nobody had thought that it would take seven years. I can finish the story and tell you it took eight years and there was a book, and nobody ever used it, I wasn’t even there when the story was finished.

So that’s a contrast between two ways of looking at forecasting problems or problems in general. One is highly intuitive and it’s a normal way to go, and it’s the causal way to go. The outside view is non-causal, it is a statistical way of looking at the problem, and by and large, I would say almost any problem that you’re looking at, starting from the outside view, will give you an anchor, will give you a ballpark, and you should start from there, and look for reasons for deviating.

Yeah, I forgot to tell you one final part of the story. When I had told … We had heard from Seymour what our future was likely to be, and sort of desperation, I asked in terms of our resources, and how well we're doing, how good we are, how do we compare to those teams you were thinking about? His answer was we were below average, but not that much. That's one part of the story.

I should add, I won't go into this, but for years, this was one of my favorite stories, it still is, but it took me years to discover who the real idiot in the story is, and of course it was I. It wasn't Seymour. I always thought Seymour had a problem that he had the information and he wasn't using it. But I had the information after that, we should have quit that day, and we didn't.

That, again, is a fairly typical thing. So you have the outside view and although, we had been given the outside view, and we believed in it, we didn't act on it. Because the inside view, the feeling that we had that we were doing well and that this was a good team, and that we were going to succeed, that overwhelmed [the new data] … which is particularly ironic because we were writing a textbook that was supposed to rid people of precisely that mistake. Not only did I not see it then, I think, I didn't see it for years. So, there are real hang-ups in this whole domain.

*Mauboussin: Let me follow up on that, I also want to pick up something you had mentioned. I heard you talk about at a conference a few weeks ago. There's one thing, we're optimistic, we tend to be optimistic about things that are important to us, so we want to call that some sort of a bias, and we were at a conference on strategy and you told a story, maybe you could tell it better than I, but basically you went to a company and they were talking about launching new products, and we know the success rate of new products is quite low. We know the success for entrepreneurs is quite low.*

*I guess the story is something along the lines of, you sort of explained this to him and the guy said, "Professor, if we listened to your advice, we never would have tried to launch these new products that have actually been successful for us." At what point do you balance this notion of optimism, which confers perseverance, versus saying, "Guys, we should just not write the textbook. Enough. We know that entrepreneurs fail at a very high rate. Let's just give up." How do we balance? Because societally, we want to people to persevere and try hard and start new things. By the same token, it may not be good for you or us as investors, or what have you.*

“There isn’t all that much risk loving among people in general and among entrepreneurs.”

I think there are lots of data that the main reason that entrepreneurs take risk...they’re not risk taking. There isn’t all that much risk loving among people in general and among entrepreneurs. What there is, is people take risks because they don’t know the odds, and that’s--if you want a theory of this thinking, that would probably be the core of the theory. People take risks because they don’t fully appreciate the risks that they are taking, and that happens a great deal.

To the individual, this optimistic bias and the outside view, the inside view tends to be optimistic, but it’s not the only source of optimism. That is quite costly to the individual on average. Of course, without it, I have a chapter in the book that certainly nobody gets to ... It’s called The Engine of Capitalism, where I talk about optimism as being the engine that keeps the whole thing going. When you look at big successes and work your way back from big successes, somebody made a mistake in doing something that they had no business trying, and it was the spectacular success of that mistake that produced the big--there is a real trade-off. Optimism is a wonderful thing to have. If you have to wish one thing to your children it’s that they be optimistic.

Optimists live longer, among other things, and generally are happier, it’s much better to be an optimist than anything else. They also persevere more. They respond better to challenges. It’s very good to have optimists. On the other hand, I, for example, do not want my financial advisor to be an optimist. I have no need for optimism on his part. There is a real tension between the two, which I’m not proposing to resolve here, but it’s in any one problem that you deal with, that tension should be on your mind.

