AI and the Mitigation of Error: A Thermodynamics of Teams

W.F. Lawless¹ and Donald A. Sofge²

¹Paine College, 1235 15th Street, Augusta, GA, wlawless@paine.edu
²Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC; donald.sofge@nrl.navy.mil

Abstract

Traditional theories of social models conceptualize teams as distributed processors, disregarding the interdependence necessary to multi-task. Yet, interdependence characterizes social behavior. Instead, traditional theory favor cooperation, a state of least entropy production (LEP), without understanding the causes, limits or consequences of cooperation. As a simple example of interdependence, foraging prey overgraze forests free of predators. In our model, interdependence creates uncertainty, tradeoffs and signals (e.g., prices, coordination, innovation). Unlike individuals, the ability of teams to multitask reflects a quantum-like entanglement that represents maximum entropy production (MEP) when solving the problems signaled by society to improve its welfare. Our model supports findings that evolution in nature is driven by the MEP from making intelligent choices. Exploiting interdependence improves team intelligence, improves performance and reduces the risk of human error; forced cooperation disorganizes it by increasing the risk of error; e.g., if team cooperation improves teamwork, widespread forced cooperation in an autocracy or bureaucracy reduces social intelligence by adding unnecessary noise to signals. In our model, competition between teams self-organizes outsiders willing to sort through the noise for signals of the choices that improve social welfare (e.g., teams in courtrooms; science; entertainment; sports; businesses). Social systems organized around competition (e.g., stronger signals from robust checks and balances) better control a society by more correctly sizing teams to solve problems with fewer errors compared to autocracies or bureaucracies. Overall, we predict, the density of MEP directed at solving problems in a society with the constraints imposed from strong checks and balances, yet able to freely self-organize its labor and capital within those constraints, is denser.

Introduction

As a work-in-progress, we report that approaches using Artificial Intelligence (AI) may soon manage complex systems with teams, including hybrid teams composed arbitrarily of humans, machines, and robots. Already, AI has been useful in modeling the defense of individuals, teams, and institutions, as well as the management of social systems such as virtual ship bridges. However, foundational problems remain in the continuing development of AI for team autonomy, especially with objective measures able to optimize team function, performance and composition. But we want to know whether once this AI problem has been solved, will we scientists be able to invert the AI solutions in order to mitigate human error.

AI approaches often attempt to address autonomy by modeling aspects of human decision-making or behavior. Behavioral theory is either based on modeling the individual, such as through cognitive architectures or, more rarely, through group dynamics and interdependence theory. Approaches focusing on the individual assume that individuals are more stable than the social interactions in which they engage. Interdependence theory assumes the opposite, that a state of mutual dependence among participants in an interaction affects the individual and group beliefs and behaviors of participants. The latter is conceptually more complex, but both approaches must satisfy the demand for predictable outcomes as autonomous teams grow in importance and number.

From an intuitive perspective, interdependence is confusing. In the abstract, we gave this simple example of interdependence, foraging prey overgraze forests free of predators; as another example, behavior changes when humans believe they are being observed.

Despite its theoretical complexity, including the inherent uncertainty and nonlinearity exposed by interdependence, we argue that complex autonomous systems must consider multi-agent interactions to develop predictable, effective and efficient hybrid teams. Important examples include cases of supervised autonomy, where a human oversees several interdependent autonomous systems; where an autonomous agent is working with a team of humans, such as in a network cyber defense; or where the agent is intended to replace effective, but traditionally worker-intensive team tasks, such as warehousing and shipping. Autonomous agents that seek to fill these roles, but do not consider the interplay between the participating entities, will likely disappoint.
Overview of the problem of human error and AI’s role in its possible mitigation. AI has the potential to mitigate human error by reducing car accidents; airplane accidents; and other mistakes made either mindfully or inadvertently by individuals or teams of humans. One worry about this bright future is that jobs may be lost; from Mims (2015),

Something potentially momentous is happening inside startups, and it’s a practice that many of their established competitors may be forced to copy if they wish to survive. Firms are keeping head counts low, and even eliminating management positions, by replacing them with something you wouldn’t … think of as a drop-in substitute for leaders and decision-makers: data.

