



Computers can do more than win at chess—a new algorithm now allows them to best humans at cooperative games like “prisoner’s dilemma.”

PhonlamaiPhoto/iStockphoto

Computers learn to cooperate better than humans

Science Mag

By [Jackie Snow](#) Mar. 28, 2017 , 1:15 PM

For the first time, computers have taught themselves how to cooperate in games in which the objective is to reach the best possible outcome for all players. The feat is far harder than training artificial intelligence (AI) to triumph in a win-lose game such as chess or checkers, researchers say. The advance could help enhance human-machine cooperation.

Twenty years ago, a supercomputer bested the then–reigning world chess champion Garry Kasparov. More recently, AI researchers have developed programs that can beat humans at more computationally demanding games, such as [Go](#) and [poker](#). But those are all winner-take-all or “zero-sum” games, in which one player wins and everybody else loses. Researchers have done less work on cooperative games in which the goal is for players to work together to optimize the outcome for everyone involved—even if logic demands that a player could improve his or her personal outcome by “betraying” the other players.

Such contests include chicken—the game in which two cars drive toward each other and swerve out of the way at the last minute—and the game theory classic the prisoner’s dilemma, in which two people are charged with a crime. Each can receive a light sentence—say 1 year—if both remain loyal to each other and deny the crime. If one prisoner betrays the other, they’ll go free while their partner gets a long term, perhaps 3 years. If both rat on each other, the prisoners get an intermediate sentence of 2 years. Play a single round, and logic demands that a player betray his partner. Play the game repeatedly, however, and people can learn to cooperate to get the lightest sentence of a year.

Jacob Crandall, a computer scientist at Brigham Young University in Provo, Utah, and colleagues wanted to see whether machines could learn to play such games. So the researchers got humans and computers together to play computerized versions of chicken, prisoner’s

dilemma, and another collaborative strategy game called “alternator.” Teams consisted of two people, two computers, or one human and one computer. Researchers tested 25 different machine-learning algorithms, AI programs that can improve their performance by automatically searching for correlations between their moves and results.

To the scientists’ chagrin, no algorithm was capable of collaborating. But then they turned to evolutionary biology for inspiration. Why not, they thought, introduce a key element of human cooperation—the ability to communicate? So they added 19 prewritten sentences—such as “I’m changing my strategy,” “I accept your last proposal,” or “You betrayed me,”—that could be sent back and forth between partners after each term. Over time, the computers had to learn the meaning of these phrases in the context of the game using their learning algorithm.

This time, one of the 25 algorithms, dubbed S# (pronounced S sharp), stood out. When given a description of a previously unknown game, it learned to cooperate with its partner in just a few turns. And by the end of the game, the machine-only teams worked together almost 100% of the time, whereas humans cooperated an average of about 60% of the time. “The machine-learning algorithm learned to be loyal,” Crandall says.

Such dependability could be a boon for algorithms that learn to make decisions for autonomous cars, drones, or even weapons on the battlefield. “[So far] cooperation [like this] hasn’t been a goal,” of most AI research, says Danica Kragic, a roboticist at KTH Royal Institute of Technology in Stockholm. Instead, she adds, most work has focused on creating autonomous technologies that can surpass human abilities, from facial recognition to playing poker. “Machines need to do more than compete,” says Crandall, who adds that research in robotics—which does a better job of emphasizing cooperation—could serve as a model for AI going forward.