Check for updates

Toward a Physics of Interdependence for Autonomous Human-Machine Systems: The Case of the Uber Fatal Accident, 2018

William Lawless *

Department of Mathematics and Psychology, Paine College, Augusta, GA, United States

Computational autonomy has begun to receive significant attention, but neither the theory nor the physics is sufficiently able to design and operate an autonomous human-machine team or system (HMS). In this physics-in-progress, we review the shift from laboratory studies, which have been unable to advance the science of autonomy, to a theory of autonomy in open and uncertain environments based on autonomous human systems along with supporting evidence in the field. We attribute the need for this shift to the social sciences being primarily focused on a science of individual agents, whether for humans or machines, a focus that has been unable to generalize to new situations, new applications, and new theory. Specifically, the failure of traditional systems predicated on the individual to observe, replicate, or model what it means to even be the social is at the very heart of the impediment to be conquered and overcome as a prelude to the mathematical physics we explore. As part of this review, we present case studies but with a focus on how an autonomous human system investigated the first self-driving car fatality; how a human-machine team failed to prevent that fatality; and how an autonomous human-machine system might approach the same problem in the future. To advance the science, we reject the aggregation of independence among teammates as a viable scientific approach for teams, and instead explore what we know about a physics of interdependence for an HMS. We discuss our review, the theory of interdependence, and we close with generalizations and future plans.

Keywords: interdependence, autonomy, teams, systems, human-machine

1 INTRODUCTION. DIFFERENT SCHOOLS OF THOUGHT AND CONTROVERSIES

Aim: The U.S. National Academy of Sciences [1,2] has determined that interdependence among teammates is the critical ingredient for a science of autonomous human-machine teams and systems (HMS), but that teams cannot be disaggregated to determine why or how they work together or to replicate them, stymying the development of a mathematics or physics of autonomous human-machine teams. Yet, surprisingly, the social sciences, which began with the Sophists over two millennia ago, still aggregate individuals to study "the nature and properties of the social world" [3]. Our aim is to overcome this barrier that has precluded the study of the "social world" to construct a physics of autonomy.

The social disruption posed by human-machine systems is more likely to be evolutionary, but it poses a dramatic change that Systems Engineers, social scientists and AI researchers must be

prepared to manage; however, the theory to bring about this disruption we suspect will prove to be revolutionary. Designing synergistic interactions of humans and machines that holistically give rise to intelligent and autonomous systems requires significant shifts in thinking, modeling, and practice, beginning with changing the unit of analysis from independent humans or programmable machines to interdependent teams and systems that cannot be disaggregated. The case study of a fatality caused by a self-driving Uber car highlighted in our review barely scratches the surface to prepare readers for this disruption. The case study is a simple model of a machine and its human operator that formed a two-agent team involved in a fatal accident; but the human teams subsequently involved in the analysis of this fatality serve as a tool to measure how far we must go intellectually to accommodate an HMS.

We hope to provide a sufficient overview of the literature for interested readers to begin advanced studies to expand their introduction to the physics of autonomy as it currently exists. We conclude with a discussion of generalizations associated with our theory of autonomy, and, in particular, the entropy production associated with state-dependent [4] changes in an autonomous HMS [5]. In our review, we contrast closed model approaches to solving autonomy problems with open model approaches. A closed model is self-contained; the only uncertainty it is able to study is in the complexity created within its own model. Open models contain natural levels of uncertainty, competition or conflict, and sometimes all three.

At its most basic level, in contrast to closed systems, the case studies explore the fundamental tool of debate used for millennia by autonomous human teams confronting uncertainty. They led us to conclude that machines using artificial intelligence (AI) to operate as members of an HMS must be able to tell their human partners whenever the machines perceive a change in the context or emotion that affects their team's performance (context change may not be detectable by machine learning alone, which is context dependent; see [6]); in turn, AI machines must be able to understand the humans interacting with them in order to assess their contributions to a team's performance from their perspective as team members (i.e., how can an HMS improve a team's choices; how can an HMS improve the effectiveness of a team's performance; etc.; in [7]). The human and AI members of the team must be able to develop goals, learn, train, work and share experiences together. As part of a team, however, once these AI governed machines learn what humans want them to learn, they will know when the human members of their team are either complacent or malicious in the human's performance of the human's roles [8], a capability thought to be possible over the next few years [9]; in that case, or if a machine detects an elevated level of emotions, the machine must be able to express its reservations about a team's decisions or processes to prevent a mistake. There is even more to be mined in the future from the case studies we review. Specifically, if a human or machine is a poor team member, what exists in the AI Engineering[1] toolbox, the

social science armature, physics, or elsewhere to aid a team in the selection of a new member of a system? Furthermore, how is the structure of an HMS with a poorly performing team member, human or machine, related to the performance of the team?