Commercial airline pilots disagree about the need to be replaced by AI (e.g., Smith, 2015).

... since the crash of Germanwings Flight 9525, there has been no lack of commentary about the failures and future of piloted aviation … Planes are already so automated, consensus has it, that it’s only a matter of time before pilots are engineered out of the picture completely. And isn’t the best way to prevent another Andreas Lubitz to replace him with a computer, or a remote pilot, somewhere on the ground? … The best analogy, I think, is modern medicine. Cockpit automation assists pilots in the same way that robots and other advanced medical equipment assist surgeons. It has improved their capabilities, but a plane no more flies itself than an operating room performs a hip replacement by itself. The proliferation of drone aircraft also makes it easy to imagine a world of remotely controlled passenger planes. … [but] remember that drones have wholly different missions from those of commercial aircraft, with a lot less at stake if one crashes.

An even greater, existential threat posed by AI is to the existence of humanity, raised separately by the eminent physicist Stephen Hawkings, the entrepreneur Elon Musk and the computer billionaire Bill Gates. Garland (2015), the director of the new film “Ex Machina”, counters their recent warnings:

... reason might be precisely the area where artificial intelligence excels. I can imagine a world where machine intelligence runs hospitals and health services, allocating resources more quickly and competently than any human counterpart. ... the investigation into strong artificial intelligence might also lead to understanding human consciousness, the most interesting aspect of what we are. This in turn could lead to machines that have our capacity for reason and sentience ... [and] a different future. ... one day, A.I. will be what survives of us.

In our research, we want to take an existential but rigorous view of AI with its possible applications to mitigate human error, to find anomalies in human operations, and to discover, when, for example, teams have gone awry, whether AI should intercede in the affairs of humans.

In our applications, we want to explore both the human’s role in the cause of accidents and the possible role of AI in mitigating human error; in reducing problems with teams, like suicide (e.g., the German co-pilot, Libutz, who killed 150 aboard his Germanwings commercial aircraft; from Kulish & Eddymarch, 2015); and in reducing the mistakes by commanders (e.g., the sinking of the Japanese tour-boat by the USS Greeneville; from Nathman et al., 2001).

Human error

Across a wide range of occupations and industries, human error is the primary cause of accidents. Take, for example, general aviation:

The National Safety Board found that in 2011, 94% of fatal accidents occurred in general aviation (Fowler, 1994);

In general aviation, the Federal Aviation Administration (FAA, 2014) attributed the accidents that occur primarily to stalls and controlled flight into terrain, that is, to avoidable human error.

Exacerbating the sources of human error, safety is the one area an organization often skimps as it tries to save money. From Gilbert (2015),

“History teaches us that when cuts are made in any industry, the first things to go are safety and training—always, every industry,” says Michael Bromwich, who oversaw the regulatory overhaul after the [Deepwater Horizon] disaster before stepping down as head of the Bureau of Safety and Environmental Enforcement in 2011 in 2011.

The diminution by an organization in its valuation of safety coupled with human error led to the explosion in 2010 that doomed the Deepwater Horizon in the Gulf of Mexico (Gilbert, 2015).

The mistakes that led to the disaster began months before the Deepwater Horizon rig exploded, investigators found, but poor decisions by BP and its contractors sealed the rig’s fate—and their own. On the evening of April 20, 2010, crew members mis-interpreted a crucial test and realized too late that a
dangerous gas bubble was rising through the well. The gas blasted out of the well and ignited on the rig floor, setting it ablaze.

Human error also emerges as a problem in the management of civilian air traffic control (ATM). ATM’s top five safety risks nearly always involve ‘Human Error’ (ICAO, 2014).

