The Thermodynamics of Autonomous Human-Machine Teams (A-HMTS): Control Or Governance?

Preprint

W.F. Lawless

Paine College, Augusta, GA, United States

Presented at the Thermodynamics 2.0 | 2020
June 22–24, 2020

International Association for the Integration of Science and Engineering (IAISAE) is a non-profit registered in Colorado, United States.

Conference Paper
IAISAE/CP-T2020-W184
June 2020

The Thermodynamics of Autonomous Human-Machine Teams (A-HMTS): Control Or Governance?

Preprint

W.F. Lawless

Paine College, Augusta, GA, United States

Suggested Citation
Lawless, W.F. 2020. The Thermodynamics of Autonomous Human-Machine Teams (A-HMTS): Control Or Governance?: *Preprint*. Superior, CO: International Association for the Integration of Science and Engineering (IAISAE). IAISAE/CP-T2020-W184. https://iaisae.org/wp-content/uploads/w184.pdf.

International Association for the Integration of Science and Engineering (IAISAE) is a non-profit registered in Colorado, United States.

This article is available at no cost from IAISAE at www.iaisae.org/index.php/publications/

Conference Paper
IAISAE/CP-T2020-W184
June 2020

Proceedings of the IAISAE 2020
International Conference on Thermodynamics 2.0
June 22-24, 2020 | Worcester, MA, USA

ICT2.0:2020-W184

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

W.F. Lawless

Professor of Math & Psychology, Paine College, Augusta, GA, USA
Summer Research Fellow, Naval Research Laboratory, Washington, DC

ABSTRACT

After Russian soldiers at an "impregnable" base in Syria were killed in late 2018 by swarms of drones, the weapons of war included partly autonomous swarms. The future, rapidly advancing, may include fully autonomous machines. But autonomy and control are incompatible, raising the issue of how to implement control. Until now, and still preferred as central to human-centered design (HCD), human control is integral, known as "humans in-the-loop." Contradicting HCD, the pace of decisions that need to be made ever faster than humans can process information for decisions motivates the development of autonomous machines and autonomous human-machine teams (A-HMTs). This new class of weapon systems uses sensors and algorithms to engage and destroy targets without manual human control, i.e.,"human out-of-the-loop." However, with autonomy in uncertain situations, control becomes what? When situations are low risk, well-defined, and with subjective opinions ignored, rational decisions govern, but under conflict or uncertainty, rational models fail. There, interdependence thrives, e.g., under uncertainty, the interdependent effects among humans is commonly expressed by debating tradeoffs to choose a path going forward. Generalizing, reducing uncertainty necessitates that A-HMTs must be able to explain to each other however imperfectly their past actions and future plans in causal terms. In that the best human teams are highly interdependent, to maximize work productivity, A-HMTs must maximize their production of entropy (MEP) to compete in highly uncertain environments. We conclude that the control of teams must give way to governance, the same method used for human-only teams.

Keywords: Interdependence; autonomy; human-machine teams; maximum entropy production (MEP)

1. INTRODUCTION

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 that any deaths had occurred among its soldiers on its base, and the U.S. denied any knowledge or involvement [1]. The drones in this attack were thought to be controlled individually over a long distance by humans assisted by gps guidance (as an example, see the animation by [2]). More and more, autonomous drones can control themselves like a flock of birds without a leader or following one, each drone monitoring the other and with the lead drone overseen by humans (e.g., [3]). But as full autonomy for machines is realized for a human-machine team working together to solve a complex problem confronted by uncertainty [4], the team will begin to face the same issues of control human teams have experienced for eons, and maybe even new ones as attempts are made to fully exploit their complementary assets and skills.