In what follows, the primary problem is to better manage the state of interdependence between humans and their machine teammates. The National Academy of Sciences review of human teams in 2015 [1] renewed interest in the study of interdependence, but the Academy was not clear about its implications, except that it existed in the best performing human teams [10], that led them to conclude that a team of interdependent teammates will likely be more productive than the same collection of individuals who perform as independent individuals [1,7]. The new National Academy of Sciences report on "Human-AI Teaming," commissioned by the U.S. Air Force, also discusses the value of interdependence, but without physics [2].

2 FUNDAMENTAL CONCEPTS, ISSUES AND PROBLEMS

In this section, we briefly review the definitions of terms used in this review (autonomy; rational; systems; closed and open systems; machine learning; social science).

2.1 Definitions: Autonomy. Autonomous Human-Machine Teams

Autonomy. Autonomous systems have intelligence tools to respond to situations that were not programmed or anticipated during design; e.g., decisions; self-directed behavior; human proxies ([2], p. 7). Autonomous human-machine systems work together to fulfill their design roles as teammates without outside human interaction in open systems, which include uncertainty, competition or conflict. Partially autonomous systems, however, require human oversight.

Autonomous human-machine teams occur in states of interdependence between humans and machines, both types able to make decisions together ([2], p. 7). Like autonomous human teams, they likely will be guided by rules (e.g., rules of engagement; rules for business; norms; laws; etc.). Theories of autonomy include human-machine symbiosis, arising only under interdependence, and addressed below.

2.2 Definitions: The Rational

Systems engineering. A concept is rational when it can be studied with reason or in a logical manner. A rational approach for traditional system engineering problems is considered to be the hallmark of engineering, such as a self-driving car. Paraphrased, from IBM's Lifecycle Management,[2] the rational approach consists of several steps: determining the requirements to solve a problem; design and modeling; managing the project; quality

---

control and validation; and then to integrate across disciplines to assure the success of a solution:

Requirements engineering: Solicit, engineer, document, and trace the requirements to determine the needs of the stakeholders involved. Build the work teams (e.g., considering the solutions available from engineering, software, technology, policy, etc.) that are able to adapt as change occurs in the needs and designs to be able to deliver the final product.

Architecture design and modeling: Model visually to validate requirements, design architectures, and build the product.

Project management: Integrate planning and execution, automate workflows, and manage change across engineering disciplines and development teams, including: iteration and release planning; change management; defect tracking; source control; automation builds; reporting; and customizing the process.

Quality management and testing: Collaborate for quality control, automated testing, and defect management.

Connect engineering disciplines: Visualize, analyze, and organize the system product engineering data with the tools that are available (viz., design and operational metrics and performance goals).

From the *Handbook of Systems Engineering* [11], engineering system products must be a transdisciplinary process; must include product life cycles (e.g., manufacturing; deployment; use; disposal); must be validated; must consider the environment it operates within; and must consider the interrelationships between the elements of the system and users.

2.3 Definitions: Artificial Intelligence and Machine Learning

At its simplest, AI is a rational approach to make systems more intelligent or "smart" by incorporating "rules" that perform specific tasks inside of a closed system (e.g., hailing a ride from Point A to Point B on a software platform from an Uber driver; [3] connecting the nearest available Uber driver and estimating costs and fees; gaining a mutual agreement between the customer and Uber driver). However, AI must also address human-machine teams operating in open systems ([2], p. 25). Machine learning (ML) is a subset of AI used to train an algorithm that learns from a "correct," tagged or curated data set to operate, say, a self-driving car being driven to learn while in a laboratory, on a safe track, or over a well-trodden, closed path in the real world (e.g., in the case of the Uber self-driving car, it was in its second loop along its closed path at the time of the fatality; in [12]; for more on Uber's self-driving technology, [4] see https://

www.uber.com/us/en/atg/technology/and   https://www.uber.com/us/en/atg/). In contrast to ML, deep learning (DL) is a subset of ML that may construct neural networks for classification with multiple bespoke layers and may entail embedded algorithms for training each layer; also, DL assumes a closed system.

2.4 Definitions: Social Science, Especially Its Application to Autonomous Teams

Social science studies "the nature and properties of the social world" [3]. It primarily observes individuals in social settings by aggregating data from individuals, and by statistical convergence processes on the data collected. We address its strengths and weaknesses regarding autonomous human-machine teams.

Social science: Strengths. Social science has several strengths that can be applied to autonomous human-machine teams. For example, the study of the cockpit behavior of commercial airline pilots separates the structure of teams from their performance [13], which we adopt. Functional autonomous human-machine teams cannot be disaggregated to see why they work ([2], p. 11), an important finding that supports the physics model we later propose. Cummings [10] found that the worst performing science teams were found to be interdisciplinary, suggesting poor structural fits, agreeing with Endsley [2], findings similar to Lewin's [14] claim that the whole is greater than the sum of its parts, which we adopt and explain. And it is similar to the emergence of synergy in systems engineering [11], which we also adopt, including symbiosis (mutual benefit).