Human error was the cause attributed to the recent sinking of a research ship (Normile, 2015):

The sinking of Taiwan’s Ocean Researcher V last fall resulted from human error, the head of the country’s Maritime and Port Bureau told local press this week. The 10 October accident claimed the lives of two researchers and rendered the dedicated marine research ship a total loss. ... Wen-chung Chi, director-general of the Maritime and Port Bureau, said that a review of the ship’s voyage data recorder and other evidence indicated that the crew should have been alerted that the ship had drifted off course.

The role of Al in reducing human error

The causes of human error could be attributed to endogenous or exogenous factors (we discuss the latter under anomalies and cyber threats). A primary endogenous factor in human causes of accidents is either a lack of situational awareness, or a convergence into an incomplete state of awareness determined by culture greatly amplified by emotions decision making (e.g., the Iranian Airbus Flight 655 erroneously downed by the USS Vincennes in 1988; e.g., Lawless et al., 2013). Many other factors have been subsumed under human error including poor problem diagnoses; poor planning, communication and execution; and poor organizational functioning.

What role can AI play in reducing human error? First, Johnson (1973) proposed that the most effective way to reduce human error is to make safety integral to management and operations. Johnson’s “MORT” accident tree analyses attempts to identify the operational control issues that may have caused an accident, and the organizational barriers that resist uncovering the deficiencies that contributed to an accident. MORT has been prized by the Department of Energy as the ultimate tool for identifying possible hazards to the safe operation of nuclear reactors.

Second, checks and balances on cognitive convergence processes permit the alternative interpretations of situational awareness that may prevent human error. Madison (1906) wrote that it is to free speech and a free press, despite all their abuses, that “the world is indebted for all the triumphs which have been gained by reason and humanity over error and oppression.” But in closed organizations, like the military in the cockpit or on a ship’s command bridge, the lack of free speech poses a threat that is associated with an increase in errors. Based on Smallman’s (2012) plan to reduce accidents in the U.S. Navy’s submarine fleet with technology that illustrates the range of opinions existing among a ship’s staff for a proposed action, in real time, AI can offer alternative perspectives that oppose the very convergence processes that permit errors to thrive (e.g., “groupthink”; Janis, 1982).

The role of Al in the discovery and rectification of anomalies. An anomaly is a deviation from normal operations (Marble et al., 2015). The factors we wish to consider are those associated with cyberattacks; e.g., the subtle takeover of a drone by a foreign agent. But, we expect, AI will be able to be used proactively for “novel and intuitive techniques that isolate and predict anomalous situations or state trajectories within complex autonomous systems in terms of mission context to allow efficient management of aberrant behavior” (Taylor et al., 2015).

The role of Al in determining team metrics and shaping team thermodynamics: In an earlier study of teams (Lawless et al., 2013), we had concluded with a bold, well-supported assertion ... that we now conclude was wrong (e.g., Lawless et al., 2015). We had come to believe at the time that for AI, the psychology of the individual is more or less a blind alley followed by many social scientists ever since James (1890). While we are not the only scientists to question the value of psychology (e.g., Witkowski & Zatoranski, 2015), we recant, somewhat.

Briefly, while teams enhance the performance of the individual (Cooke & Hilton, 2015), the added time to coordinate and communicate among a team’s members is a cost and also a source of error. We propose that well-performing teams can decrease the likelihood of human error, poorly performing teams can increase human error. We consider static and bistable agents.

With Shannon’s static, stable agents, the information theory community appears to be more focused on the mutual information shared at the input and output of two variables, forming a channel. Instead, for individual agents forming a team, we devised a measure of team efficiency based on the information transmitted by a team. We concluded that (Moskowitz et al., 2015, p. 2), in general, a “high degree of team player interdependence is desirable (mapping to good performance). However this is not always the case, certain tasks may require players to be independent.” We also concluded that the “larger the spectral radius of the topology of a team, the more efficient it will be [and the easier for knowledge to spread], all other things being equal.” (pp. 7-8)

Bistable agents produce greater complexity. To form teammates, we reconsidered agents as bistable to capture the independent observation and action behaviors of humans, or the multiple interpretations of reality common to humans (Lawless et al., 2015). Bistability transforms interdependence into a phenomenon similar to quantum entanglement.
Building on our concept of efficiency, but with bistable agents to replicate humans as observers and actors, we noted that Shannon’s team entropy is at a minimum for a team slaved together. Contradicting Shannon except when observation and action are perfectly aligned, or when multiple interpretations converge in perfect agreement, in our model, like the independent brain systems that observe and act interdependently as a team, we set the baseline entropy for a perfect team at the least entropy production (LEP); below, we will find that LEP is associated with maximum cooperation.

