

**THE THERMODYNAMICS OF AUTONOMOUS HUMAN-MACHINE TEAMS (A-HMT):
CONTROL OR GOVERNANCE?**

W. F. Lawless

Professor, Paine College, Augusta, GA, USA
Summer Research Fellow, Naval Research Laboratory, Washington, DC, USA

ABSTRACT

In late 2018, a “formidable” Russian base in Syria was attacked by a “highly sophisticated” swarm of drones; Russian soldiers were killed and aircraft destroyed, but Russia denied deaths occurred, and the U.S. denied involvement. The drones in this attack were thought to be controlled individually over a long distance by humans guided by GPS. More autonomous drones can control themselves like a flock of birds without a leader or following one. But as full autonomy for machines is realized for a human-machine team working together to solve problems confronted by uncertainty, the team will begin to face the same issues of control human teams experience.

Control has been a part of the human-centered design (HCD) that has dominated Systems Engineering for over two decades; HCD, known as “human in the loop,” is the preferred process for Systems Engineering. But human centered activities may not be independent, even for simple acts like driving a car, making HCD once thought to be harmful. Looking towards the future, autonomous systems are considered potentially harmful. For HCD, a common refrain is that “we must always put people before machines, however complex or elegant that machine might be.” Autonomy raises the bar. In the U.S., “Lethal autonomous weapon systems (LAWS) are a special class of weapon systems that use sensor suites and computer algorithms to independently identify a target and employ an onboard weapon system to engage and destroy the target without manual human control of the system.” This concept of autonomy is known as “human out of the loop” or “full autonomy.”

With their goal of control, by rejecting the cognitive model, physical network scientists and game theorists have dramatically improved the predictability of behavior in situations where beliefs are suppressed, in low risk environments, or for economic beliefs under certainty, but the predictability by these models fails in the presence of uncertainty or conflict, exactly where interdependence theory thrives; e.g., the

interdependent effects in debating the possible tradeoffs to choose the next path. Facing uncertainty, debate exploits the bistable views of reality that exist to explore interdependently the tradeoffs that test, or search, for the best paths forward. Generalizing, reducing uncertainty for a system necessitates that human and machine teammates are both able to explain however imperfectly each other's past actions and future plans in causal terms.

In that no single agent can determine social context alone, resolving uncertain contexts requires a theory of interdependence to build and operate safely and ethically A-HMT systems. The best science teams are interdependent. A team's intelligence is its interdependent interactions among its teammates. We extend these findings to the open-ended debates that explore tradeoffs seeking to maximize a system's production of entropy (MEP) in highly competitive uncertain environments.

Finally, human teams cannot be controlled in the technical sense of controlling a swarm of drones. We generalize this insight to conclude that the control of teams must give way to governance.

Keywords: traditional rational theory, interdependence theory, control, governance