*Mauboussin: I have another topic I’d like to draw out, which I think is also very much related to this big data, and that is that as we embrace big data, it suggests often in many organizations change. For example, one of the simple themes of Moneyball might be the numbers guys competing with the scouts, and if you embrace the numbers, then the scouts are basically losers, and the numbers guys are the winners. Obviously one of your also major contributions is this concept of loss aversion, and there are going to be losers and winners. How do we think about the notion--Can you speak about the notion of loss aversion and changing organizations, and to what degree people will or will not embrace some of the changes suggested by big data as a consequence of these true senses of loss?*

I think the main contribution that Amos Tversky and I made during the study of decision making is a sort of trivial concept, which is that losses loom larger than gains. When people look at disadvantages and advantages, the disadvantages are more salient and attract more attention, and are weighted more in decision making than the advantages. It's a big difference, it's been measured in many contexts. As a very rough guideline, if you think two to one, you will be fairly close to the mark in many contexts.

We call that loss aversion. In the context of change and of reform in general, I mean, reforming government, reforming the civil service, change in an organization, there is one thing that is guaranteed when you're making a big change. There will be losers and there will be winners. Some people will derive some advantage from the change, other people will derive some disadvantage. You can know ahead of time that the losers, the potential losers, will fight harder than the potential winners, and when you look at that, that's almost invariably the case, in part because the losers know what they're about to lose and the winners are not sure. In part, because losses loom larger than gains.

"People who initiate the reform do not fully appreciate the resistance that they will encounter."

What typically happens in changes and in reforms of many kinds is that the people who initiate the change, the people who initiate the reform do not fully appreciate the resistance that they will encounter. There is a piece of research that is highly relevant to that, but I like a lot. There is an effect in decision making that's called the endowment effect. That is that if ... And it was first studied, I think, by my friend Jack Knetsch in the context of sandwiches. My selling price for a sandwich I own is higher than my buying price for the same sandwich when I do not own it, and it's a very strange result, but that's how he first observed it. Substantially higher.

People tend--for the sandwich it is particularly bizarre because you could get another, but when people own a good and they have to give it up, giving up is more painful than getting something.

But, it turns out that Jack Knetsch did a bit of research. I'm not sure it has been replicated and I'm really almost afraid that it won't hold up because it's such a nice result. When you have an advisor selling a sandwich on behalf of somebody else or buying a sandwich on behalf of somebody else, that effect is gone, there is no loss aversion. That's really important.

Loss aversion is emotional, the reluctance is emotional, and if I'm making a decision on behalf of somebody else, I don't feel that emotion, which means, by the way, that advisors are likely to be more rational in the long run because loss aversion is costly. It may also be true of leaders of organizations. When they make a change, they see the situation after the change, everything will be better than it is now. They do not appreciate the losses, the changes that people will experience, and the fact that the losses are going to be painful, and they end up compensating the losers because, quite often, that's the only way of getting something through. This, I think, is one of the major reasons why reforms and changes are almost invariably more expensive than anticipated. It's because that aspect of the change is rarely anticipated.

*Mauboussin: Do you have any prescriptions as to how to mitigate that effect?*

Take the outside view, that would be my prescription. I mean, if you look at other attempts to do similar things, you will see it there. You will see it there more clearly than you are likely to see it in your own case.

*I want to ask another question about decision making within organizations. I know one idea that has captured your attention a bit is this concept you're calling noise. I think noise has a lot of meanings to different people, but perhaps you could explain when you're thinking about this concept of noise, what that means, why it's interesting to you, and again, what the significance is to organizations and decisions.*

"One thing that has intrigued me most is this concept of noise."

In the last few years, I've been doing consulting and the one thing that has intrigued me most is this concept of noise. I'll spend a few minutes and I'll tell you about it.

We know why Paul Meehl was right when in 1954, he said that formulas, very simple formulas, are better than judges. And we know it from a very interesting study that was done. It was done with clinical psychologists but I'm pretty sure that it will apply more broadly. You have people who are using a profile of information to make predictions about a set of cases. They could predict how much a person will earn or some quantitative criteria.