Control has been part of the human-centered design (HCD) that has been dominant in Systems Engineering for more than two decades [5]; HCD, also known as "human in the loop," is today the preferred process for Systems Engineering [6]. However, because a sequence of human-centered activities may not be independent, even for the simple acts entailed in driving a car, HCD was once considered harmful [7]. Presently, looking towards the near future, autonomous systems are considered potentially harmful [8]. For humans interacting with technology, a common refrain is that "we must always put people before machines, however complex or elegant that machine might be." [9] Autonomy raises the bar for complexity, difficulty and threat. 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." [10] This concept of autonomy is known as "human out-of-the-loop" or "full autonomy."

Full autonomy is considered a difficult, complex task to achieve [11]. However, it is also consider to be very controversial (e.g., [8] also includes quotes of the warnings made by the United Nations).

With their goal of control (one of the leading proponents is Barabási; see [12],13],[14]), by rejecting the cognitive model, not only physical network scientists but also game theorists [15] dramatically improve the predictability of behavior in situations where beliefs are suppressed or ignored, in low risk environments, or for economic beliefs in highly certain environments, but the predictability by these models fails in the presence of uncertainty or conflict ([16]; see also [17]), exactly where interdependence theory thrives [4]; e.g., the interdependent effects of debating the possible tradeoffs to solve a problem. For example, facing uncertainty, debate exploits the bistable views held by opposing viewpoints in reality existing naturally and arising spontaneously that humans use to explore interdependently the tradeoffs that test, or search, for the best paths moving 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 (compare the warning in [18] by Pearl, a distinguished Artificial Intelligence scientist, followed by his equally strong warning reiterated in [19]).

In that no single human or machine agent can determine uncertain social contexts alone [20], resolving the uncertainty in a context requires a theory of interdependence to build and operate safely and ethically autonomous human-machine systems. From the literature, the best science teams are fully interdependent (see the broad-based summary of research by the National Academy of Sciences; in [21]). A team's intelligence has been located in the interdependent interactions among its teammates [22]. 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 but uncertain environments [4].

In that the best human teams are highly interdependent [23], to maximize work productivity, thermodynamically, A-HMTs must be able to interdependently maximize their production of entropy (MEP; in [24]) to be able to compete to gain an advantage in highly uncertain environments (see the Introduction in [25]). To achieve MEP requires intelligence [26] in the effective and efficient use of tradeoffs by a team or organization to minimize its structural dissipation of wasted entropy (SEP), akin to focusing a telescope [4]: as SEP decreases, MEP can increase.

2. MATERIALS AND METHODS

This paper is an extended abstract that reviews the theory and empirical research recently completed in late 2019. Here we review a case study introduced in [4] and [20] and discussed herein.

3. RESULTS AND DISCUSSION

Case Study. We consider the human and machine as a system. AI's Machine Learning (ML) can learn patterns sufficiently well-enough to be able to drive a car. However, a self-driving Uber car in 2018 struck and killed a pedestrian. The car acquired the pedestrian in its path ahead at 6s before impact, selected its brakes at 1.3s earlier than impact, but its brakes had been disconnected by Uber's engineers to improve the car's ride. The car's human operator acquired the pedestrian 1s early and hit the brakes 1s after impact, the car striking and killing the pedestrian (see [28] and [29]).

There are several lessons to be drawn from this Case Study. First, albeit with its brakes disconnected, the Uber car performed as it had been designed to perform. Second, the performance of the Uber car was faster with its decision and superior in its performance compared to its human operator. Third, however, the self-driving Uber car was unable to alert its human operator immediately at the same time its sensors had acquired an impending obstacle in the road (i.e., the pedestrian). But by failing to alert its human operator seconds earlier than the human operator became aware of the pedestrian's presence in the roadway, the Uber self-driving car was a poor team player [20]. The latter is an example of dysergy, the opposite of synergy.

