Why Humans Distrust Algorithms – and How That Can Change

Feb 13, 2017   North America   Podcasts, Research, Video

Mathematical models have been used to augment or replace human decision-making since the invention of the calculator, bolstered by the notion that a machine won't make mistakes. Yet many people are averse to using algorithms, preferring instead to rely on their instincts when it comes to a variety of decisions. New research from Cade Massey and Joseph Simmons, professors in Wharton's department of operations, information and decisions, and Berkeley J. Dietvorst from the University of Chicago finds that control is at the core of the matter. If you give decision-makers a measure of control over the model, they are more like to use it. Massey and Simmons spoke to Knowledge@Wharton about the implications of their research.

An edited transcript of the conversation follows.

Knowledge@Wharton: Could you give us a brief summary of this research? This paper is a follow-up to something you had done recently.

Joseph Simmons: We're studying a phenomenon called "algorithm aversion," which is the tendency for people to not want to follow specific evidence-based rules when they make decisions, even though a lot of the research shows that's exactly the way you should be making judgments and forecasts. A lot of people just want to rely on their gut or go by the seat of their pants. They don't want to rely on consistent, evidence-based rules — and they should.

We've been studying for a couple of years now why or under what circumstances don't they want to rely on these algorithms. Our second paper is about how to get people to be more likely to rely on algorithms. We basically found that if you tell people, "You can go with an algorithm that is going to give you some advice, or you can go with your own opinion,"

and you ask them, "What do you want to do?" — they're actually OK with saying, "I'll use the algorithm."

However, once you give them some practice and let them see how their algorithm performs, all of a sudden they don't want to use it anymore. That's because they see the algorithm make mistakes. Once they see algorithms or computers make mistakes, they don't want to use it anymore, even though the algorithm or computer is going to make a smaller mistake or more infrequent mistakes than they are going to make.

> "A lot of people just want to rely on their gut or go by the seat of their pants. They don't want to rely on consistent, evidence-based rules — and they should."
>
> –Joseph Simmons

Knowledge@Wharton: The algorithms are supposed to be perfect.

Simmons: Right. People want algorithms to be perfect and expect them to be perfect, even though what we really want is for them to simply be a little better than the humans. Our first paper is kind of pessimistic and shows that once people see the algorithms do their thing, they don't want to use them. Our second paper shows that you can get people to use algorithms as long as you give them a little bit of control over them. You say, "The algorithm tells you that this person is going to have a GPA of 3.2. What do you think their GPA is going to be?" They don't want to just go with the 3.2 that the algorithm [predicts]. But if you say, "You can adjust it by .1," then their response is: "OK, I'm fine to use the algorithm." We find as long as you give people a little bit of control over these things, they're more likely to use them. And that's pretty good news.

Cade Massey: We operationalize this in an experimental context, but we're motivated by real world context. Some of the early ideas for this research came from working with companies where we would go in with models for decision–making about hiring and recruiting new employees. Based on many years' worth of data and some pretty good analytics, we were sure that we had the best advice going. Yet those organizations were reluctant to use [these models] because they wanted to

rely on just their intuition.

It's very common in hiring, it's very common in performance evaluation, it's even common increasingly in some fields where they automate decision-making, like how to manage a hedge fund or what should the sales forecast be for some product. Those are all places where increasingly automatically generated forecasts or advice is available. We call it an algorithm. And the final decision-maker has discretion over whether they listen to that advice, use their own [instincts] or use a blend.

Knowledge@Wharton: Your key takeaway was that people are less averse to using algorithms if they have some control. But there is a conclusion that surprised you in how much control you had to give them to make them feel better. Tell us about that.

Massey: We were agnostic on how much control would be necessary to get them to buy in. The downside to giving them control is they start degrading the algorithm. In most domains, they're not as good as the model. The more of their opinion is in there, the worse it performs. In some sense, you'd like to give them as little control as possible and still have them buy in. We didn't know what the answer to that would be. We got early evidence that it wasn't going to be very much; then we started testing the limits of it and found that we could give them just a little bit of control. You know, move something around 5% or so and they would be much more interested in using the algorithm. If you give them more, it doesn't increase the lift at all. Give them a little bit, and it's about the same as giving them moderate influence.

Simmons: What's nice about that is when they adjust the algorithms, they make them worse. But if they can only adjust it [a little bit], they can only make it that much worse. And since they are more likely to use it in that case, their final judgments will wind up being correlated with the algorithm close to perfectly. We can't get people to use algorithms 100%, but we can get them to use algorithms 99%, and that massively improves their judgments.

"We can't get people to use algorithms 100%, but we can get them to use algorithms 99%, and that massively improves their judgments."