Social science: Weaknesses. Traditional social science has little guidance to offer to the new science of autonomous systems [15]. The two primary impediments with applying traditional social science to an HMS are, first, its use of closed systems (e.g., a laboratory) to study solutions to the problems faced instead of the open field where the solutions must operate ([2], p. 56); and second, the reliance by social scientists on the implicit beliefs of individuals as the cause of observed behaviors, or the implicit behaviors of individuals derived from aggregated beliefs.

Implicit beliefs or behaviors are rational. Either works well for limited solutions to closed problems (e.g., game theory). The difficulty with implicit beliefs or behaviors is their inability to generalize. First, with the goal of behavioral control [16], by adopting implicit beliefs, physical network scientists, game theorists [17], Inverse Reinforcement Learning (IRL) scientists (e.g. [18]) and social scientists (e.g. [19]) have dramatically improved the accuracy and reliability of applications that predict human behavior but only in situations where alternative beliefs are suppressed, in low risk environments, or in highly certain environments; e.g., the implicit preferences based on the actual choices made in games [17] do not agree with the preferred choices explicitly stated beforehand ([20], p. 33). In these behavioral models, beliefs have no intrinsic value.

In contrast, second, often based on surveys or mental tasks, cognitive models discount the value of behavior [21], improving the correlations between cognitive concepts and cognitive beliefs about behavior, but not actual behavior; e.g., self-esteem beliefs correlate strongly with beliefs about academics or work (e.g. [22]),

---

[3]https://www.feedough.com/uber-business-model/
[4]On December 7th, Uber sold its self-driving unit to Aurora Innovation Inc. [113].

TABLE 1 | The failure of concepts to generalize to build new theory: The case of social science in closed systems. Column two contrasts the leading concepts in social science by its founding social scientist(s); Column three shows the scientist who toppled the leading concept. The table highlights the inability of social scientists to build new theory from prior findings.

Leading theory	Leading theory and theorist	Theory invalidated by:
Self-Esteem	Diener [22]; hailed by the American Psychological Association	[23]
Ego-Depletion	[25]	[26]
Implicit Attitudes Theory (racism)	[27]	[28]
Superforecasters	[30]	[32]
[31], PNAS	Shu et al. [31]; includes Ariely and Bazerman, the chief developer and promoter of the "Honesty" scale	Berenbaum [29], Editor in Chief, PNAS, retracted

but not with actual academics or work (e.g. [23]). Implicit behavior models, however, lead to predictions that either fail on their face or from a lack of replication. These failures are reviewed in **Table 1**. They have led to Nosek's [24] replication project in an attempt to avoid searing headlines of retractions; however, Nosek's replication project does not overcome the problem with generalizations.

Table 1 illustrates that "findings" in traditional social science with data from individuals in closed systems cannot be generalized to new findings. From the first row of **Table 1**, proposed by Diener [22], high self-esteem has been hailed by the American Psychological Association (APA) as the best psychological state that an individual can achieve, but the concept was found to be invalid by Baumeister et al. [23]. Later, Baumeister developed ego-depletion theory [25], a leading concept in social psychology until it was found to be invalid by [26]. Implicit attitudes theory, the concept undergirding implicit racism, was proposed by Greenwald et al. [27], but later found to be invalid by Tetlock's team [28]. The next failure to generalize, is the leading theory in social psychology developed by Tetlock for predictions to be made by the public and businesses known as superforecasting; however, the first two forecasts made by his highly trained international superforecasters were that the United Kingdom's Brexit would not occur in 2016, and that Donald Trump would not be elected President; both happened. After years of giving TED talks on the value of honesty, [5] Ariely published his new "honesty" scale in the Proceedings of the National Academy of Sciences (PNAS), which recently was retracted by the Editor of PNAS [29]. Apparently, Ariely fabricated the data for his honesty scale.

Social science includes economics. As an example from economics of the same problem with models of closed systems of individual beliefs, Rudd [33] has written that "Mainstream economics is replete with ideas that "everyone knows" to be true, but that are actually arrant non-sense." Rudd focuses on inflation, concluding that expectations (based on surveys) of inflation are not related to the inflation that actually

occurs. Rudd's conclusion is part of an ongoing series of arguments about the causes of inflation. For example, from two Nobel Laureates, first has been "the failure of many economists to get inflation right" [34]; and second, inflation has also been attributed to the fear of a wage-price spiral driving expectations [35]. But other economists such as Larry Summers, a leading economist, have predicted that the extraordinary fiscal expenditures during the pandemic would likely cause inflation [36]. Summers was initially contradicted by a proponent of the new economics, known as Modern Monetary Theory, but MMT, which holds that inflation is unlikely from excessive government expenditures, is now on the defensive from the existence of rapidly rising government expenditures associated with inflation [37].