There are a number of things that you can do, and you know the criteria, you know the outcome that people are trying to predict, so you can look at the accuracy of the individuals, then you can apply the statistical model, you can look at the information, and you can predict the criterion, the statistical model beats the judges, that's not surprising. Now you do something else. For each of these judges, you build a model, and what the model does is it predicts the prediction that the individual will make. This is nothing to do with the outcome. You build a linear model that predicts what the judge will say.

Now, in a holdout sample, in a new sample, you compare the accuracy of the judge to the accuracy of the model of his or her own model. The model of the judge is better than the [actual] judge. If you think about this, in a majority of cases, this is really an important result because it tells you where it comes from, why people are inferior to formulas, and people are inferior to formulas because a formula, when you give it the same input, twice, it will always have the same output. This is not true of people. People vary and they vary over time, so, I mean, one, I think of is x-ray readers, radiologists. When they look at the same x-ray twice, so I read, they reach different conclusions 20% of the time. That's noise. That's what I would call noise.

In many organizations, you have many functionaries that are making decisions on behalf of the organization, and at least in principle, they're interchangeable. An obvious example is the justice system. You have judges and we assume that the same defendant in principle, it ought to be the case that [they are treated the same]. Now this happens in many organizations. For example, you have credit rating agencies. You have different individuals rating. What you would want is you would want all the individuals to be interchangeable. It turns out they're not.

And so, we did research in a financial services company, quite a large one and a well-known one, and we did research on two categories of employees, and I can't tell you more for obvious reasons. You'll see why, actually, in a minute. These are decisions that people make by looking at written material. They pore over material and then they put out quantitative judgment in dollars. You have many of these individuals making--and, in principle, it's random who gets what. We ran an experiment and it was very brave of them to let us run the experiment. We ran an experiment where we presented the same material to 40 people, five different cases. The interesting thing is we asked the executives, suppose we take two individuals at random, how much will they differ in percentages?

It's odd but the answer typically, and you'll feel the same, actually, 5%-10%. You expect that well-trained individuals will come up with the same number. The real answer is 50%--45% to 50%. It's a huge problem, actually, because those errors, that variability, is costly. You can analyze how it is. When error is costly, noise is costly.

"When error is costly, noise is costly."

 

We tend to think of mistakes or errors as bias, that it's a systematic error, but nonsystematic error is costly. For the statisticians here, if you have a square loss function, then noise and bias are actually additive. It's the same.

What was very striking was that here's an organization, it has a big and costly problem, and it doesn't know it. It doesn't know it because it very rarely happens because organizations, unlike Google, most organizations do not experiment, so it had never occurred to them to run that experiment.

They thought that their employees agree with each other, but they don't. Furthermore, experienced employees disagree with each other just as much as novices, so experience does not bring convergence. What it does bring is increased confidence. That's the point I was making earlier.

That, we are now trying to figure out, what are the limits of this problem? Because I think probably there are many organizations that have that kind of problem, as we have as individuals. I'll give you another example just to stretch the thinking about that. You have financial advisors who deal with clients, and they have a list of clients and they have to prioritize the list depending on various characteristics of the clients. Do they do it the same way? Probably not. And if they don't do it the same way, they can't all be right. If there is a solution, you would want them to converge on that solution. Noise is not advantageous in that case. And, I can't think of many examples. There is no evolution when there is no selection, where error and noise are essential. When there is no selection, noise is generally costly.

*We're going to open it up in just a moment. This is a good time to think about your questions. Before I open it up, I do want to ask. Certainly one of the themes of today or big data in general is to think about forecasting. Thinking about the future. You recently warmly endorsed a book by Philip Tetlock called Superforecasting. I just wonder if you could share with us a bit what you found interesting or useful about that book. Yeah, what you found interesting about that book and maybe a quick synopsis of the journey that Phil's been on and where he is today.*

Phil Tetlock is a social scientist, he's a psychologist, he's also a friend of mine, I mean, full disclosure. In 2005, he published a really important book on political and strategic forecasting. In that book, he looked at pundits and CIA analysts and people whose business it is to make, to predict the political and strategic future, and he had them make long-term predictions, and then he waited ten years.

