Rationality and Robots

The Achilles heel of neoclassical economic theory has always been the assumption that humans are rational beings: that they will exercise, eat right, and save for retirement, that they won’t pay more for a cup of coffee—or a car, house, or share of stock—than it’s worth. But as behavioral economists routinely demonstrate, people’s decisions are inescapably influenced by psychology, emotion, societal forces, and cognitive biases.

Artificial intelligence, however, is a different story. Computers are rational in ways humans can never be, and recent years have seen rapid progress in AI research and achievement. Drones and self-driving cars are oft-discussed examples, but computer scientists have also been developing machines to conduct automated negotiations, to reason about consumer preferences, to make optimal buying choices and predict when prices will change. And as machines are increasingly put to work in economic contexts—setting sales prices for goods, competing in online auctions, executing high-speed market trades—a convergence is taking place between neoclassical economic tradition and cutting-edge computer science.

Which makes sense, says David C. Parkes, Colony professor and area dean of computer science. After all, what AI scientists and engineers are striving toward is a “synthetic homo economicus,” that mythical agent of neoclassical economics whose choices are perfectly rational. He calls this emerging robot species *machina economica*.

In a *Science* paper coauthored this past July with Michael P. Wellman, a University of Michigan computer scientist, Parkes considers the newfound relevance of neoclassical economics and asks what changes these AI advances may necessitate for both new theory and the design of economic institutions that mediate daily interactions. “We’re not just asking whether the neoclassical theories of economics will be more useful for AI systems than for human systems” (better, that is, at predicting machines’ thinking and behavior) “and how AIs will differ from people,” Parkes says, but “whether we’re beginning to understand how to design the rules by which AIs will interact with each other.” The latter is increasingly urgent, as the task of reasoning shifts from people to machines that learn humans’ preferences, overcome their biases, and make complex cost-benefit trade-offs. (For instance, algorithms are already estimated to drive more than 70 percent of U.S. stock-market trades.) “How will norms change,” he asks, “if, whenever I want to buy something, I let my software agent talk to your software agent?”

Real-world examples offer some guidance. Parkes points to the online auctions in which buyers bid for advertising space on Google search-results pages: high bidders at the top of the page, low bidders at the bottom. Those auctions used to fol-
Right now, low a “first-price” mechanism, in which advertisers paid whatever amount they bid. But “what happened very quickly,” Parkes explains, “was that people developed these bidding robots that tried to bid just high enough to keep the same place on the page.” Bidding wars ensued, leading to wasteful computation as the bids were constantly adjusted. The software systems running the search engines were completely overloaded.

So Google began to hold “second-price” auctions, in which advertisers paid the next-highest bid rather than their own price. That made counter-speculation less useful and it became sensible to be straightforward about what price each advertiser was willing to pay. The sawtooth cycles of sharply rising and falling bids stabilized. “That’s the kind of design question you can ask,” Parkes says. “If my world consists of rational or almost-rational economic agents, how might we change the rules by which resources are allocated or prices set, so that we can make things more stable and well-behaved? And try to actually simplify the reasoning.”

Numerous challenges lie ahead. Not least is the limit to computational capability. “We don’t claim that AI will ever be perfectly rational,” Parkes points out, “because we know that there are always intractable computational problems. And AI may deviate from rationality in its own ways, differently from people, and we’re just beginning to understand what that might mean.”

Another perennial challenge is the interface with intractably irrational humans. That cuts both ways, though, Parkes notes: for all their rationality, computers lack common sense, and their human designers sometimes fail to anticipate interactions that bring on unexpected consequences. (A robot price war in 2011 caused an out-of-print biology text about flies to be listed for $23 million on Amazon.)

The new phenomena of AI economic systems may in fact require a new science, Parkes says. “For instance, how would you verify not only that a system is doing the right thing, but that it will always do the right thing? We’d also have to agree as a society what ‘right’ means: Should it be fair? Should it be welfare-maximizing? Should it respect laws?” New laws might be required, he continues: “Who would be liable if your agent makes a transaction that leads somebody to die in a chain of consequences that would have been very hard to anticipate?”—for example, by proactively buying a drug for possible future profit, and in the process depriving someone who needs the medication right away. “So,” he says, “things are quite complicated.”

Numerous challenges lie ahead. Not least is the limit to computational capability. “We don’t claim that AI will ever be perfectly rational,” Parkes points out, “because we know that there are always intractable computational problems. And AI may deviate from rationality in its own ways, differently from people, and we’re just beginning to understand what that might mean.”

Another perennial challenge is the interface with intractably irrational humans. That cuts both ways, though, Parkes notes: for all their rationality, computers lack common sense, and their human designers sometimes fail to anticipate interactions that bring on unexpected consequences. (A robot price war in 2011 caused an out-of-print biology text about flies to be listed for $23 million on Amazon.)

The new phenomena of AI economic systems may in fact require a new science, Parkes says. “For instance, how would you verify not only that a system is doing the right thing, but that it will always do the right thing? We’d also have to agree as a society what ‘right’ means: Should it be fair? Should it be welfare-maximizing? Should it respect laws?” New laws might be required, he continues: “Who would be liable if your agent makes a transaction that leads somebody to die in a chain of consequences that would have been very hard to anticipate?”—for example, by proactively buying a drug for possible future profit, and in the process depriving someone who needs the medication right away. “So,” he says, “things are quite complicated.”

David C. Parkes
Website:
http://econcs.seas.harvard.edu

Mimicking Organs

Could tiny, translucent chips that mimic human organs replace animal testing for drug development? That reality may be coming, according to researchers at the Wyss Institute who have developed organs-on-chips: flexible polymer microchips (about the size of a computer flash drive) that provide a window into the tissue structures, functions, and mechanical motions of lungs, intestines, kidneys, and other organs.

Designed by Folkman professor of vascular biology and Wyss Institute director Donald E. Ingber, Tarr Family professor of bioengineering and applied physics Kevin “Kit” Parker, and former Wyss technology development fellow Dan Huh, these chips have potential to accelerate drug discovery, decrease drug-development costs, and create a future of personalized medicine to treat a wide range of diseases including cancer, liver failure, pulmonary thrombosis, and asthma. Testing drugs in animals has proven problematic due to fundamental biological differences, Ingber says, and the traditional method of testing human tissue in the laboratory can also be challenging, because cells often die or fail to work normally once removed from their context in the body.

The microchips “are much more realistic models of the human body than flat layers of cells grown in petri dishes,” Ingber explains. “We have a window on molecular-scale activities inside living, human cells, in the physical context of a tissue and an organ that we can watch in real time”—insights “that are very difficult to get with an animal.”

Instead of moving electrons through silicon, the contents of the translucent chips push small quantities of chemicals past cells from lungs, intestines, livers, kidneys,