As another example from economics, Leonard [38] concludes that the Federal Reserve's Open Market Committee (FOMC) often misunderstands the US economy by failing to produce intended results, partly due to its adherence to consensus seeking (also known as group think or minority control, often under the auspices of a strong leader of the FOMC; in [5]). As a last example this time with economic game theory, the use of war games in the fleet results in "preordained proofs," per retired General Zinni (in [39]); that is, choose a game for a given context to obtain a desired outcome.

2.5 Definitions. Open Systems and Interdependence

A different model than closed systems of individuals is needed to be replaced by open systems of teams [2]. The approaches to design and operation in the future, however, must also include autonomy and the autonomous operations of human and machine teams and systems. [6] That likely means that these models for autonomy must address conflict and uncertainty, both of which impede or preclude the rational approaches that are only understood in closed systems [40], like game theory [17].

---

[5]See Ariely giving a TED talk on "How to change your behavior for the better," including honesty at https://www.ted.com/talks/dan_ariely_how_to_change_your_behavior_for_the_better.

[6]To meet the digital future, the Systems Engineering Research Center (SERC) is developing a roadmap for Systems Engineering; e.g., https://www.researchgate.net/publication/340649785_AI4SE_and_SE4AI_A_Research_Roadmap.

As an overly simplistic example: Should the well-trained arm of a human implicated in a fatality be taken to court (a human comparison with ML and DL), the actions of the arm would have to be explained. In this simple case, explanations would have to be provided in a court of law by the human responsible for the arm (e.g. [15]). Explainability for machines is hoped to be provided eventually in real time by AI, but it is neither available today nor at the time of Uber's fatal accident (e.g., [41, 42]). In the case of Uber's pedestrian fatality, the explanation was provided at great expense and after more than a year of intense scrutiny by the National Transportation Safety Board [43].

Another example further prepares us to move beyond the case study. By following simple rules, the work output from a team of uniform workers digging a ditch, or machines in a military swarm, can be aggregated; e.g., "many hands make light work" [1]. With Shannon's [44] rules of communication between two or more independent agents, it is straight forward to model. In contrast, a team or system constituted with orthogonal roles (e.g., a small restaurant with a waiter, cashier and cook), while more common, is much harder to model for the important reason that their perceptions of reality are different. Consider a bi-stable illusion that generates two orthogonal or incommensurable interpretations (e.g., the bi-stable two-faces candlestick illusion). A human perceiving one interpretation of the illusion cannot perceive its bi-stable counterpart simultaneously [45]. Thus, the information collected from two or more workers coordinating while in orthogonal roles can lead to zero correlations, precluding the convergence to a single story from occurring [46]; e.g., despite over a century of being the most successful theory with prediction after successful prediction, quantum theory does not abide by intuition and it resists a rational interpretation (e.g. [47]).

A distinction is necessary. Quantum mathematics is logical, rational and generalizable, however, the interpretations derived from its results are neither logical, rational nor generalizable, an important distinction.

More relevant to our case study, rational approaches, beliefs and behaviors, whether implicit or observed, fail in the presence of uncertainty [48] or conflict [40], exactly where interdependence theory thrives [5]. Facing uncertainty, interdependence theory predicts that free humans engage in debate to exploit the bistable views of reality that naturally exist to explore the tradeoffs that test or search for the best paths going forward, bringing to bear experience, goals, ability to negotiate, fluidity of the situation, all interdependently integrated to confront the uncertainty faced. Thus, in the development of human-machine systems, an environment for interdependence, shared experience, and team learning from training is necessary. This idea also extends to actual teaming; the human and machine must continually test and reevaluate their interdependence *via* jointly developed knowledge/skills/abilities. In particular, reducing the uncertainty faced by a team or system requires that human and machine teammates are both able to explain to each other, however imperfectly, their past actions, present status, and future plans in causal terms [41,42].

Literature. For the human-autonomous vehicle interaction, there are generally accepted concepts and theories in the literature for what technical prerequisites have to happen for humans and machines to become team players; e.g. mutual predictability, directability, shared situation awareness and calibrated trust in automation [49, 50]. In aviation, fly-by wire systems have been implemented that enable human-machine interaction, some of which are being tested in vehicles (e.g., conduct by wire, H-mode; [7] in [51]; and [52]).

