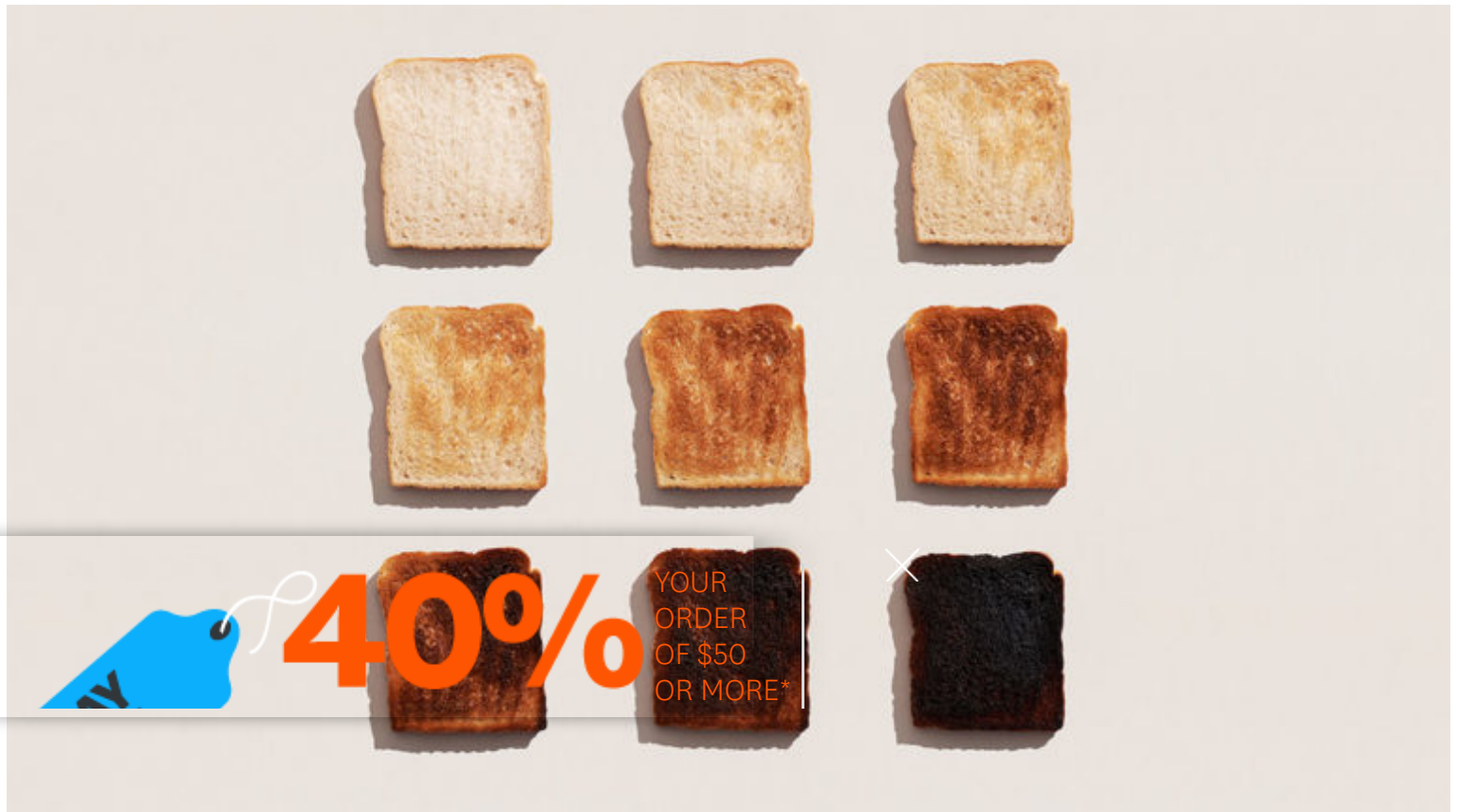Harvard
Business
Review

DECISION MAKING

Why "Many-Model Thinkers" Make Better Decisions

by Scott E. Page

NOVEMBER 19, 2018

"To be wise you must arrange your experiences on a lattice of models."

— Charlie Munger

Organizations are awash in data — from geocoded transactional data to real-time website traffic to semantic quantifications of corporate annual reports. All these data and data sources only add value if put to use. And that typically means that the data is incorporated into a model. By a model, I mean a formal mathematical representation that can be applied to or calibrated to fit data.

Some organizations use models without knowing it. For example, a yield curve, which compares bonds with the same risk profile but different maturity dates, can be considered a model. A hiring rubric is also a kind of model. When you write down the features that make a job candidate worth hiring, you're creating a model that takes data about the candidate and turns it into a recommendation about whether or not to hire that person. Other organizations develop sophisticated models. Some of those models are structural and meant to capture reality. Other models mine data using tools from machine learning and artificial intelligence.

