Jean-François Bonnefon
Toulouse School of Economics

"Experimental Ethics for Driverless Cars"

Monday, November 30, 2015
2:00-3:30pm
CIT Building, Lubrano Conference Room 477

The wide adoption of self-driving, Autonomous Vehicles (AVs) promises to dramatically reduce the number of traffic accidents. Some accidents, though, will be inevitable, because some situations will require AVs to choose the lesser of two evils: for example, running over a group of pedestrians or sacrificing their own passenger by driving into a wall. It is a formidable challenge to define the algorithms that will guide AVs confronted with such moral dilemmas. These moral algorithms will need to accomplish three potentially incompatible objectives: being consistent, not causing public outrage, and not discouraging buyers. We argue that to achieve these objectives, manufacturers and regulators will need psychologists to apply the methods of experimental ethics to situations of unavoidable harm. We report five studies showing the classic signature of a social dilemma: people are morally favorable to utilitarian AVs, programmed to minimize the death toll in case of unavoidable harm - but they are uncomfortable with driving one themselves.

Jean-François Bonnefon is a cognitive psychologist at the Toulouse School of Economics (France). He is the author of more than 100 publications on the rationality of human thinking and behavior. His research appeared in a broad range of scientific outlets, from computer science to philosophy, linguistics and economics. His recent work applies the insights of moral psychology and behavioral economics to the new challenges of machine ethics and human-computer cooperation.

For more information on this talk and the HCRI Speaker Series, contact hcri@brown.edu or visit hcri.brown.edu.