2.6 Preliminary Implications for Theory

Lewin [14] founded the discipline of social psychology. His key contribution identified the importance of interdependence in what was then known as "group dynamics." Two of his followers developed a full theory of interdependence centered around Von Neumann's and Morgenstern's [53] idea of games [54,55], but the lead contributor, Thibaut, died and Lewin's student, Kelley [56], gave up on being able to account for why human preferences established before a game failed to predict the choices made during actual games. Jones [20] asserted that ". . . most of our lives are conducted in groups and most of our life-important decisions occur in contexts of social interdependence," but that the study of interdependence in the laboratory was "bewildering." Subsequently, assuming that only i.i.d. data was of value in the replication of experiments (where i.i.d. stands for "independent and identically distributed" data; in [57]), Kenny et al. [58] devised a method to remove the statistical effects of interdependence from experimental data, somewhat akin to treating quantum effects as "pesky" [46]. After the National Academy of Sciences [1] renewed interest in the study of interdependence in 2015, the Academy's review of human-machine teams concluded that the interdependence among team members precluded the attribution of a team's performance to the "disaggregation" of its contributing members ([2], p. 11), directly contradicting Von Neumann's theory of automata, but directly supporting our physics of interdependence. How, then, can it be studied is the goal of this article.

3 CURRENT RESEARCH GAPS. A CASE STUDY

Purpose. The purpose of this case study is to explore some of the implications of teamwork, such as metrics of structure or performance of teams, that can be applied in an AI engineered system by reviewing one of the first autonomous human-machine systems that failed, resulting in a fatal accident. We then attribute the cause of the accident to a lack of interdependence between the human operator and the machine.

3.1 NTSB Report

The following is summarized from the NTSB [59] report on automation.[8]

On 18 March 2018, a 49-year-old female pedestrian walking a bicycle was fatally struck by a 2017 Volvo XC90 Uber vehicle operating an Automated Driving System (ADS) then under development by Uber's Advanced Technologies Group (ATG).

At the time of the pedestrian fatality, the ATG-ADS had used 1 lidar and 8 radars to measure distance; several cameras for detecting vehicles, pedestrians, reading traffic lights and classifying detected objects; various sensors that had been recently calibrated for telemetry, positioning, monitoring of people and objects, communication, acceleration and angular rates. It also had a human-machine interface (HMI) tablet and a GPS used solely to assure that the car was on an approved and pre-mapped route before engaging the ADS. The ADS allowed the vehicle to operate at a maximum speed of 45 mph (p. 7), to travel only on urban and rural roads, and under all lighting and weather conditions except for snow accumulation. The ADS system was easily disengaged; until then, almost all of its data was recorded (the exception noted below of lost data occurred whenever an alternative determination of an object was made by ADS; e.g., shifting from an "object" in the road to an oncoming "vehicle" ahead).

The ADS constructed a virtual environment from the objects that its sensors detected, tracked, classified and then prioritized based on fusion processes (p. 8). ADS predicted and detected any perceived object's goals and paths as part of its classification system. However, if classifications were made and then changed as happened in this case (e.g., from "object" to "vehicle" and back to "object"), the prior tracking history was discarded, a flaw since corrected; also, pedestrians outside of a crosswalk were not assigned a predicted track, another flaw since corrected.

When ADS detected an emergency (p. 9), it suppressed any action for one second to avoid false alarms. After the 1-s delay, the car's self-braking or evasion could begin, a major flaw since corrected (p. 15). If a collision could not have been avoided, an auditory warning was to be given to the operator at the same time that the vehicle was to be slowed (in the case study, the vehicle may have also begun to slow because an intersection was being approached).

Using the recorded data to replay the accident, before impact, radar first detected the pedestrian 5.6 s before impact; lidar made its first detection at 5.2 s, classified the object as unknown and static, changed to a static vehicle at 4.2 s on a path predicted to be a miss, reclassified to "other" and static but back again to vehicle between 3.8 and 2.7 s, each re-classification discarding its previous prediction history for that object; then a bicycle but static and a miss at 2.6 s; then unknown, static and a miss at 1.5s; then a bicycle and an unavoidable hazard at 1.2 s, the categorization of a hazard immediately initiating "action suppression"; after the 1 s pause, finally an auditory alert was sounded at 0.2 s; the operator took control at 0.02 s before impact; and the operator selected brakes at 0.7 s after impact.

3.2 NTSB Notes

- The indecisiveness of the ADS was partly attributed to the pedestrian not being in a crosswalk, a feature the system was not designed to address (p. 12), since corrected.
- The ADS failed to correctly predict the detected object's path, and only determined it to be a hazard at 1.2 s before impact, causing any action to be suppressed for 1.0 s but, and as a consequence of the impact anticipated in the shortened time-interval remaining before impact, exceeding the ADS design specifications for braking and thus not enacted; after this self-imposed 1.0s delay, an auditory alert was sounded (p. 12).
- For almost 20 min before impact, the HMI presented no requests for its human operator's input (p. 13), likely contributing to the human operator's sense of complacency.

3.3 NTSB Lessons Learned

Several lessons were learned and discussed in the NTSB report.

