Researchers Reveal the Neuronal Computations Governing Strategic Social Interactions in the Human Brain

May 19 2008

In a strategic game, the success of any player depends not just on his or her own actions, but on the behavior of every other player in the game. To be successful, players must not only pay attention to what other players do, but also how they are thinking.

Understanding how the brain functions during this strategizing is at "the core of studies of adaptive social intelligence," says John P. O'Doherty of the California Institute of Technology and the subject of a recent series of brain studies by O'Doherty and his colleagues that offer new insight into how the brain works in social situations.

O'Doherty, an associate professor of psychology, along with graduate student Alan N. Hampton and Peter Bossaerts, William D. Hacker Professor of Economics and Management and professor of finance at Caltech, had volunteers participate in a simple two-player strategy game. In the game, volunteers were assigned to play either the role of an employer or an employee, and were isolated from one another. One of them was placed within a functional MRI machine, which measured brain activity in real time.

During each trial, the participant acting as the employee had to choose to either work or not work, and the employer had to decide to check up on his employees ("inspect"), or to leave them alone ("not inspect"). What each party selects will depend both on their own goals and on their anticipation of the behavior of the other participant. For example, the employer, who is busy with other work, does not want to waste time checking on his employees, but does want to occasionally inspect them to make sure they're not slacking off--and to let them know he's watching. His employees, however, prefer to shirk their duties as much as possible when not being inspected, but would rather be found working when the employer decides to check on them.

To provide real incentive for the volunteers, they were given a small monetary award based on the outcome of each trial. An employer's maximum payoff, $1.00, occurs when his employees are working, and he does not check on them; employees in this situation earn nothing. An employer's worst payoff--$0.00--is assigned either when his employees are working and his mistrust leads him to inspect anyway, or when his employees are shirking and he does not inspect; in both scenarios, the employees earn $.50. In the final case, employees who are caught shirking get nothing, while their boss gets $.25.

The game is set up such that there is no possibility of a tie, with an equal payoff for both players. "It's a competitive situation," O'Doherty explains, "so each person has to keep anticipating and predicting the behavior of the other person, to outguess them"--and maximize their own profit.

"The whole point of this game is the idea that in order to do well, you have to predict what the other player is going to do, and for that you need to know what the other player thinks you are going to do," says O'Doherty.

This type of thought, in which a person creates a mental representation of the thoughts of another person, is known as mentalizing. "We're trying to understand the rules that the brain uses to make these representations. How do I take my perception of what you've done and then use that to work out what to do next? How does the brain transform information and then produce behavior?" O'Doherty says.

O'Doherty and his colleagues used a simple mathematical model that can solve such a game by taking into account the history of the opponent's choices to work out what that opponent is likely to do next. They found that subjects' actual choices could be predicted well by such a model. Furthermore, a number of brain areas previously implicated in mentalizing, such as the superior temporal sulcus (STS) and medial prefrontal cortex (mPFC), showed changes in their activity over time. These changes are predicted by the mathematical model, suggesting that the brain itself uses mathematical operations similar to that encapsulated in the model to solve the task.

O'Doherty and his colleagues found that activity in the mPFC changed depending on the subjects' success in past trials, while activity in STS reflected how that success compared with how well they thought they'd do. Furthermore, activity in the two brain areas appeared to be linked. "If the subjects were surprised by their prediction success"--if, say, they did more poorly than they had expected--"we saw increased activity in the STS. At the same time, there was increased correlation between activity in the STS and the mPFC," O'Doherty says.

This suggests, he says, that "the STS modulates the activity of the mPFC, telling it to refine its expectations, which can ultimately lead to a change in the subjects' behavior."

"This research is getting at the essence of how the brain functions in social situations," he says. Such studies could, therefore, eventually shed light on disorders, such as autism, that involve problems with social interaction.

The paper, "Neural correlates of mentalizing-related computations during strategic interactions in humans," was published in the May 6 issue of the *Proceedings of the National Academy of Science*. The work was partially funded by the Gordon and Betty Moore Foundation.

Source: Caltech