Will there be a regime change? Will there be financial crisis? Will there be a transition to democracy? Whatever.

Basically, the conclusion of the book was that people cannot do it. Another conclusion was the more they think they can do it, the less they can do it. That is, the more overconfident the people were, and the more they have a theory about what's going on, the less they can do it. Those were some very good conclusions, and I drew heavily on that work in my own.

In recent years, Phil has turned to a different project encouraged by IARPA, the Intelligence Advanced Research Projects Agency. There was a contest among academic institutions among groups of academics at different universities, there was a contest for organizing predictions--so, a forecaster tournament of forecasting accuracy where people were to assign probabilities to events. This was short-term, six weeks to a few months, and how do you improve forecasting accuracy? There were many, several groups, and Phil Tetlock and his wife, Barb Mellers, they headed one group, and they won hands down.

Not only did they win overall, I mean, they have, by the way, those were very large number of forecasts. What they did is they advertised for people who were interested in playing that game, and they, I think, in the first year, had 3,000 people who made sort of weekly predictions about assigning probabilities to political events. They worked in teams and they also made individual judgments. At the end of the year, they identified the top 2%, and those later were called the Superforecasters. And they've been following those top 2% of forecasters who are just very good at it.

That includes a wide range of people from—I mean, there are many people with quantitative ability and expertise and so on, but there's, I think, a pharmacist somewhere in Alaska or maybe ... who beats the CIA in terms of the accuracy of her probabilistic predictions. That's interesting to me at multiple levels. The first place because I was very impressed by the first book and so I was skeptical that he would actually find it, and I wasn't very optimistic when Phil started.

The fact that short-term prediction is possible is not revolutionary. I mean, we would expect that long-term prediction still isn't probable. What there is in the book is an analysis of how they do it, what makes somebody into a forecaster, and those ideas are pretty simple, but they're also widely applicable.

We had mentioned several of them, so it's a mixture of the inside view and the outside view. It's clearly disciplined intuition, it is clearly an attempt to make independent judgments and then to collate it. Many of the standard ideas that when you apply them, it turns out it really improves your ability to understand problems in the real world.

I think it has applications. I'm more skeptical, by the way, about the applications to the political strategic world because I'm not sure that the decision makers are equipped to deal with their decisions as gambles, but in the financial industry, as an adjunct or replacement to judgment, where judgment is used and not big data, then it's quite interesting, I think.

*Thank you. With that. We'll open it up. Yeah, just grab a mic. Josh, yeah, grab a microphone and we'll ...*

*Speaker 4: You mentioned the inconsistency both within individuals given the same data point and then, of course, the divergence between individuals within an organization. At what level of aggregation does that dissipate? I'm thinking in particular of Scott Page's work on diversity prediction theorem and some of the prediction market stuff that you touched on with Tetlock.*

*Mauboussin: All right. Does wisdom of crowds help us address some of them?*

There are certain things that are washed out in aggregation, and so, for example. In personnel assessment, which is a topic in which I've been interested recently, and performance assessment in organizations. There was...where ratings are used, there are two ... All of you are familiar with it. There's forced ranking is used by about half of the large companies and ratings and used in other companies. Where people are rated by a single manager, most of the variance, more variance is attributable to differences among managers than to differences among the people that they rate. There are huge differences in how different managers use the same rating scale. That gets washed out in aggregation, so I read recently about how they do things in Google, and there, there are multiple ratings. You would expect that sort of variation to be eliminated.

On the wisdom of the crowd, Superforecasters clearly beat prediction markets. There is the wisdom of the crowd. It's not entirely clear how far wisdom of the crowd--at least to me, I haven't seen compelling evidence that it's far superior to simple averaging of independent opinions.

Superforecasters have been compared to prediction markets, and they're better in that context.