- The operator was distracted by her personal cell phone ([12], p. v); [9] the pedestrian's blood indicated that she was impaired from drugs and that she violated Arizona State's policy by jaywalking.
- Uber had inadequate safety risk assessments of its procedures, ineffective oversight in real-time of its vehicle operators to determine whether they were being complacent, and exhibited overall an inadequate safety culture (p. vi; see also [60]).
- The Uber ADS was functionally limited, unable to correctly classify the object as a pedestrian, to predict her path, or to adequately assess its risk until almost impact.
- The ADS's design to suppress action for 1 s to avoid false alarms increased the risk of driving on the roads and prevented the brakes from being applied immediately to avoid a hazardous situation. Volvo's ADS was partially disabled to prevent conflicts with its radar which operated on the same frequency as the radar for Uber's ATG-ADS (p. 15).
- By disconnecting the Volvo car's own safety systems, however, Uber increased its systemic risk by eliminating the redundant safety systems for its ADS, since corrected (p. vii).
- According to NTSB's decision, although the National Highway Traffic Safety Administration (NHTSA) had published a third version of its automated vehicles policy, NHTSA provided no means to a self-driving company of evaluating its vehicle's ADS to meet national or State safety regulations, or to provide a company with the detailed guidance to design an adequate ADS to operate safely. NTSB recommended that safety assessment reports submitted to NHTSA, voluntary at the time of NTSB's final report, be made mandatory (p. viii) and uniform across all states; e.g., Arizona had taken no action by the time that NTSB's final report was published.

---

[8]In this section, page numbers in parenthesis refer to the NTSB [59] report.

[9]In this section, page numbers in parenthesis refer to the NTSB [12] report.

3.4 Three More Case Studies

Several other case studies could be addressed; e.g., Tesla's advanced driver Autopilot failed to detect a truck's side as it entered the roadway [61]; a distracted Tesla driver's autopilot drove through a stop sign, but the car did not alert its distracted driver [62]; and, for the first time, vehicle manslaughter charges have been filed against the driver of a Tesla for misusing its autopilot by running a red light and crashing into another car, killing two persons [63]. These case studies signal that a new technology has arrived and that we must master its arrival with physics to generalize it to an autonomous HMS that is safe and effective.

3.5 Current Research Gaps Summarized

Putting aside issues important to NTSB and the public, from a human-machine team's perspective, by both being independent of each other, the Uber car and its human operator formed an inferior team [15]. Human teams are autonomous, the best being highly interdependent [10], and not exclusively context dependent (currently, however, machine learning models are context dependent, operating in fully defined and carefully curated certain contexts; in [6]). For technology and civilization to continue to evolve [64], what does autonomy require for future human-machine teams and systems? Facing uncertain situations, the NTSB report indirectly confirmed that no single human or machine agent can determine context alone, nor, presently, unravel by themselves the cause of an accident as complex as the Uber fatality (see also [15]); however, resolving uncertainty requires at a minimum a collective goal, a shared experience, and a state of interdependence that integrates these with information from the situation; moreover, autonomy needs the ability to adapt to rapid changes in context [2], and, overall, to operate safely and ethically as an autonomous human-machine system resolves the uncertainty it faces. We know that the findings of Cummings [10] contradict Conant's [65] generalization of Shannon to minimize the interdependence occurring in teams and organizations. And to reduce uncertainty and increase situation awareness, trust and mutual understanding in an autonomous system necessitates that human and machine teammates are able to explain to, or debate with, each other, however imperfectly, their views of reality in causal terms [41,42]. To operate interdependently, humans and machines must share their experiences in part by training, operating and communicating together. To prevent fatalities like those reflected in the case studies requires interdependence. Otherwise, functional independence will lead to more mistakes like those explored by the NTSB about the Uber self-driving car's pedestrian fatality.

3.5.1 A Deeper Analysis

In summary, despite interdependence having originated in social science, by focusing on the i.i.d. data derived from independent individuals in closed system experiments [57], the different schools in social science have been of limited help in advancing the science of interdependence. For example, if the members of a team when interdependent are more productive than the same individuals in a team but who act independently of

each other [1,2,7,10], then studying how to increase or decrease the quantity of interdependence and its effects becomes a fundamental issue. However, although Lewin [14] founded social psychology to study interdependence in groups, an Editorial by the new editor of the Journal of Personality and Social Psychology (JPSP): Interpersonal Relations and Group Processes, JPSP seeks to publish articles that reflect that "our field is becoming a nexus for social-behavioral science on individuals in context" [66]; Leach's shift towards independence further removes the Journal's founding vision away from the theory of interdependence established by Lewin [14]. Fortunately, the Academy has rejected this regressive shift [1,2].