With the bistable agent model (Lawless et al., 2016), by increasing the energy expended in teamwork, we concluded that the robust intelligence necessary to succeed appears unlikely to occur for agents acting independently, reducing autonomy and thermodynamic effectiveness (e.g., productivity). By multitasking (MT) together, a team of agents is more effective than the same agents independently performing the same tasks. For a team of bistable agents, however, bias reduces its robust intelligence and autonomy. Instead, robustness requires a MT team observing reality to contradict an opposing team (two teams of MT’s best capture situational awareness), implicating the value of competition in determining reality, but these two teams are also insufficient for robustness. Robust intelligence requires three teams: two opposing MT teams to construct Reality plus a neutral team of freely moving bistable independent agents attracted or repelled to one or another team to optimize the thermodynamic forces that determine team effectiveness and efficiency. Thus, given two competitive teams, adding a third team to constitute a spectrum of neutral bistable agents able to invest freely (properties, ideas, works), act freely (joining and rejecting either team), and observe freely makes the greatest contribution to robust intelligence, to mitigating mistakes, and to maximizing effective and efficient autonomy.

**Entropy of Teamwork: Multitasking**

Beset by the lack of replicating important social science research (e.g., in psychology Nosek, 2015; in economics Chang & Li, 2015) reduces the generalizations that social science has to improve society that are not normative. Compounding the failure to replicate, discounting the value of interdependence, the defining characteristic of social behavior, most of social science is focused on the individual, including economics (Adhieh, 2009) and the law (Zamir & Teichman). In fact, social psychologists recommend that interdependence be removed or avoided (e.g., social psychology, in Kenny et al., 1998; information theory, in Conant, 1976). While game theory incorporates interdependent choices (Von Neumann & Morgenstern, 1953), it negates its value with toy problems (Rand & Nowak, 2013) designed to overstate the value of social cooperation, an atheoretical result that models least entropy production (LEP). And yet, the most basic tool of traditional social science, the self-reported observations by individuals, has left social scientists unable to comprehend how to resolve their inability to find sizable correlations between self-reports and behavior (Zell & Krizan, 2014). Despite this uncertainty with the replication and meaning of its basic data, social scientists strongly believe that cooperation is superior to competition, but some of them at least admit “the evolution of the human mind is a profound mystery” (e.g., Pinker, 2010, p. 8999).

Traditionally, teams have been organized around a division of labor, like distributed processing where a job is divided into n processors (Traub, 1985), overlooking the large benefits derived from exploiting interdependence with multitasking (Bartel et al., 2013). Individuals are poor at multitasking (Otto & Sentana, 2015), the function of teams; e.g., like the independent roles for the players cooperating together to form a baseball team. Heretofore, interdependence has long been an unsolved theoretical complexity; e.g., Von Neumann and Morgenstern (1953, p. 148) feared that if Bohr was correct about the interdependence between action and observation, it would make a rational model of the interaction “inconceivable”. More importantly, by basing our model on interdependence, we calculate that the reduction in the degrees of freedom (dof) for the individuals who constitute a team is similar to quantum entanglement.

