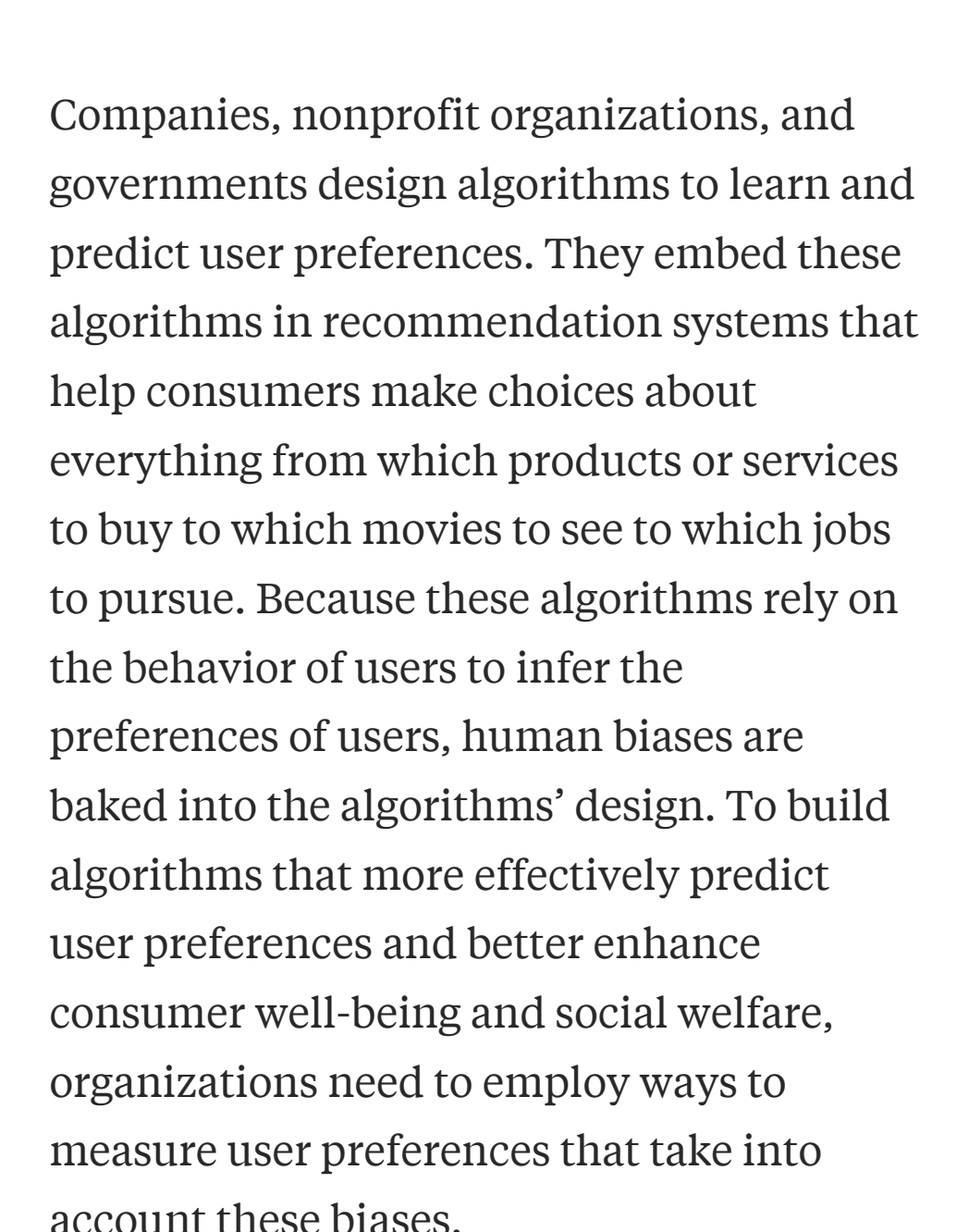


Algorithms

Why Algorithm-Generated Recommendations Fall Short

by Carey K. Morewedge

January 09, 2024



Michael Blann/Getty Images

Summary. The online systems that make recommendations to us often rely on their digital footprint — our clicks, views, purchases, and... [more](#)

Companies, nonprofit organizations, and governments design algorithms to learn and predict user preferences. They embed these algorithms in recommendation systems that help consumers make choices about everything from which products or services to buy to which movies to see to which jobs to pursue. Because these algorithms rely on the behavior of users to infer the preferences of users, human biases are baked into the algorithms’ design. To build algorithms that more effectively predict user preferences and better enhance consumer well-being and social welfare, organizations need to employ ways to measure user preferences that take into account these biases.