The most sophisticated organizations — from Alphabet to Berkshire Hathaway to the CIA — all use models. In fact, they do something even better: they use many models in combination.

Without models, making sense of data is hard. Data helps describe reality, albeit imperfectly. On its own, though, data can't recommend one decision over another. If you notice that your best-performing teams are also your most diverse, that may be interesting. But to turn that data point into insight, you need to plug it into some model of the world — for instance, you may hypothesize that having a greater variety of perspectives on a team leads to better decision-making. Your hypothesis represents a model of the world.

Though single models can perform well, ensembles of models work even better. That is why the best thinkers, the most accurate predictors, and the most effective design teams use ensembles of models. They are what I call, many-model thinkers.
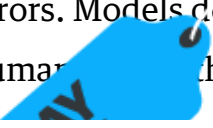
In this article, I explain why many models are better than one and also describe three rules for how to construct your own powerful ensemble of models: *spread attention broadly, boost predictions,* and *seek conflict*.

The case for models

First, some background on models. A model formally represents some domain or process, often using variables and mathematical formula. (In practice, many people construct more informal models in their head, or in writing, but formalizing your models is often a helpful way of clarifying them and making them more useful.) For example, Point Nine Capital uses a linear model to sort potential startup opportunities based on variables representing the quality of the team and the technology. Leading universities, such as Princeton and Michigan, apply probabilistic models that represent applicants by grade point average, test scores, and other variables to determine their likelihood of graduating. Universities also use models to help students adopt successful behaviors. Those models use variables like changes in test scores over a semester. Disney used an agent-based model to design parks and attractions. That model created a computer rendition of the park complete with visitors and simulated their activity so that Disney could see how different decisions might affect how the park functioned. The Congressional Budget office uses an economic model that includes income, unemployment, and health statistics to estimate the costs of changes to health care laws.

In these cases, the models organize the firehose of data. These models all help leaders explain phenomena and communicate information. They also impose logical coherence, and in doing so, aid in strategic decision making and forecasting. It should come as no surprise that models are more accurate as predictors than most people. In head-to-head competitions between people who use models and people who don't, the former win, and typically do so by large margins.

Models win because they possess capabilities that humans lack. Models can embed and leverage more data. Models can be tested, calibrated, and compared. And models do not commit logical errors. Models do not suffer from cognitive biases. (They can, however, introduce or replicate human biases that is one of the reasons for combining multiple models.)

Combining multiple models

While applying one model is good, using many models — an ensemble — is even better, particularly in complex problem domains. Here's why: models simplify. So, no matter how much data a model embeds, it will always miss some relevant variable or leave out some interaction. Therefore, any *model will be wrong.*

With an ensemble of models, you can make up for the gaps in any one of the models. Constructing the best ensemble of models requires thought and effort. As it turns out, the most accurate ensembles of models do not consist of the highest performing individual models. You should not, therefore, run a horse race among candidate models and choose the four top finishers. Instead, you want to combine diverse models.

For decades, Wall Street firms have used models to evaluate investment risk. Risk takes many forms. In addition to risk from financial market fluctuations, there exist risks from geopolitics, climactic events, and social movements, such as occupy Wall Street, not to mention, risks from cyber threat and other forms of terrorism. A standard risk model based on stock price correlations will not embed all of these dimensions. Hence, leading investment banks use ensembles of models to assess risks.

But, what should that ensemble look like? Which models does one include, and which does one leave out?

The first guideline for building an ensemble is to look for models that focus **attention** on different parts of a problem or on different processes. By that I mean, your second model should include different variables. As mentioned above, models leave stuff out. Standard financial market models leave out fine-grained institutional details of how trades are executed. They abstract away from the ecology of beliefs and trading rules that generate price sequences. Therefore, a good second model would include those features.

The mathematician Doyne Farmer advocates agent-based models as a good second model. An agent-based model consists of rule-based "agents" that represent people and organizations. The model is then run on a computer. In the case of financial risk, agent-based models can be designed to include much of that micro-level detail. An agent-based model of a housing market can represent each household, assigning it an income and a mortgage or rental payment. It can also include behavioral rules that describe conditions when the home's owners will refinance and when they will declare bankruptcy. Those behavioral rules may be difficult to get right, and as a result, the agent-based model may not be that accurate — at least at first. But, Farmer and others would argue that over time, the models could become very accurate.