4 POTENTIAL DEVELOPMENTS. THE MOVE TOWARDS A THEORY OF INTERDEPENDENT AUTONOMY

In systems engineering, structures have here-to-fore been treated as static, physical objects more often designed by computer software with solutions identified by convergence (i.e., Model-Based Engineering, Slide-16; in [67]). In this model, function is a structure's use (SL16), and dynamics is a system's behavior over time (SL26). However, we have found that the structure of an autonomous team is not fixed; e.g., adding redundant, unnecessary members to a fluid team adversely reduces the interdependence between teammates and a team's productivity [5]. In fact, in business mergers, it is common for teams to discard or replace dysfunctional teammates to reach an optimum performance, the motivation for organizations sufficiently free to be able to gain new partners to improve competitiveness, or to spin-off losing parts of a complex business.

4.1 Potential Developments. The Move Towards a Theory of Interdependent Autonomy

We next consider whether there is a thermodynamic advantage in the structure of an autonomous human-machine participants in a team interdependent on their team's performance.

To better make the point, we begin with a return to the history of interdependence, surprisingly by a brief discussion of quantum theory. In 1935 (p. 555), Schrödinger wrote about quantum theory by describing entanglement:

> ... the best possible knowledge of a *whole* does not necessarily include the best possible knowledge of all its *parts,* even though they may be entirely separate and therefore virtually capable of being 'best possibly known' ... The lack of knowledge is by no means due to the interaction being insufficiently known ... it is due to the interaction itself. ...

Similarly, Lewin [14], the founder of Social Psychology, wrote that the "whole is greater than the sum of its parts."

Likewise, from the *Systems Engineering Handbook* [11], "A System is a set of elements in interaction" [68] where systems ". . . often exhibit emergence, behavior which is meaningful only when attributed to the whole, not to its parts" [69].

There is more. Returning to Schrödinger (p. 555),

> Attention has recently been called to the obvious but very disconcerting fact that even though we restrict the disentangling measurements to *one* system, the representative obtained for the *other* system is by no means independent of the particular choice of observations which we select for that purpose and which by the way are *entirely* arbitrary.

If the parts of a team are not independent [8], and if the parts of a whole cannot be disaggregated ([2], p. 11), does a state of interdependence among the orthogonal, complementary parts of a team confer an advantage to the whole [15]? An answer comes from the science of human teams: Compared to a collection of the same but independent scientists, the members of a team of scientists when interdependent are significantly more productive [1,10]. Structurally, to achieve and maintain maximum interdependence, a team must not have superfluous teammates [5]; that is, a team must have the least number of teammates necessary to accomplish its mission. The physics of an autonomous whole, then, means a loss of independence among its parts; i.e., the independent parts must fit together into a "structural" whole, characterized by a reduction in the entropy produced by the team's structure [5]. Thus, for an autonomous whole to be greater than the sum of its individual parts, unlike most practices in social science (the exception being commercial airliner teams; in [13]) or systems engineering, structure and function must be treated interdependently [5].

For a ground state, when a team's structure is stable and existing at a low state of emotion, **Eq. 1** captures Lewin's [14] notion that the whole, S, is greater than the sum of its parts, (S_i), and System Engineering's conjecture of the emergence of synergy, both occurring when the whole produces less entropy than the sum of its parts:

$$S_{Whole} \leq \sum_{i=1}^{n} S_i \qquad (1)$$

In contrast, an excited state occurs with internal conflict in a structure [70], when teammates are independent of each other, or when emotion courses through a team as happened with the tragic drone strike in Afghanistan on 29 August 2021 [71], then the whole becomes less than the sum of its parts as all of a team's free energy is consumed by individuals heedless of their rush to judgment, captured by **Eq. 2**:

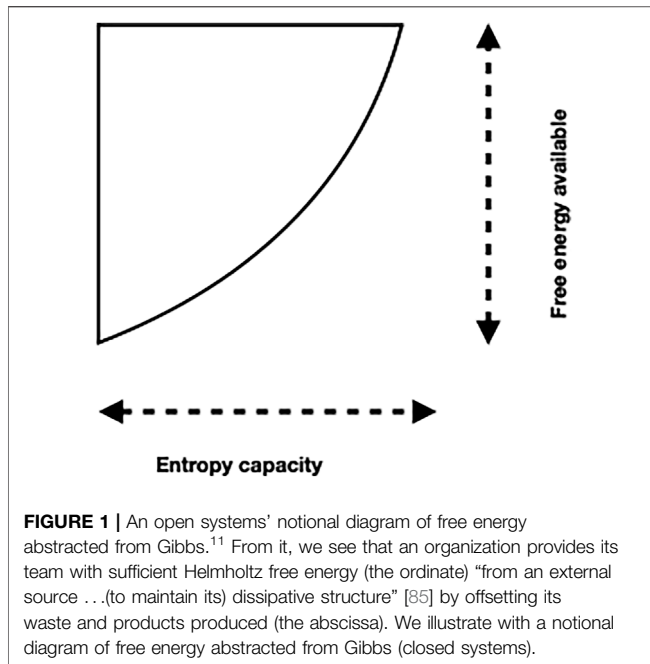$$S_{Whole} \geq \sum_{i=1}^{n} S_i \qquad (2)$$