Consider a new app that promises to revolutionize your dinners. It uses a proprietary algorithm to recommend foods tailored just for you. Eager to try it, you quickly sign up. But the recommendations aren’t what you expected. Monday, it’s pizza. Tuesday, a burger. Wednesday, fried chicken. Thursday, barbecue. Friday, steak frites. Confused, you call the company to ask what’s going on.

“Our algorithm analyzes your past food orders and picks out your favorites,” they explain. “Then we only recommend those favorites, nothing else. It’s the perfect personalized meal plan!”

This is how algorithms are built. This is why algorithms are failing us.

Algorithms that help us make the best decisions — that connect us with the best ideas, experiences, jobs, people, and products — should enrich our lives. But algorithms — such as those that curate [social media](#), [allocate health care](#), and [price car insurance](#) — are failing to deliver on this promise. Algorithms built for and by [governments](#) and nonprofits, such as algorithms used to [forecast criminal activity](#) and [score college entrance exams](#), are also falling short.

One core reason for this disappointing performance is modern algorithms that assist and replace human decision-making, such as recommendation systems, are built on a psychological model of what the user is doing. In a new paper in *Nature Human Behavior*, my colleagues and I argue that a fundamental constraint is what algorithms are designed to learn from: our behavior. Algorithms rely on our clicks, views, purchases, and other digital footprints to infer our preferences. These “[revealed preferences](#)” can identify our preferences for things we didn’t know we wanted such as a trip to Montenegro, dinner at Le Bernardin, or a reality TV show about a restaurant in West Hollywood. Still, the revealed preferences of individual users are an incomplete and, at times, misleading measure of the “[normative preferences](#)” that compose their true goals and values.

Biases that Affect Decision-Making

Psychologists and behavioral economists have documented many [anomalies in human decision-making](#). When these psychological biases influence our behavior, algorithms will wrongly conflate our revealed and normative preferences. Here are three examples:

Fast Thinking

We often lack the knowledge, time, ability, or motivation to decide rationally. We then rely on associative intuitions and habits and are influenced by contextual factors like the [default we are recommended](#). These decision strategies are usually good but can create [systematic biases in our decision-making](#). Algorithms trained on our habits may learn preferences we no longer endorse (e.g., [most smokers want to quit](#)). Algorithms trained on our behavior will [reflect our biases and structural inequities](#) in systems that we never endorsed and may be hidden to us. When Amazon made a hiring algorithm, it learned to preference male candidates over female candidates from the human hiring decisions on which it was trained — [a gender bias](#) that had gone unnoticed until it was codified by the algorithm.

Wants, Not Shoulds

We hold conflicting desires — to indulge in the [present or wait for the future](#), to take for ourselves or share, to exploit what we know or take a risk and try something new. Algorithms observe and base recommendations on user preferences when decisions are made. These algorithms learn the conflict resolutions that are most immediately gratifying (wants) even if we aspire to longer-term resolutions (shoulds). [Netflix users](#) put high-brow films like *Planet Earth* on their watch lists but end up binging on soaps like *Bridgerton* and action movies like *Extraction*.

Social Norms and the Status Quo

On platforms like Amazon and YouTube and Facebook, algorithms steer us to see and buy what people like us are seeing and buying. We rely on recommendation systems and lists (e.g., best sellers) to find options in platforms’ large catalogs. Recommendation systems help us find what we like, but they also [change our preferences](#) and [reduce the diversity](#) of what we see and buy, increasing the market share of popular and existing options. Without considerable engineering, recommendation systems would suggest *Harry Potter* everywhere, whether relevant (e.g., to users watching *Lord of the Rings*) or irrelevant (e.g., to users buying *Mastering the Art of French Cooking*).

There are several measures that organizations can take to build more effective algorithms.

Audit algorithms for human bias.

