How artificial intelligence is changing economic theory

July 17 2015, by Leah Burrows

David C. Parkes, the George F. Colony Professor and Area Dean of Computer Science. Credit: Eliza Grinnell/Harvard Paulson School

A century of economic theory assumed that, given their available options, humans would always make rational decisions. Economists even

had a name for this construct: *homo economicus*, the economic man.

Have you ever met a human? We're not always the most rational bunch. More recent economic theory confronts that fact, taking into account the importance of psychology, societal influences and emotion in our decision-making.

So, are the theories that are predicated on homo economicus extinct? David C. Parkes, the George F. Colony Professor and Area Dean of Computer Science at Harvard John A. Paulson School of Engineering and Applied Sciences, doesn't think so. Humans may not always make rational decisions, but well-conceived algorithms do.

In a paper out today in the journal *Science*, Parkes and co-author Michael Wellman, of the University of Michigan, argue that rational models of economics can be applied to artificial intelligence (AI) and discuss the future of machina economicus.

At first glance, neoclassical economic theory and AI seem like strange bedfellows. Where and how do they overlap?

Parkes: The idea of rationality is a shared construct between AI and economics. When we frame questions in AI, we say: what are the objectives, what should be optimized and what do we know about the world we're in? The AI/economics interface has become quite fertile because there is a shared language of utility, probability, and reasoning about others.

Take, for example, the revelation principle in economics, which is a theory by which the design of economic institutions, such as markets, can be restricted to those where it is in the best interest of participants to

truthfully reveal their utility functions. Today's Internet advertising systems, which are populated by artificial trading agents, are an operational version of this [economic theory](#). Search engines are designing interfaces where advertisers reveal their budget constraints and goals, and these systems then provide the algorithms to fit those needs. You don't see many mechanisms like this in human societies but we may see them more and more in AI systems.

Where does current economic theory fall short in describing rational AI?

Machina economicus might better fit the typical economic theories of rational behavior, but we don't believe that the AI will be fully rational or have unbounded abilities to solve problems. At some point you hit the intractability limit—things we know cannot be solved optimally—and at that point, there will be questions about the right way to model deviations from truly rational behavior.

Poker is great example of a complicated reasoning problem: a lot of information is missing, you don't know the cards of the other players, you're uncertain about the card that will be dealt next and you're reasoning against another reasoning agent.

Recently, researchers developed an algorithm that effectively solves Heads Up Limit Texas Hold'em, applying game theory from economics. Researchers have developed an AI that has attained perfect rationality in this setting, and have done so using a number of general-purpose techniques. But this has come about after decades of research, and only for a restricted, two player version of poker.

But perfect rationality is not achievable in many complex real-world settings, and will almost surely remain so. In this light, machina

economicus may need its own economic theories to usefully describe behavior and to use for the purpose of designing rules by which these agents interact.

Besides poker, what would a rational AI system do better than a human?

One of the more complicated things people do is buying and selling property. It's actually really hard to describe to your real estate agent what you're looking for. Your broker may have some ideas, but real feedback only comes when showing you properties. It's an inefficient system. AI researchers have been developing the idea that an AI would elicit your preferences initially by direct query. An AI can show you a comparison between two houses and ask which one you prefer. As you answer, the AI can build a model of your preferences, adaptively eliciting information until it can reason that it knows enough about your preferences that the AI would be able to go out and— if it was very successful—know what house to buy on your behalf. Even if it were only moderately successful, the AI would bring back a couple of options that did a good job of optimizing your preferences within the market.

How would humans interact with these rational machines?

One way a machine can understand you and your preferences is to observe how you act. There is an approach called inverse reinforcement learning—economists call it revealed preference—where if an AI sees the decisions you make every day, it can begin to understand something about you. What your trade offs are, how you spend your time, what you like to wear and when, who do you like to talk to, who do you not like to talk to, etc. By observing your behavior, the AI can begin to build a model of your preferences. Then, you can imagine over time the AI

could start acting on those preferences and interposing itself—hopefully not in a creepy way.

Researchers have been developing techniques that look at your electronic information stream—your emails, voicemails, social media use—and learn about such things as your work environment, its hierarchy, who your manager is, and who reports to you. From there, the AI can decide which communications you actually need to see when. It can know when you're in a meeting and only interrupt you when someone important calls. It can do that based on modeling the value to you of information versus the cost of interruption. As the AI begins to do more for you, it can learn based on the choices you make. Of course, this presupposes that you're making rational choices on your own behalf, so your revealed preferences may not be the same as what your preferences should be.

What are the biggest challenges in building machina economicus?

The problems AIs will be solving, whether they are in a market or social context, are complex problems, especially where there are other participants in the system.

Optimal behavior will often depend on the behavior of others, making this quite different from reasoning about an environment where you are the only actor. If an AI is acting to buy, sell or exchange information or to set the price for something, it needs to reason about what other AIs are doing in the system as well.

Are there dangers in building something that is too rational?

Rationality can lead to unintended consequences. If you tell an AI car to

get into the city as quickly as possible, it might run some lights because its optimizing and reasoning about the probability of getting caught versus getting to its destination quickly.

Analogies like this ring true to the stock market as well. At the moment, we're living in a time where the presence of fast, algorithmic trading algorithms is leading to concerns about the fairness and efficiency of the stock markets. However, AI can also make markets more efficient, by doing a better job of matching supply and demand and allocating resources to those who need them and better understanding preferences and societal considerations.

There are also important questions about how rapid progress in AI will affect the workplace and the broader economy, both in the U.S. and globally, and this is an area that economists and policy makers should be looking at and are looking at.

More information: "Economic reasoning and artificial intelligence." *Science* 17 July 2015: DOI: 10.1126/science.aaa8403

Provided by Harvard University