Preface

Trust is a key issue in the development and implementation of autonomous systems working with humans. Humans must be able to trust the actions of the machines to want to work with them, and machines must develop or establish trust in the actions of human coworkers to ensure effective collaboration. There is also the issue of autonomous agents, robots and systems trusting one another and humans.

But trust can mean different things in different contexts. For flight control systems on airplanes, trust may mean meeting rigorous criteria regarding structural qualities of the airplane, flightworthiness, and a provably stable control system. In the context of humans interacting with humanoid robots, trust may more closely relate to the interdependence between the human and robot in correctly reading and interpreting each other's voice commands and gestures and observed actions, and the likelihood that both the robot and human will do what is expected of each other. In the context of an autonomous automobile carrying passengers, trust in the system may be the expectation that the system will respond correctly not only to foreseen road and traffic conditions, but also to unusual circumstances (for example, gridlock; alternative route planning; a child running into the street; running out of gas on the highway; an engine catching fire; hearing a fire engine or ambulance with siren blaring; or a flat tire causing the vehicle to swerve). Interdependent trust also includes system controllers and society. System controllers, human or machine, must be able to control at the individual, group and system levels; and society must be willing to entrust its citizens, including the elderly and young, to the system.

This symposium will explore the various meaning aspects and meanings of trust between humans and machines in various situational contexts, and the social dynamics of trust in teams or organizations composed of autonomous machines working together with humans. We will seek to identify and/or develop methods for engendering trust between humans and autonomous machines, to consider the static and dynamic aspects of trust, and to propose metrics for measuring trust.

This AAAI symposium addresses these specific topics and questions: What are the connotations of "trust" in various settings and contexts? How do concepts of trust between humans collaborating on a task differ from [human and machine], [machine and human], and [machine and machine] trust relationships? What metrics exist for trust between individuals, and how well do these translate to trust relationships between humans and autonomous machines? What metrics for trust currently exist for evaluating machines (possibly including such factors as reliability, repeatability, intent, and susceptibility to catastrophic failure) and how may these metrics be used to moderate behavior in collaborative teams including both humans and autonomous machines? How do trust relationships affect the social dynamics of human teams, and are these effects quantifiable? What validation procedures could be used to engender trust between a human and an autonomous machine? What algorithms or techniques are available to allow machines to develop trust in a human operator or another autonomous machine?

Other topics included are computational models of trust in autonomous system; trust model between a single human and a single robot; the effect of trust on team social dynamics; verification and validation of autonomous system behaviors; human requirements for trust in machines; methods for engendering trust between humans and machines ; metrics for established trust; metrics for deception in humans and machines; and other computational and heuristic models of trust relationships, and related behaviors, in teams of humans and machines.

Keynote speakers at the symposium include John Lee (University of Wisconsin), Missy Cummings (ONR/MIT), Jeff Bradshaw (IHMC), Holly Yanco (University of Massachusetts, Lowell), Ron Diftler (NASA/JSC), and Jim Hansen (NRL Monterey).

*Don Sofge, Geert-Jan Kruijff, and William F. Lawles*s
Symposium Cochairs