–Joseph Simmons

Knowledge@Wharton: If I'm a business owner or someone who is going to be charged with using one of these algorithms, how might I apply this research in real life?

Massey: The overarching lesson would be that you don't simply impose a monolithic model or black box model and say, "This is how you use judgment. This is how you should codify your decision-making." People will fight that. You want to let them have discretion. That's going to look different in different places. Consider a graduate school making admissions — they rank their applicants and, at some point, they cut a line and make exceptions. They move people around. You can automate some of that process. Even if you use some of their judgment to provide inputs to the model, you can use an automatic model to say, "These are the folks that you should take."

On one hand, you could say, "Here is the model. This is what it says; take it or leave it. We're going to automate the process." You're basically going to have a revolt on your hand. But if you say, "Here is a model that is advisory. We suggest that you consider it. If you want to move things around, move them around." We've worked with schools in exactly this way, and what you find is that they're a little skeptical early on. They lean on the model some, and over time they are practically using the entire model as it is, even though they have discretion to change as much as they want.

Knowledge@Wharton: I would also think that the presentation would be important to make sure that people know they have this control.

Simmons: I think the important thing is to avoid an all-or-nothing framing — like having to stick with the algorithm 100% of the time. If people think that based on how you have described it, they are going to push back. But if you can frame it as, "Even 99% of the time, we are going to go with the algorithm, but you have the option to change the algorithm or to

not go with the algorithm at a given moment," that's going to make people a lot more amenable to using it.

Another context in which this might matter is self-driving cars. You can imagine people not feeling comfortable being in a self-driving car if they have no control whatsoever. But if they say, "Well, there's this sort of thing you can do. It's a little bit difficult and unusual, but there's this thing you can do to gain control over the car in circumstances where you might need to do it. We found that people never need to use this, but it does exist" — we would predict in that circumstance, people would be much more amenable to getting in a self-driving car because there is some control. Autopilot is usually safer than real pilots, but people want a pilot there, even though lots of plane crashes are due to pilot error. They feel better about that. I think our research speaks to that a bit.

"People want algorithms to be perfect ... even though what we really want is for them to simply be a little better than the humans."

–Joseph Simmons

Knowledge@Wharton: Are there other stories in the news that might apply to this research?

Massey: How about election forecasts?

Simmons: Yes — so back in November, we had a presidential election that surprised the world. There was a bunch of people out there predicting, based on past polling information, what the election was going to look like. Probably the most famous case is Nate Silver, who writes for FiveThirtyEight.com. He said there was a 70% chance that Hillary Clinton would win the election and a 30% chance that Donald Trump would win. Of course, Donald Trump won, and there was a lot of pushback against Nate Silver at the end of it. He was wrong, we think, in part because the model was wrong. The thing is, it wasn't necessarily wrong because 30% happens 30% of the time. When individual pundits go out there and say one thing is going to happen, they don't get as much blowback as when a person who uses statistics and an algorithm — which people expect to be right 100% of the time — gets it wrong. I think the blowback that we've seen in the direction of Nate Silver has

been in line with what we have found before.

Massey: It circles back to that first paper, where people are just much harder on models and algorithms when they err than they are on people. They are just more forgiving. We have explored it a little bit, but the bottom line is that [models and algorithms] are held to a higher standard.

Knowledge@Wharton: Does that make sense? No person or thing is perfect.

Massey: There are a variety of reasons we think people do this. One is that they believe that people can improve over time whereas a model is relatively fixed. Both of those things are not necessarily true. Models can improve over time, and people don't necessarily improve over time. The psychology of that is compelling but not necessarily correct. There certainly would be some settings where people can improve more than a model, but we think that people have that intuition more than they actually should.

Knowledge@Wharton: Is there anything that sets this research apart from other work in this area?

Massey: We're not the first to talk about the difference between model judgment and human judgment. It's been established for decades now that models are quite good. We're relatively early in to trying to understand why that is and how you fix [perceptions about] it.

"You can't prescribe anything until you better understand why it exists."

–Cade Massey

Simmons: Not many people have documented previously the reasons why people are averse to using algorithms. There's been some anecdotal research, there's been some writings about how people don't like these things, but no one's really looked at it systematically before.

Massey: And again, back to the motivation. The motivation was we work with organizations, we want them to use more models, we need to know how to break down that bias. You can't prescribe anything until you better understand why it exists.

Knowledge@Wharton: What's next for this research?

Massey: We continue to play with some factors that might contribute to the people's reluctance to use algorithms, but we also want more real-world tests of it. If we work with professionals with real money at stake, do they fall into the same biases? Are there ways for us to help them? We have a couple of organizations that we have talked to over time, and they are interested in running experiments on their employees or customers to see if what we see in the lab takes place in the field.