Symposium "Norms and commitments in human-robot cooperative interactions"

12th Collective Intentionality conference, Neuchâtel, on July 13-16, 2020.

Convenor:
Elisabeth Pacherie (Institut Jean Nicod, PSL, Paris)
Speakers:
Ingar Brinck (Lund University)
Raul Hakli (University of Helsinki)
John Michael (Centreal European University, Vienna)

General presentation

Recent developments in social robotics aimed at improving cooperative human-robot interactions have brought to light a number of important challenges to that enterprise. These challenges fall into two main categories: motivational and predictive.

Humans are social animals that exhibit a robust motivation to engage with others that has a variety of sources, both endogenous (e.g. need to belong or general pro-social tendency) or exogenous (e.g. social pressure). This motivation plays a crucial role in explaining why people engage in joint action with human partners and why they may remained engaged in a joint action even when more attractive options emerge. In contrast, humans appear much less motivated to engage in joint actions with robots, exhibiting negative attitudes (e.g., the *Uncanny Valley Effect*) and forms of distrust toward robots..

Besides motivation, successful cooperative interactions are also premised on prediction. Agents need to coordinate their actions at various levels and to do must be able to make accurate predictions regarding their partner's actions and their consequences. In human-human interactions a variety of processes and devices, ranging from automatic processes of motor resonance all the way to explicit communication and commitments, help us predict the actions of our partners. However, research on human-robot interaction suggests that prediction can be a serious challenge for human-robot interactions. Robots are often cognitively opaque to their human partners. For instance, several findings in psychology and neuroscience suggest that humans interact differently and do not deploy the same range of predictive processes when their partner is a robot rather than a human (Sahaï et al., 2017; 2019). In addition, the gap between the expected and the actual capabilities of the robot may also impact the predictive capacities of humans.

The main purpose of this symposium will be to assess the extent to which an appeal to norms and commitments in human-robot cooperative interactions may help mitigate these motivational and predictive challenges, what the benefits or drawbacks of such an approach could be, and how it compares to other strategies, such as strategies that emphasize the need to built robots whose appearance and behavior appeal to human positive emotions and exploit human tendencies towards anthropomorphism or strategies that emphasize the need to endow robots with an extensive range of human-like social cognition abilities.

Ingar Brinck: Social norms in Human-Robot interaction

Abstract:

Social norms are spontaneously emergent patterns of coordinated behaviour that organize how individuals behave towards each other in accordance with social expectations about what and how an individual ought to do in a given situation. They improve how agents collectively

manage, and reduce the cognitive costs associated with interaction generally. Basing HRI in social norms can be expected to have advantages that would include meeting the motivational and predictive challenges. However, social norms present a challenge for HRI, being notoriously difficult to implement. I will discuss the advantages of an approach to HRI based in social norms, granted that the implementation problem can be handled in a satisfactory way, and compare it to other approaches. Then, I will briefly address the implementation problem and sketch a way of dealing with it.

Raul Hakli: Trusting a robot collaborator

Abstract:

It is often thought that collaboration requires that the collaborators can trust each other. I will study what it could mean to trust robots in the context of human-robot collaboration. In philosophical literature, it is commonplace to distinguish between reliance and trust. When we rely on someone to do something we act on the premiss that they will do it. Reliance on machines like robots is nonproblematic, but proper trust requires more than mere reliance. Roughly, it requires some kind of a normative expectation that the trusted will do what they ought to do. I will look at various theories of trust that spell out this normative component in terms of motivations, commitments, or obligations, respectively. I will study what it might mean to trust a robot in light of such accounts. I will argue that while it is psychologically possible to trust robots, such trust may not be fully appropriate in the sense that the normative expectations do not seem to be applicable to robots. When we say that we trust robots, it might be better to understand that to mean that we rely on them to work as expected but we also have certain beliefs about the design, implementation, and use of those robots. Hence, the normative component of trust is ultimately applied to people instead of the robots themselves.

John Michael: The Sense of Commitment in Human-Robot Interaction

Abstract:

In this talk I spell out the rationale for developing means of manipulating and of measuring people's sense of commitment to robot interaction partners. A sense of commitment may lead people to be patient when a robot is not working smoothly, to remain vigilant when a robot is working so smoothly that a task becomes boring and to increase their willingness to invest effort in teaching a robot. Against this background I will present a set of studies that have been conducted to probe various means of boosting people's sense of commitment to robot interaction partners, and raise discuss the implications for our psychological and normative understanding of commitment.