Trust and Interdependence in Controlling Multi-Agent Multi-Tasking Autonomous Teams

William F. Lawless[1] and **Donald A. Sofge**[2]

[1]Paine College, 1235 15th Street, Augusta, Georgia, wlawless@paine.edu

[2]Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC, donald.sofge@nrl.navy.m

Abstract

In this report we address the role of trust in autonomous systems, and our progress in developing a theory of interdependence for the efficient control of hybrid teams and systems composed of robots, machines and humans working interchangeably. Sentient multi-agent systems require an aggregation process like data fusion. But conventional use of fusion for the control of UxV systems hinges on convergences to form patterns, increasing uncertainty. Present solutions appear to indicate stability for cooperative contexts and instability for competitive ones, in line with our theoretical expectations.

Introduction

Trust is a key issue in the development and implementation of autonomous systems working with and for humans. Humans must be able to trust the actions of the autonomous machines to want to work with them, and autonomous machines must be able to develop or establish trust in the actions of human co-workers. This trust between and among hybrid agents must be extended in a manner that ensures efficient and effective communication, collaboration and the free flow of information without increasing barriers between robots, machines and humans.

Trust can mean different things in different contexts. For flight control systems on airplanes, trust may mean meeting rigorous criteria regarding the structural qualities of an airplane, flight worthiness, and control system stability. In the context of an autonomous automobile carrying passengers, trust in the system may be the expectation that the autonomous robot will respond correctly not only to foreseen road and traffic conditions, but also to unusual circumstances (e.g., gridlock; alternative route planning; a child running into the street while chasing a ball; running out of gas on the highway; an engine catching fire; hearing and seeing an approaching fire engine or ambulance with siren blaring; or a flat tire causing the vehicle to swerve unexpectedly).

In the context of teams where multi-tasking occurs with hybrid teams, trust may more closely relate to the management of the interdependence among teammates in correctly sensing, reading and interpreting each other's voice commands, gestures and observed actions to increase the likelihood that the hybrid teammates do what is expected of each other. 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 a multi-tasking hybrid system composed of autonomous agents and humans working together. However, the control of interdependent teams has not yet been solved (Jamshidi, 2009). But when it is solved, we expect to find that bidirectional trust becomes an interdependence among sentient agents, each capable of reacting to one other's actions in hybrid teams, systems and society.

When does trust arise? In a single agent or system composed of independent agents, trust occurs when an agent, including humans, is performing satisfactorily over a range or set, S, of behaviors, from the maximum of underperformance (infimum) to the minimum of over-performance (supremum), a range that we designate as governed by set-points (from Diener, 1984). The entropy, H, increases as n becomes equiprobable to $p = 1/n$, giving $H = \log n$. Based on Shannon's information theory, competitive systems, composed of independent players, generate more information than cooperative systems, where one agent is dependent on another. But for interdependent agents, we expect Shannon's theory of information to be replaced with Von Neumann's (Gershenfeld, 2000). Under competitive situations among sentient agents, alternative viewpoints spontaneously arise under uncertainty to form a superposition of both.

When alternative viewpoints about reality arise, what has trust to do with multi-robot or hybrid systems? For a

human group or firm (of multi-taskers) seeking to increase its competitiveness to gain an advantage for its clients with an approach that reduces entropy (for knowledge, H=0; in Conant, 1976); e.g., if an algorithm can predict excellence in the medical choices that are made, it will increase trust and value in the services provided by that team.

Providing intelligent second opinions raises an important issue with advanced robots working interdependently with humans. As stated earlier, human teams work together to solve two broad classes of problems: those solvable problems that require cooperation for efficiency, consuming the information already available by following existing algorithms, laws, or procedures to reduce uncertainty and increase stability; and those unsolved or intractable problems that require competition, generating new information to solve a complex problem(s), but increasing uncertainty, instability and disruption in the process. The latter is characterized by a competition between alternative viewpoints.