In our model, the perfect team is composed of independent agents working interdependently; by exploiting interdependence, the reduction in dof helps the team to generate LEP from internal cooperation among teammates to maintain structure, the small expenditures of energy on structure allowing a team to maximize entropy production (MEP) on the problems the teams was formed to address, thereby driving social evolution. In fact, “evolution (progress) in Nature demonstrates … maximization of the entropy production” (Martyushev, 2013, p. 1152). This is further brought home with the recent work on Entropic Intelligence (Wissner-Gross & Freer, 2013) to stress that the most intelligent choices follow curves that maximize entropy (where F is the entropic force; T the reservoir temperature; $S(X)$ the entropy associated with macrostate $X$; and $X_0$ the present macrostate):

$$ F(X_0) = T\nabla_X S(X)|_{X_0} $$

(1)

The key to understanding equation (1) is that the search for the paths to the macrostate $X$ that are most likely to maximize MEP must avoid “excluded” volumes (Wissner-Gross & Freer, 2013, p. 168702-1). Non-equilibrium systems, such as the intelligence displayed by human and machine teams (i.e., hybrid teams) and higher organizations of teams, measured by productivity, have a bias for instantaneous MEP (Martyushev & Selezn, 2006) at each instant in time across all possible paths.
through configuration spaces that seek to avoid “excluded” volumes (Wissner-Gross & Freer, 2013). That this information is dynamically revealed (Ziebart et al., 2010) suggests a role for intelligent observers in instantly seeking the path to the microstate at each instant that is most likely to promote MEP (Wissner-Gross & Freer, 2013). Should this state be recognizable and determinable (Lawless et al., 2015) leaves open two questions: “How?” and “How to avoid excluded volumes?”

We have also applied our model to understand the disorganization from social systems that widely enforce cooperation in a fashion that reduces intergroup competition, in effect increasing the size of “excluded” volumes: gangs; authoritarians (e.g., China); and large bureaucracies (e.g., the U.S. Department of Energy, or DOE). Increasing the size of excluded volumes occurs when teams are fragmented by enforced consensus decision rules (Lawless et al., 2008). For example, instead of using MEP to address its problems, present-day China is creating uncertainty that impedes a random search for MEP as it, instead, expends energy to tear apart its self-organized social structures, thereby reducing its social intelligence and social welfare [9]. Autocracies work by decreasing intergroup competition; e.g., in politics; in the law; and in environmental practices. While competition leads to centers of self-organization able to generate MEP (Lawless et al., 2016), open competition also generates the instability and uncertainty that have long offended socialists (Herzen, 1982). Why is competition necessary?

To answer this question with our model, we envision that the frisson from competition attracts the onlookers (e.g., neutral voters in a political campaign; jury members in the courtroom; computer users seeking new hardware) who are willing to sort through the noise to find or disambiguate the signal for the options, paths and choices that better provide the resources needed by a society to improve its social welfare (Lawless et al., 2016). The surprise with our approach is that social systems organized around competition (checks and balances) are better able to control the many elements of a society than authoritarian regimes (e.g., on the inability of China to control the lower-level managers of its economy, see Talley & Peker, 2015).

Results

By being inefficient in optimizing the size of teams able to solve a problem they deem important, authoritarian regimes are unable to reach the levels of MEP comparable to free-market economies e.g., authoritarian decision-makers prefer consensus decisions, otherwise known as minority control (Lawless et al., 2008).

As an example of consensus decision making from the EU (WP, 2001, p. 29):

The requirement for consensus in the European Council often holds policy-making hostage to national interests in areas which Council could and should decide by a qualified majority.

As another example, DOE-WIPP data for DOE’s Citizen Advisory Boards (CAB) following majority-rule (MR) versus consensus-rule (CR) decision-making indicated that teamwork and compromise between factions was more prevalent under MR but that teams were more fragmented under CR (Lawless et al., 2008). Specifically, we found that four of five MR-CABs accepted the recommendations of DOE’s scientists regarding the operation of WIPP’s transuranic waste repository compared to three of four CR-CABs that rejected the same advice. This result indicates that majority rule teams are more highly interdependent than consensus ruled teams; that they sort through the nose of social interaction; and that they make more intelligent decisions.

Similarly, recent research indicates that the best teams of scientists are highly interdependent (Cummings, 2015).