Companies can use A/B testing and dig into internal data to reveal when and why user behavior and the normative preferences of users diverge. Researchers, with the cooperation of a health care system, found that the organization’s [algorithm prioritized white over Black patients for access to additional services](#) because the health care system spent more on similarly healthy white patients than on those who are Black and used spending as a proxy for health. A [market value analysis algorithm](#) used by [municipal governments](#) in the United States to guide investment in and the continuation of public services tends to deprioritize neighborhoods with a higher concentration of Black and poorer residents — in part, because it incorporates data like homeownership rates, which are influenced by structural biases due to historically prejudiced lending practices.

Even without access to internal company data, scientists, governments, and non-profits can still audit algorithms for human bias by [observing individual users](#), [examining the API of platforms](#), and [using bots](#), and with intervention studies. Facebook users who were paid by researchers [in an experiment](#) to deactivate their account for four weeks spent more time with their families and reported feeling better, and many abandoned the platform after the payments stopped. These results suggest many users had spent time on Facebook out of habit, not because they enjoyed it or found it satisfying.

Improve algorithm design to better reflect users’ normative preferences.

Algorithm designers could tune algorithms from reflecting wants toward reflecting more shoulds by expanding the time horizon of observed behavior. When Meta pared down the volume of onsite notifications users received to only the most relevant notifications, for instance, [the intervention](#) initially reduced visits to Facebook in the short term, but eventually visits recovered and grew over the long term.

When people make choices between alternatives (e.g., an apple and a chocolate bar), the path their mouse takes can reveal conflicting preferences between those options. Mouse-tracking technologies that analyze mouse trajectories [could reveal](#) conflicts between wants and shoulds that are missed by a reliance on choice data (e.g., click-through rates). Of course, tuning algorithms entirely towards shoulds could kill user demand, but finding a better balance between wants and shoulds should benefit firms and users.

Train algorithms on different user data.

Algorithms are typically trained on one portion, or a fold, of a data set and then validated on other data that serves as a hold-out sample. Algorithm designers could [selectively train algorithms on data from users](#) who exhibit more deliberate choices and better decision-making and report desired outcomes (e.g., less loneliness, greater happiness or satisfaction, and so on). Designers could train algorithms on users who spend more time or consider more data before making choices. For instance, instead of training autonomous vehicles on everyone, designers could train AV’s on safe drivers. Designers could train social media algorithms on the users who are happiest, most satisfied with their experience, or who engage with content from credible sources. If those users don’t exist, designers could create simulations of users who behave in ways that reflect users’ normative preferences and train algorithms on that simulated data.

Craft algorithms that rely less exclusively on behavior and more directly on stated preferences.

Integrating human-in-the-loop design practices in algorithms benefit many applications of machine learning. Designers can solicit user preferences through surveys and interviews and blend this with the preferences revealed in user behavior by including their measures in the basket of objectives that algorithms are designed to optimize. If we tell Netflix to show us new science documentaries, its algorithm shouldn’t just serve us the sitcoms and action films in our viewing history. Users can often tell when an algorithm is failing them, and their explicit feedback can [improve recommendations](#). To understand how users want platforms to regulate private spaces in virtual reality for bullying and harassment, Meta collaborated with researchers to conduct [deliberative polling studies](#) with survey respondents from 32 countries.

It is time for companies, governments, and scientists to invest in the behavioral science of algorithm design. Designers cannot code their way out of human psychology. Algorithm design should move beyond revealed preferences. Algorithms could reflect the people we aspire to be, not just who we have been.

Carey K. Morewedge is a professor of marketing and Everett W. Lord Distinguished Faculty Scholar at the Questrom School of Business at Boston University.

Read more on [Algorithms](#) or related topics [Behavioral science](#) and [Consumer behavior](#)

Recommended For You

38 Smart Questions to Ask in a Job Interview

PODCAST **How to Build a Start-Up - Without VC Funding**

10 Common Job Interview Questions and How to Answer Them

Use Strategic Thinking to Create the Life You Want

Partner Center

Harvard Business Review

Start my subscription!

Explore HBR

HBR Store

About HBR

Manage My Account

Follow HBR

Explore HBR

Managing Data Science: An eight-week newsletter series on how to make analytics and AI work for your organization.

Sign Up