*Mauboussin: They also found Superforecasters who work in teams work better than Superforecasters working by themselves, which is also an interesting finding. Billy, I've got a question here. Yeah.*

*Speaker 5: Hi, Professor Kahneman. I asked you about this during the break, but in biomedical research and psychological research, there's a real crisis in non-replicable research. I just wonder in your field, as well, and I just wonder how do you think that that applies to these larger questions of predictive algorithmic systems or disciplined intuition or whatever? These are structured published research in top journals, right?*

I'm a psychologist and we are in the midst of a sort of crisis on reproducibility of our own results because there was a study published in Science a couple of months ago that really was quite disappointing in terms of replication. This morning, I must say that my key impression was that Hal Varian has the best job in the world. I mean, I just imagine he lives in heaven so far as I'm concerned. Because he doesn't have a problem with reproducibility, he has very, very large samples. The problem of reproducibility is that the samples are too small relative to the size of the effects that are measured.

What happens when samples are small, is that people develop-- Researchers develop very bad habits, and the bad habits are to protect themselves against finding nothing, they try many things, and then they report selectively and they fool themselves, and the result is reproducibility is low, it's quite low in medical research. I don't remember the statistics but they're worse than psychology. There is a real problem. What seemed very clear today was that when you have huge data, that problem, the problem of ... Look, I mean, it's the same as in regression. The problem with regression is not that something happens in the second time that you measure it. The error is in the first time.

It's how stable the results are, how true the results were when you measured them. At least big data won't have that problem. The stability of relationships and other samples in the past, the future will be like the past. That's a problem, but my impression was that at least when the data is huge, and the questions are relatively limited, they can be quite ...

The question can be quite complex. The search universe, relative to this amount of data, it's a very favorable ratio. I was very envious of the guys who do big data.

*Speaker 6: Hi, how are you? I wanted to go back to the point that you made about using the category to shape decision making more than this specific example, certainly from an insider's point of view. How do you think about situations where you may be creating a new category or at least there's not an obvious comparator. For example, to take a very simplistic example from the consumer world, think about pricing the first cup of Starbucks coffee. You wouldn't have gone to the category of 7-Eleven coffee said, "Well, 99 cents seems about right." How do you think about those sorts of situations?*

In that particular example, that's what, I suppose, market research is for. You do research, there are no guarantees that the market research will be accurate, but it would certainly be better than guessing. The question that you're asking about unique events and unusual events, that has been in the background of today. For example, in the discussion this morning about abduction, sort of the new hypothesis, the new idea. Certainly, there are people, and everybody thinks of Steve Jobs in that context always as somebody who had intuitions that turned out right, and the question in my mind following this morning was whether that kind of performance could be duplicated in big data.

I couldn't see any reason why not. Ultimately, I think intuitively, we bridge those gaps and we're not even aware of the fact that we don't have a good comparison case, but ultimately, big data are going, there is a hope or a fear with big data that they'll be going beyond where we go with confidence with our [own] intuition.

*Mauboussin: Maybe I can offer one reason with big data. Maybe he was just lucky.*

Where.

*Mauboussin: Steve Jobs.*

Yeah. I think unquestionably he was lucky and he didn't hit it every time. Yeah. I mean, it's a topic that Michael and I are both very concerned with, which is the role of luck, and we've both emphasized the role of luck. It's very hard for me.

It's interesting because think of what we have been talking about--statistical thinking, causal thinking. It's very difficult when you think of Steve Jobs and you've read about him, and he's in the cinemas, as well. It's difficult not to see a causal system in front of you, so that makes it very difficult to view what happened as an instance of luck. It could be luck.

*Mauboussin: It's like your famous equation, your Brockman equation. Maybe you should share what your--The Brockman equation is good, it's good.*

John Brockman is sort of an intellectual impresario who also is an agent for authors, and he asks a yearly question and publishers ask it of 150 people, quite often interesting people, to give brief answers to--and publishes the result. I think there is one that's just out, and they're not turning out to be bestsellers. A few years ago, before they weren't bestsellers, his question was, "What is your favorite formula?" I answered. My favorite formula was about success, and I wrote success equals talent plus luck and great success equals talent plus a lot of luck. That's the formula that Michael refers to.