4. CONCLUSION

With our literature review and the Case Study that we have provided, it is apparent that self-driving cars as well as self-flying drones are becoming a reality. Presently, the level of autonomy for human-machine teams is poor, but rapidly changing and improving. Once machines are able to explain themselves to their human teammates and to understand the explanations by their human teammates in turn, control for autonomous members human teams cannot be controlled in the technical sense of controlling a swarm of drones. Each teammate will have its own skills; should the roles and skills be orthogonal (complementary) to each other as is common with highly specialized human teams (e.g., at a minimum, a surgical team may consist of a surgeon or two; an anesthesiologist supported by a nurse anesthetist; an operating room nurse; a surgical technician; multiple physician assistants; and several medical students as observers; it may even possibly include a representative from a company that makes medical equipment in the operating room; from [29]), direct control will no longer be feasible. We generalize this insight to conclude that the control of teams must give way to governance [30].

ACKNOWLEDGEMENTS

This work was completed during the Summer of 2020 while the author was a Fellow at the Naval Research Laboratory in Washington, DC.

REFERENCES

[1] Grove, T. (2018, 1/15), "Drone Attacks on Russian Bases in Syria Expose Security Holes. The attacks reveal a fresh threat to Moscow from insurgent rebel groups in Syria even after their broad defeat by Russia and its allies," Wall Street Journal, retrieved 6/20/202 from https://www.wsj.com/articles/drone-attacks-on-russian-bases-in-syria-expose-security-holes-1516017608

[2] RMD (2020, 3/20), "Mind of the swarm. Amazing new technology allows drones to flock together as they fly," Raytheon Missiles & Defense, retrieved 6/30/2020 from https://www.raytheonmissilesanddefense.com/news/feature/mind-swarm

[3] Vásárhelyi, G., Virágh, C., Somorjai, G., Nepusz, T., Eiben, A.E. & Vicsek, T. (2018), Optimized flocking of autonomous drones in confined

