Eugene Wigner Colloquium

joint event of GRK 1558 and SFB 910

Prof. Jacob Crandall

Masdar Institute of Science and Technology, United Arab Emirates

"Regulating Highly Automated Machine Ecologies"

Much of the world's systems and critical infrastructure now consist of human-machine ecologies. Such societies are or will likely be composed of complex networks of people and autonomous machines. In this talk, I will discuss two aspects of my research group's work with respect to such societies. First, I will discuss our work toward developing a theory of the regulation of human-machine ecologies. These human-machine societies are strongly influenced by two major forces: automation and regulation. On one hand, increasingly sophisticated control algorithms govern the behavior of individual machines in the society. On the other hand, a regulatory authority (typically a person or organization) is tasked with setting rules and incentives that promote system-wide stability and efficiency. Thus, we have studied how regulatory power and algorithmic sophistication jointly impact the regulability and controllability of these systems.

Second, we have studied how cooperative relationships can be established and maintained between individual nodes of the human-machine network. Efforts in Artificial Intelligence (AI) and Machine Learning have traditionally focused on matching or outperforming humans in difficult cognitive tasks (e.g. face recognition, personality classification, driving cars, or playing video games) or defeating humans in strategic zero-sum encounters (e.g. Chess, Checkers, Jeopardy!, or Poker). In contrast, less attention has been given to developing autonomous machines that establish mutually cooperative relationships with humans even when the self-regarding preferences of humans and machines are somewhat in conflict. We have developed a new learning system that we have developed that rivals human cooperation in two-player repeated interactions. This is the first general-purpose algorithm that is capable, given a description of a previously unseen game environment, of learning human-level cooperation within short timescales. It does so without pre-programming of well-known, game-specific strategies (e.g. tit-for-tat), thus enabling human-AI cooperation in scenarios previously unanticipated by algorithm designers.

GRK1558 research training group

Thursday, 07.01.16 · 16:15h · EW 202

Technische Universität Berlin · Institut für Theoretische Physik · Hardenbergstraße 36 · 10623 Berlin
www.itp.tu-berlin.de/grk1558 · www.itp.tu-berlin.de/sfb910