*Speaker 7: That was great and I have two questions, I guess. One bears on this abduction point, and I think where does the following fit into your thinking? I think it's distinct from data. I think about things like think about general relativity, and so what happens here is we have a few things. One is tensors, an interesting mathematical form, we have non-Euclidean geometry, the intuition that you can treat time as if it were a spatial dimension. None of this has to do with the size of the data set. It has to do with a certain kind of schema, some formal structure mathematical computation algorithm, which is not arrived at by observing regularities in data sets. That's, I think, what explains revolutions in science.*

*Einstein says, "You know, Minkowski, that's useful. I can borrow that schema." Where do these kinds of structures appear in your framework?*

Many years ago, when I was a graduate student, I was exposed to somebody who thought he knew what creativity was, and he built a test, which is still in use, it's called the remote association test. The remote association test is you're given three words, and they're all three linked to some common concept, and I'm not going to try to think of example. Actually, in my book, I did cite a couple of the examples.

I've always thought, I mean, I asked him how well will this predict the creativity of architects and other people that were studying at the time, and he answered very arrogantly, "Well if the criterion is good, the correlation should be perfect because this is creativity."

He convinced me, and I think that creativity is putting remote things together, it's seeing connections that are there, and that are real, and you recognize it once the connection has been made, but we often said, "Oh, I would never have seen it." This is our experience of the creativity of others is that they have seen a pattern that we can recognize after the fact, that we would never have seen it. That was in the background of my remarks earlier, that I'm not sure that that pattern would not be uncovered in big data. In principle, I think it should be discoverable in big data. There is no magic. It was there. He put together things that existed. It didn't come out of nowhere.

The same is true, I'm quite convinced, of Jobs. Clearly, you can't say that Einstein was luck. I mean, that possibility isn't there, but all the elements were there, he just, nobody else could put them together, or didn't put them together. Could is an impossible word in that context, and he did, but no magic.

*Speaker 8: Your work in psychology became, after a few decades, let's say, almost a textbook in economics. Meaning that one discipline became a mainstream in another discipline a few decades afterwards. Relating to what Doyne Farmer talked about today, I wonder what you think it would take to economics these days to be able to be open to other disciplines related to the things that came up today about new type of theories and how to also the way to absorb big data into economic theory and empirics.*

First of all, I'm not an economist, so I would be completely free not to answer the question at all because it's not my field. I do have a couple of remarks on this. In the first place, I think this idea of economics as a closed discipline and as very resistant to change doesn't seem to be true. I mean, I have a prize in economics, and I was considered a heretic 25 years ago. That's very, very quick when that happens. I want to cite the Nobel Prize that was given yesterday to my friend, Angus Deaton. He's an economist, and I have learned more about how to resist my own causal intuitions from him than I have from anybody else because he was trained, and he attributed that to his discipline as an economist to be very, very careful about causality.

There is a mode of thinking about social science that other social sciences can borrow from that work. I'm not on the complaining side when it comes to economics.

*Speaker 9: Has your work ever gone into ethics? My question is motivated by your comment on loss aversion and when a system or a corporation is changed, and it affects the losers more than the winners. Does that impose an ethical framework or consideration on how do we factor that in? Should we, as a society, be looking for better solutions or does it say anything about income inequality of distributions?*

Let me try for a short answer. Obviously, loss aversion has to be ethically relevant. In fact, it is strongly relevant to the intuition that people have about the fairness of behaviors. Many years ago, Richard Thaler and Jack Knetsch and I did a study of perceptions of fairness in the market, and it's all about losses. There are constraints on what people are allowed to do in imposing losses on others. There are much fewer constraints, you don't have to share your profits, but you cannot impose losses just in order to make a profit for yourself if you care about being perceived as fair. Clearly, all of this is highly relevant.