Smallman (2009) has concluded that convergences can be challenged with alternative viewpoints. His system tracks agreement and disagreement among users in a system (e.g., submarine). Yet no known method to compute and display alternative viewpoints exists yet. But by constructing orthogonal pro-con vectors during the sensory fusion process, a tool to display alternative viewpoints could mitigate mistakes in the control of hybrid systems (Lawless et al., 2011).

We hope to build on Smallman's (2009) work to reduce convergence processes. The standard JDL fusion model also uses convergence processes; e.g., Llinas et al. (2004) highlight the value of belief consistency (p. 6) to build a "community consensus" (p. 13). But alternative beliefs are permitted in the JDL Fusion model (Steinberg et al.,1999). Thus, replacing Smallman's *non-computational* approach with a mathematical model based on orthogonal beliefs advances the science of fusion and decision making.

Theory

Needed for hybrid or pure robot teams is a transactions or exchange model that tracks bidirectional sensory effects and interdependent uncertainty. The end result should be collective control theory; e.g., the waggle dance performed by interdependent bees exemplify the exchanges known as quorum sensing (Sasaki & Pratt, 2011).

As social uncertainty increases, bistable interpretations, a mixture of reality and illusions, spontaneously arise among agents. Reactance, ρ, against illusions serves to drive social oscillations; e.g., volatility (stock markets, mobile phone churn, divorce rates). Here, ρ becomes the seed to create new organizations (from IBM comes Apple, MS). But how to model (e.g., illusions, debate, oscillations, resonance)?

Consider a social model of competitive debates (e.g., in politics, courtrooms, science). Let two speakers each represent an organization, with an audience of neutrals in front of both. To model debate, the two different views do not commute; i.e., $[A,B]=iC;$ where i represents phase space, and C a gap in the Knowledge, K, of Reality, R, where K implies that H-> 0 in the limit. A social model of debate is with an inverted Prisoners Dilemma Game: D-D improves social welfare (competition), C-C reduces it (cooperation); i.e., successful debate increases social welfare (increasing social ΔA; further, the winning organization out-gains ΔA). **Conjecture**: Despite open conflict (+H), Democracies solve problems better than autocracies.

Conclusion

For autonomous agents, a focus on one thing at a time implies interdependent tradeoffs under uncertainty, reducing the ability to multitask. The purpose of autonomous teams is to multitask. But for users to trust a multi-tasking team, it must know how the team performs under both cooperative and competitive situations. In this effort we explore the role of trust and interdependence among agents in cooperative and competitive contexts.

Acknowledgements: This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U. S. Army Research Office under contract/grant number W911NF-10-1-0252.

References

Conant, R. C. 1976. Laws of information which govern systems. IEEE Trans. Systems, Man, and Cybernetics 6: 240-255.

Diener, E. 1984. Subjective well-being, Psychological Bulletin, 95(3): 542-575.

Gershenfeld, N. 2000. The physics of info. techn. Cambridge U.

Jamshidi, M. 2009. Control of system of systems. Intelligent Control Systems. T. Nanayakkara, Sahin, F., & Jamshidi, M. (Eds.). London, UK, Taylor & Francis. Vol 2 (Ch. 8)

Lawless, W. F., Angjellari-Dajci, F., Sofge, D. Grayson, J., Sousa, J. L. & Rychly, L. 2011. A New Approach to Organizations: Stability and Transformation in Dark Social Networks, J. Enterprise Transformation, 1:4, 290-322.

Llinas, J., Bowman, C. L., Rogova, G., Steinberg, A. N., Waltz, E., White, F. E. 2004. Rev. and Extensions to JDL Data Fusion Model II. Proc 7th Int'l Conf Info Fusion, Sweden, pp. 1218–30.

Sasaki, T., Pratt, S. 2011. "Emergence of group rationality from irrational individuals." Behavioral Ecology 22: 276-281.

Smallman, R., Roese, N. J. 2009. Counterfactual thinking facilitates behavioral intentions. Journal of Experimental Social Psychology, 45 (4), pp. 845–852.

Steinberg, A.N., Bowman, C. L., White, F. E. 1999. Revisions to the JDL data fusion model. Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE 3719.