environments, *Science Robotics*, 3(20): eaat3536. DOI: 10.1126/scirobotics.aat3536
[4] Lawless, W.F. (2019), The Interdependence of Autonomous Human-Machine Teams: The Entropy of Teams, But Not Individuals, Advances Science, *Entropy* 2019, *21*(12), 1195; https://doi.org/10.3390/e21121195
[5] Cooley, M. (1990), Human-centred systems, in Howard H. Rosenbrock (Editor), Designing Human-centred Technology: A Cross-disciplinary Project in Computer-aided Manufacturing, London: Springer-Verlag, Chapter 10.
[6] Fouad, H.Y., Raz, A.K., Llinas, J., Lawless, W.F. & Mittu, R. (accepted, 2020), Finding the Path Toward Design of Synergistic Human-Centric Complex Systems, IEEE Systems, Man & Cybernetics.
[7] Norman, D.A. (2005, July-August), Human-centered design considered harmful," Interactions, pp. 14-19, retrieved 4/28/2020 from https://dl.acm.org/doi/pdf/10.1145/1070960.1070976
[8] Editors (2019), "Ready for weapons with free will?" New York Times, from https://www.nytimes.com/2019/06/26/opinion/weapons-artificial-intelligence.html."
[9] Cooley M. (2008) On Human-Machine Symbiosis. In S.P. Gill (Ed.), Cognition, Communication and Interaction. Transdisciplinary perspectives on interactive technology. London: Springer, Chapter 26.
[10] CRS (2019, 12/19), Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, Congressional Research Service, retrieved 4/28/2020 from https://fas.org/sgp/crs/natsec/IF11150.pdf
[11] Mattis, J., Secretary of Defense (2018) Nuclear Posture Review 2018, Office of the Secretary of Defense, Department of Defense; retrieved 2018 from https://www.defense.gov/News/SpecialReports/2018NuclearPostureReview.aspx
[12] Barabási, A.-L. (2009). "Scale-free networks: A decade and beyond." Science **Science 325**: 412-413.
[13] Barabási, A. L. (2012). Network science: Understanding the internal organization of complex systems. Invited talk 2012 AAAI Spring Symposium Series, Stanford University, AAAI Publications.
[14] Liu, Y.Y. & Barabási, A.L. (2016), Control principles of complex systems, Reviews of Modern Physics, 88(3): 1-58, DOI: 10.1103/RevModPhys.88.035006
[15] Amadae, S.M. (2016), "Rational choice theory. POLITICAL SCIENCE AND ECONOMICS," Encyclopaedia Britannica, from https://www.britannica.com/topic/rational-choice-theory
[16] Mann, R.P. (2018), Collective decision making by rational individuals, PNAS, 115(44): E10387-E10396; from https://doi.org/10.1073/pnas.1811964115
[17] Hansen, L.P. (2013), Uncertainty inside and outside of economic models, The Nobel Prize, retrieved 4/25;2020 from https://www.nobelprize.org/prizes/economic-sciences/2013/hansen/lecture/
[18] Pearl, J. (2002), "Reasoning with Cause and Effect", AI Magazine, 23(1): 95-111; from https://aaai.org/ojs/index.php/aimagazine/article/download/1612/1511
[19] Pearl, J. & Mackenzie, D. (2018, 5/18), "AI Can't Reason Why. The current data-crunching approach to machine learning misses an essential element of human intelligence", Wall Street Journal, from https://www.wsj.com/articles/ai-cant-reason-why-1526657442
[20] Lawless, W.F., Mittu, R., Sofge, D.A. & Hiatt, L. (2019), Editorial (Introduction to the Special Issue), "Artificial intelligence (AI), autonomy and human-machine teams: Interdependence, context and explainable AI," *AI Magazine*, 40(3): 5-13.
[21] Cooke, N.J. & Hilton, M.L. (Eds.) (2015), Enhancing the Effectiveness of Team Science. Authors: Committee on the Science of Team Science; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Research Council. Washington (DC): National Academies Press.
[22] Cooke, N. (2020, 3/24), Effective human-artificial intelligence teaming, AAAI-2020 Spring Symposium, Stanford, CA. see the Blurb with links to the agenda at https://aaai.org/Symposia/Spring/sss20symposia.php#ss03
[23] Cummings, J. (2015). Team Science Successes and Challenges. National Science Foundation Sponsored Workshop on Fundamentals of Team Science and the Science of Team Science (June 2), Bethesda MD (https://www.ohsu.edu/xd/education/schools/school-of-medicine/departments/clinical-departments/radiation-medicine/upload/12-_cummings_talk.pdf).
[24] Martyushev, L.M. (2013), Entropy and entropy production: Old misconceptions and new breakthroughs, *Entropy*, 15: 1152-70.
[25] Lawless, W.F., Mittu, R. & Sofge, D.A. Editors (2020), Human-machine shared contexts. Elsevier.
[26] Wissner-Gross, A. D., and C. E. Freer (2013), Causal Entropic Forces, Physical Review Letters: 110(168702): 1-5.
[27] NTSB (2018, 5/24), "PRELIMINARY REPORT HIGHWAY HWY18MH010. The information in this report is preliminary and will be supplemented or corrected during the course of the investigation", National Transportation Safety Board, from https://www.ntsb.gov/investigations/AccidentReports/Pages/HWY18MH010-prelim.aspx
[28] NTSB (2019b, 11/19), 'Inadequate Safety Culture' Contributed to Uber Automated Test Vehicle Crash - NTSB Calls for Federal Review Process for Automated Vehicle Testing on Public Roads, National Transportation Safety Board, retrieved 4/11/2020 from https://www.ntsb.gov/news/press-releases/Pages/NR20191119c.aspx
[29] From Johns Hopkins Medicine: Surgical Team; retrieved 7/22/2020 from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/the-surgical-team
[30] Williamson, O.E. (2009), Transaction Cost Economics: The Natural Progression. Nobel Lecture. Retrieved 6/7/2020 from https://www.nobelprize.org/prizes/economic-sciences/2009/williamson/lecture/