There is a major distinction that comes up in decision making and it comes up in ethical discussions, as well. The distinction is between final states and changes. You can think in terms of final states like what is the ideal distribution of a good, or you can think of the current situation and evaluate ways of changing from the current situation. You don't get to the same conclusions when you're thinking of the ideal state and when you're thinking of changes, and the ethical intuitions that we have are primarily about changes, but we do have intuitions about ideal distributions, we have intuitions about changes, and they don't fit together. I don't think, when you look carefully at human ethical intuitions, our intuition, that they are not consistent, but you cannot ignore losses in thinking about ethics. You cannot merely consider an ideal world because you've got to get there, and getting there involves gains and losses.

*Speaker 10: Hi. You talked about how confidence does not equate to validity. Is there a logical basis for the value human beings place on confidence in that we elect people who seem confident to run companies and run countries. Is there a logical basis for that given what you've said how it actually doesn't mean anything.*

That's really a beautiful question, which was why do we put so much value on confidence? There has been a lot of work in recent years among psychologists on how we form impressions of other people. There seem to be two major dimensions in the impressions we form, and one of them is warmth from good to bad, warm to cold, and the other one is competence or dominance. It turns out that competence or dominance is very important to us when we value it. People, it is one of the, when people are exposed to pictures of individuals, they form an impression of competence and strength at the same speed that they form an impression of likability, which is really less than a second.

Confidence is part of that complex, it's part of that, and we want that. We want the people that we depend on to be competent and to be confident, and that is true about our leaders, it is true about our parents. We're afraid of competence and dominance from strangers and people that we don't like, but we very much need it from people who lead us. There is a huge desire for confident leaders. There is a huge desire for intuitive leaders. If you think of a national leader, you think of two who reach the same decision, and one of them reaches it quickly and the other slowly, we tend to be more attracted to the one who reaches the conclusion quickly.

This is very deep that wishful confidence, and also, naturally, there is a wish for ourselves to be confident, to feel confident, but it's fed that is when I act confidently I'm rewarded for it by other people, similarly, by the way, for optimism. Optimism is rewarded by other people.

*Speaker 11: What's your thoughts on why some rules tend to work, have a high probability of working such as stopping at a red light, but other rules, like you were talking about with employee valuations following a set rules don't tend to necessarily work very well? What can executives do in advance to increase the odds that the rules they're setting fall in the higher probability range?*

Again, the issue is one of clarity and clarity of feedback, unequivocal feedback. The example of stopping at a red light, that's one extreme of the distribution, where you know if you violated that rule or not. When it's - the less clear it is, whether you violated a rule or not, the more likely it is that the rule will be violated.

*Speaker 12: This is more of a comment, but I would be interested in your reaction. The question on competence, it occurs to me that in many small scale situations, you want somebody confident, the captain of a ship, for example, you want somebody who knows what they're doing. There are a lot of small scale situations. It seems to me it's only a few special situations, maybe major political questions, financial advice, and things like that where people are confident but not competent.*

You use a very interesting phrase in your question. You said, "We want somebody who knows what he's doing." But confidence is we want somebody who looks, who presents himself as if he knows what he's doing. That's not exactly the same thing, but we do want that, as well.

*Speaker 13: Somewhat along the same lines, in your discussion of the entrepreneur harking back to that. The idea that entrepreneurs basically are people who made mistakes and got lucky. There's an alternative explanation, which is that they are people who are not only confident of their ability, but confident that whatever mistakes they stumble into, they will be able to find a way through. Okay? That may be the ultimate, I mean, that's an ultimate form of optimism, and it seems to be a decision tree optimism. Would you comment on that?*

It's certainly the case that leaders and entrepreneurs, they don't view themselves as gamblers. They view themselves as captains of a ship in a stormy sea. It's a very, so it's clear that this is their view. I did not mean to say that the entrepreneurs who are successful were just lucky. Obviously, they had to be very talented. What I did say was that in general, they tend to overestimate their odds, and that in many cases, if they didn't overestimate their odds, they wouldn't take the gamble. The best example is a study that was done a long time ago about small businesses, owners of small businesses like restaurants and laundromats and so on. Where people are asked--where the survival rate is known, it's about 1/3 for five years in the United States, so small business has a 2/3 probability of not existing five years after it's set up. So when you ask people who open small business, what they think their odds of success are, they're very high. 85% and up. Some of them are certain they will succeed.

