Researchers give robots the capability for deceptive behavior

September 9 2010

Georgia Tech Regents professor Ronald Arkin (left) and research engineer Alan Wagner look on as the black robot deceives the red robot into thinking it is hiding down the left corridor. Credit: Georgia Tech Photo: Gary Meek

A robot deceives an enemy soldier by creating a false trail and hiding so that it will not be caught. While this sounds like a scene from one of the Terminator movies, it's actually the scenario of an experiment conducted by researchers at the Georgia Institute of Technology as part of what is believed to be the first detailed examination of robot deception.

"We have developed algorithms that allow a robot to determine whether it should deceive a human or other intelligent machine and we have designed techniques that help the robot select the best deceptive strategy to reduce its chance of being discovered," said Ronald Arkin, a Regents

professor in the Georgia Tech School of Interactive Computing.

The results of robot experiments and theoretical and cognitive deception modeling were published online on September 3 in the *International Journal of Social Robotics*. Because the researchers explored the phenomena of robot deception from a general perspective, the study's results apply to robot-robot and human-robot interactions. This research was funded by the Office of Naval Research.

In the future, robots capable of deception may be valuable for several different areas, including military and search and rescue operations. A search and rescue robot may need to deceive in order to calm or receive cooperation from a panicking victim. Robots on the battlefield with the power of deception will be able to successfully hide and mislead the enemy to keep themselves and valuable information safe.

"Most social robots will probably rarely use deception, but it's still an important tool in the robot's interactive arsenal because robots that recognize the need for deception have advantages in terms of outcome compared to robots that do not recognize the need for deception," said the study's co-author, Alan Wagner, a research engineer at the Georgia Tech Research Institute.

For this study, the researchers focused on the actions, beliefs and communications of a robot attempting to hide from another robot to develop programs that successfully produced deceptive behavior. Their first step was to teach the deceiving robot how to recognize a situation that warranted the use of deception. Wagner and Arkin used interdependence theory and game theory to develop algorithms that tested the value of deception in a specific situation. A situation had to satisfy two key conditions to warrant deception -- there must be conflict between the deceiving robot and the seeker, and the deceiver must benefit from the deception.

Once a situation was deemed to warrant deception, the robot carried out a deceptive act by providing a false communication to benefit itself. The technique developed by the Georgia Tech researchers based a robot's deceptive action selection on its understanding of the individual robot it was attempting to deceive.

The black robot intentionally knocked down the red marker to deceive the red robot into thinking it was hiding down the left corridor. Instead, the black robot is hiding inside the box in the center pathway. Credit: Georgia Tech Photo: Gary Meek

To test their algorithms, the researchers ran 20 hide-and-seek experiments with two autonomous robots. Colored markers were lined up along three potential pathways to locations where the robot could hide. The hider robot randomly selected a hiding location from the three location choices and moved toward that location, knocking down colored markers along the way. Once it reached a point past the markers, the robot changed course and hid in one of the other two locations. The

presence or absence of standing markers indicated the hider's location to the seeker robot.

"The hider's set of false communications was defined by selecting a pattern of knocked over markers that indicated a false hiding position in an attempt to say, for example, that it was going to the right and then actually go to the left," explained Wagner.

The hider robots were able to deceive the seeker robots in 75 percent of the trials, with the failed experiments resulting from the hiding robot's inability to knock over the correct markers to produce the desired deceptive communication.

"The experimental results weren't perfect, but they demonstrated the learning and use of deception signals by real robots in a noisy environment," said Wagner. "The results were also a preliminary indication that the techniques and algorithms described in the paper could be used to successfully produce deceptive behavior in a robot."

While there may be advantages to creating robots with the capacity for deception, there are also ethical implications that need to be considered to ensure that these creations are consistent with the overall expectations and well-being of society, according to the researchers.

"We have been concerned from the very beginning with the ethical implications related to the creation of robots capable of deception and we understand that there are beneficial and deleterious aspects," explained Arkin. "We strongly encourage discussion about the appropriateness of deceptive robots to determine what, if any, regulations or guidelines should constrain the development of these systems."