Hypothesis: The open systems that create and separate information from noise (higher T) better determine macrostate paths than the closed systems which obfuscate information (lower T) by increasing the size of its “excluded” volumes (Lawless et al., 2016). The driving force for social evolution, more likely in democracies than autocracies, competition between groups (Lawless et al., 2015) forces teams to survive by becoming effective and then efficient. Democracies with healthy checks and balances not only generate noise but the solutions to problems; under majority rule (MR), opposing viewpoints are more likely to lead to a solution, suggesting that opposed teams are fighting through the combination of noise and information to find the signal for the information that indicates an better path forward. In contrast, the fragmentation from consensus rules (CR) increases the degrees of freedom of a team, leading to less success (Lawless et al., 2008); e.g., that autocracies censor public discourse leads to the misallocation of team resources.

Authoritarian regimes are unable to reach the levels of MEP comparable to free-market economies by being inefficient in optimizing the size of teams able to solve a problem they deem important; e.g., Sinopec oil company uses about 548 thousand employees to produce about 4.4 million barrels of oil per day whereas Exxon uses about 82 thousand employees to produce about 5.3 million barrels of oil per day (Lawless et al., 2016).

Authoritarians, surprisingly, are also ineffective at gaining widespread social cooperation with the law by not adhering to the value of the checks and balances afforded with the law; thus, autocracies are less effective at gaining the same level of cooperation than democracies which operate with more effective checks and balances.
(Lawless et al., 2016). Less effective with matters of the law increases team fragmentation and internal competition.

As an example, Indonesia is the 105th freest country in 2015 (e.g., www.heritage.org). The following example exemplifies the inability of Indonesia in confronting and resolving its environmental affairs (Otto & Sentana, 2015):

Indonesia is preparing [its] Navy ships to evacuate citizens suffering from a toxic haze that has spread throughout the region and grounded flights as far away as the Philippines and Thailand.

As another example, the military productivity of Iran, an autocratic theocracy, is less effective than the military productivity of Israel, a constitutional democracy (see www.globalfirepower.com).

Discussion and Conclusion

The central planning by autocracies leads to resource misallocation, increasing entropy but not entropy focused as MEP, thereby increasing waste far more than societies governed by free-markets. Central planning also reduces the competition that drives innovation (Ridley, 2015). For example, Europe is turning against biotech science (Lynas, 2015).

From our work-in-progress, we conclude that Authoritarian regimes are unable to reach the levels of MEP comparable to free-market economies by being inefficient in optimizing the size of teams able to solve a problem that has been deemed important.

Unlike rational decision-making (lower T), the key to causal entropy forces is to search for and to find a point (higher T) of instability (Wissner-Gross & Freer, 2013) common to decision making by those human groups free to explore the configuration space for solutions, and where choices have the potential to head in maximum manifold directions at any instant in order to solve the problem faced by a team. This instability is more likely to occur with decision making in democracies than in autocracies.

Applying equation (1) to hybrid teams of humans, machines and robots, our model exploits interdependence to improve teams, their decisions, and, by generalizing, social intelligence (Cummings, 2015). Finally, we expect to find in future research that the density of MEP directed at solving problems in a society able to freely self-organize its labor and capital is denser.

In this paper, we studied how AI might serve as a metric of team performance to guide interventions that reduce a team’s errors, whether for human, machine or robot teams. We generalized team fitness with a comparative measure of MEP, where low MEP increases errors. In the future, we contrast whether teammate redundancy (network theory) increases team efficiency or reduces fitness (interdependence theory). From network theory, Centola and Macy (2007, p. 716) conclude: “As redundancy increases, the network becomes more efficient if some of the redundant ties are randomly rewired to create new bridges.” However, based on Exxon’s and Sinopec’s similar production rates compared to the vastly increased number of teammates for Sinopec, we can conclude the opposite, that an over-fitted team creates opportunities for corruption, not efficiency.

References


ICAO (2014, 8/8), “Integration of Human Factors in research, operations and acquisition”, presented at the International Civil Aviation Organization’s Second Meeting of the APANPIRG

www.globalfirepower.com/countries-listing-middle-east.asp

www.heritage.org/index/country/indonesia