When you ask them what is the probability that a business like yours will succeed, it's much lower, it's about 2/3. But you can see that if somebody opens an Italian restaurant, clearly they're optimistic about the Italian restaurant. Otherwise, they wouldn't do it. It is also the case that there is that over-estimate.

There is another source of data, which I find really quite interesting. There is an institution in Canada that will assess startups and innovations and inventions for their commercial potential. If you have an invention, you can send it to them, and they will rate it for you, and they're really very good at it, especially rating things as hopeless.

If they rate something as hopeless, it really is hopeless, and then you can, over the years, they have accumulated data on what are the reactions of people upon being told that their invention is hopeless. Approximately 50%, I think, carry on anyway and fail, but, of course, it's the same perseverance that causes others to succeed.

*Speaker 14: I have one question, actually, and that is in your opinion, it's been said, for example, in many systems that your success is in part a function of the context of the environment in which you find yourself. From a physicist's perspective, you would say that in some cases, the susceptibility of a system, the small perturbations can be infinite. In other situations, it's zero or almost zero. What if Jobs had been born ten years earlier or ten years later? What if he had not met Steve Wozniak? Okay. This is where luck enters into it, to some degree. What about all the pairs of people in their garage trying to reinvent the Internet today as opposed to Brin and Page?*

Kahneman: I don't know. What if the owners of Google had been offered $1 million instead of $750,000-- I think that was the amount that was offered, that would have been a big difference to a lot of things. Michael has a book on luck and skill, and I discuss it in exactly the same way. Yeah. Success has a lot of luck into it. We can say it differently because that can be misleading. Talent is necessary but it's not sufficient, so whenever there is significant success, you can be sure that there has been a fair amount of luck.

*Speaker 15: In terms of your thoughts, I guess we haven't really sort of characterized it as this, but when you look at these, it seems like an issue of competence and confidence, complexity seems to be the issue in terms of, I guess, what discerns the good from the meaningful. Here we are in 2015, we have very polarized sort of political landscape, and it seems that a lot of the people just don't understand complexity. How do we get people up to speed or do you have any perceptions in terms of really in this idealistic sort of motif, how to bring complexity to the masses?*

I feel bad about ending this conversation on such a note, but my answer is no. I don't see how to do this. I mean, I really don't see how it's possible.

In the book I wrote, I distinguish those two ways of thinking about things, faster and slower. When you're talking to the public at large, and you want to get action or you want to get something embraced and so on, you have to speak to their fast thinking. You have to have a story that is engaging that people can relate to.

Communicating complexity, by communicating scientific evidence, scientists are deluded to some degree about themselves, and certainly about the public, about the compelling power of evidence. Evidence is not all that compelling. That's what we see all around us. People who have strong beliefs without any evidence, and who can resist any evidence to the contrary, and complexity is really not what people are after, and so I'm not optimistic at all. I could have said simply no, but I'm just saying it more slowly. Thank you.

*Mauboussin: Thank you. Thank you, everybody. I do want to give another thank to all of our core organizers, John Rundle, who also is our wonderful MC for the day. Marty Liebowitz, thanks again, Marty, for not only your help in putting this all together, but also your hosting the event. Chris Wood, who unfortunately was unable to join us from SFI. I would just say on behalf of all of my colleagues at Santa Fe Institute, thank you all for attending today. Certainly, let us know if you'd like to learn more about SFI, we would certainly welcome that, welcome your interest, and we hope to have the opportunity to exchange more ideas in the future. Have a great afternoon, everybody.*