We care less about whether agent-based models would outperform other standard models than whether agent-based models will read signals missed by standard models. And they will. Standard models work on aggregates, such as Case-Shiller indices, which measure changes in prices of houses. If the Case-Shiller index rises faster than income, a housing bubble may be likely. As useful as the index is, it is blind to distributional changes that hold means constant. If income increases go only to the top 1% while housing prices rise across the board, the index would be no different than if income increases were broad based. Agent based models would not be blind to the distributional changes. They would notice that people earning $40,000 must hold $600,000 mortgages. The agent based model is not necessarily better. It's value comes from focusing attention where the standard model does not.

The second guideline borrows the concept of **boosting**, a technique from machine learning. Ensemble classification algorithms, such as random forest models consist of a collection of simple decision trees. A decision tree classifying potential venture capital investments might say "if the market is large, invest." Random forests are a technique to combine multiple decision trees. And boosting improves the power of these algorithms by using data to search for new trees in a novel way. Rather than look for trees that predict with high accuracy in isolation, boosting looks for trees that perform well when the forest of current trees does not. In other words, look for a model that attacks the weaknesses of your current model.

Here's one example. As mentioned, many venture capitalists use weighted attribute models to sift through the thousands of pitches that land at their doors. Common attributes include the team, the size of the market, the technological application, and timing. A VC firm might score each of these dimensions on a scale from 1 to 5 and then assign an aggregate score as follows:

Score = 10*Team + 8*Market size + 7*Technology + 4*Timing

This might be the best model the VC can construct. The second best model might use similar variables and similar weights. If so, it will suffer from the same flaws as the first model. That means that combining it with the first model will probably not lead to substantially better decisions.

A boosting approach would take data from all past decisions and see where the first model failed. For instance, it may be that be that investment opportunities with scores of 5 out of 5 on team, market size, *and* technology, do not pan out as expected. This could be because those markets are crowded. Each of the three attributes –team, market size, and workable technology – predicts well in isolation, but if someone has all three, it may be likely that others do as well and that a herd of horses tramples the hoped for unicorn. The first model therefore would predict poorly in these cases. The idea of boosting is to go searching for models that do best specifically when your other models fail.

To give a second example, several firms I have visited have hired computer scientists to apply techniques from artificial intelligence to identify past hiring mistakes. This is boosting in its purest form. Rather than try to use AI to simply beat their current hiring model, they use AI to build a second model that complements their current hiring model. They look for where their current model fails and build new models to complement it.

In that way, boosting and attention share something in common: they both look to combine complementary models. But attention looks at what goes into the model — the types of variables it considers — whereas boosting focuses on what comes out — the cases where the first model struggles.

Boosting works best if you have lots of historical data on how your primary model performs. Sometimes, we don't. In those cases, seek **conflict**. That is, look for models that disagree. When a team of people confronts a complex decision, it expects — in fact it wants — some disagreement. Unanimity would be a sign of group think. That's true of models as well.

The only way the ensemble can improve on a single model is if the models differ. To borrow a quote from Richard Levins, the "truth lies at the intersection of independent lies." It does not lie at the intersection of correlated lies. Put differently, just as you would not surround yourself with "yes men" do not surround yourself with "yes models."

Suppose that you run a pharmaceutical company and that you use a linear model to projects sales of recently patented drugs. To build an ensemble, you might also construct a systems dynamics model as well as a contagion model. Say that the contagion model results in similar long-terms sales but a

slower initial uptake, but that the systems dynamics model leads to a much different forecast. If so, it creates an opportunity for strategic thinking. Why do the models differ? What can we learn from that and how do we intervene.

In sum, models, like humans, make mistakes because they fail to pay attention to relevant variables or interactions. Many-model thinking overcomes the failures of attention of any one model. It will make you wise.
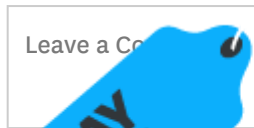

Scott E. Page is the Leonid Hurwicz Collegiate Professor of Complex Systems, Political Science, and Economics at the University of Michigan and an external faculty member of the Santa Fe Institute. He is the author of *The Model Thinker* (Basic Books, 2018).

This article is about DECISION MAKING

⊕ **FOLLOW** THIS TOPIC

Related Topics: DATA

Comments

Leave a Co

POST

2 COMMENTS

Sophia Li   5 days ago

Thanks for the great article. to link with my job, it is true for hiring a candidates that when we think of the right candidate for a dept.,we always form a MODEL not only from our point of view, meantime we need to have a deep understanding of the hiring managers's MODEL, the both side model will make the hiring decision more effective.

REPLY

∨ **JOIN THE CONVERSATION**

POSTING GUIDELINES

We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.