Misdirection and Counter-deception in Robot Teams: Methods and Ethical Considerations

Ronald C. Arkin

An effective team requires trust, dependability, cohesion, and capability. These attributes are the same for teams of robots. When multiple teams with competing incentives are tasked, a strategy, if available, may be to weaken, influence or sway the attributes of other teams and limit their understanding of their options. Such strategies are found in nature and in sports such as feints, misdirection, etc. Here, we focus on one class of strategies for multi-robots: intentional misdirection using shills/confederates, and ethical considerations associated with deploying such teams. The use of intentional misdirection must be anticipated as robot teams become more autonomous, distributed, networked, numerous, with more capability for critical decisions.

This NSF-funded project studies strategies to enable robots and robot teams to model, generate, cope, and counter misdirection. It offers a novel approach to resilience in these teams to possible disruption. Computational models provide a framework for understanding, producing, and countering misdirection in robotic systems. They have been designed using behaviors at the individual and team levels, building on decentralized methods of control and communication.

The relationship of misdirection and its relationship to intelligence is well documented. Indeed, the Turing test, a hallmark measure of artificial intelligence, is based on confusing a human with a computer. Deception is believed to play a significant role in Human-Human interaction, and thus also has a place in Human-Robot Interaction: *"The development of deception follows the development of other skills used in social understanding" [Vasek]. "Another price you pay for higher-order intentionality is the opportunity [for] ... deception " [Dennett].*

Robot teams that use misdirection provide the ability to confuse, obscure, and execute novel behaviors that no single agent could provide. Heterogeneity arises not only of a single deceiving agent, but also shills, which support misdirection indirectly. The goal may include: inducing a misperception of intent; masking the movement of the team; or a miscalculation of the numbers of the team and dispositions.

We have conducted extensive prior work on robot deception for individual robots, including using these agents to feign strength where there is none, feint or mislead, and provide support for those in distress among others. This research has led to the development of the first taxonomy of human-robot deceptive activities, including misdirection. Here, we consider *team* misdirection, from organizational models drawn from sports, the military, biology and other relevant disciplines.

Ethical considerations have played an important role and continue to do so in our ongoing research. Limited, if any, research has been conducted on coordinated robot team misdirection, especially when using robots of differing capabilities, let alone the ethical aspects of group deception. Several researchers have studied deception in a robotics context, but few have considered the ethical consequences. While benefits are clearly apparent to the team performing the deception, ethical questions surrounding misdirection or other forms of deception are real and this talk will address these issues.