Interdependence theory guides us to conclude that the intelligent interactions of teammates requires that the teammates be able to converse in a bidirectional causal language that all teammates in an autonomous system can understand; viz., intelligent interactions guide the team to choose teammates that best fit together. In the limit as the parts of a whole become a whole [15], the entropy generated by an autonomous team's or system's whole structure must drop to a minimum to signify the well-fitted team, allowing the mission of the best teams to maximize performance (maximum entropy production, or MEP; in [72]); e.g., by overcoming the obstacles faced [73]; by exploring solution space for a patent [74]; or by merging with another firm to reduce a system's vulnerability. [10] In autonomous systems, characterizing vulnerability in the structure of a team, system or an opponent was the job that Uber failed to perform in a safety analysis of its self-driving car; instead, it became the job that NTSB performed for the Uber team. But as well, the Uber self-driving car and its operator never became a team, remaining as independent parts of a whole (viz., **Eq. (2)**); nor did the Uber car recognize that its operator had become complacent and that the Uber car needed to take an action to protect itself, its human operator and the pedestrian it was about to strike by stopping safely [9]. Unfortunately, even with intelligence being designed into autonomous cars, vehicles are still being designed as tools for human drivers and not as collaborative human-machine teams. Until the car and driver collaboratively learn, train, work and share experiences together interdependently, such mistakes will continue to occur.

This review fits with a call for a new physics of life to study "state-dependent dynamics" (e.g., an example may be quantum biology; in [4]), another call for a new science of social interaction [75], for how humans interact socially with machines (e.g., the CASA paradigm, where human social reactions to computers was studied, in [76–78])), and another to move beyond i.i.d. data [57] in the pursuit of a new theory of information value [79]. The problems with applying social science and Shannon's information theory to teams and systems are becoming clearer as part of an interdisciplinary approach to a new science of autonomous human-machine teams and systems, leading us to focus on managing the positive and negative effects of interdependence. One of the end results, for which we strive in the future, is the new science of information value [80].

In sum, as strengths of interdependence, we have proposed that managing interdependence with AI is critical to the mathematical selection, function and characterization of an aggregation of agents engineered into an intelligent, well-performing unit, achieving MEP in a complementary tradeoff with structure, like the focusing of a telescope. Once that state occurs, disaggregation for analysis of how the parts contribute to a team's success is not possible ([2], p. 11). Interdependence also tells us that each person or machine must be selected in a trial-and-error process, meaning that the best teams cannot be replicated, but they can be identified [5]; and, second, the information for a successful, well-fitted team cannot be obtained in static tests but is only available from the dynamics afforded by the competitive situations in the field able to stress a team's structure as it performs its functions autonomously; i.e., not every good idea for a new structure succeeds in reality

---

[10]For example, Huntington Ingalls Industries has purchased a company focused on autonomous systems [114].

FIGURE 1 | An open systems' notional diagram of free energy abstracted from Gibbs.[11] From it, we see that an organization provides its team with sufficient Helmholtz free energy (the ordinate) "from an external source …(to maintain its) dissipative structure" [85] by offsetting its waste and products produced (the abscissa). We illustrate with a notional diagram of free energy abstracted from Gibbs (closed systems).

(e.g., a proposed health venture became "unwieldy," in [81]). This conclusion runs contrary to matching theory (e.g. [82]) and rational theory [40]. But it holds in the face of uncertainty and conflict for autonomous systems (cf [48]), including autonomous driving (e.g. [83]).

4.2 A Brief Model of Team Entropy Introduced, With Generalizations

In **Figure 1**, we have proposed a model of the entropy production by teams from the application of the free energy available to an autonomous human-machine team to be able to conduct its teamwork and perform its mission (e.g., [46,84]).

Interdependence between structure and performance implies that a limited amount of free energy is available for a team to care for its teammates and perform its mission. For an open-system model of teams, we propose that interdependence between structure and performance uses the free energy available to create a trade-off between uncertainty in the entropy produced by the structure of an autonomous human-machine system and uncertainty in its performance [46,84]:

$$\Delta(structure) * \Delta(performance) \sim C \qquad (3)$$

In **Eq. 3**, uncertainty in the entropy produced by a team's structure times uncertainty in the entropy produced by the performance of the team is approximately equal to a constant, C. As structural costs minimize, the emergence of synergy occurs as the team's performance increases to a maximum, increasing its power.

The predictions from **Eq. 3** are counterintuitive. Applying it to concepts and action results in a tradeoff: as uncertainty in a concept converges to a minimum, the overriding goal of social scientists, uncertainty in the behavioral actions covered by that concept increase exponentially, rendering the concept invalid, the result that has been found for numerous concepts; e.g., self-esteem [23]; implicit attitudes [28]; ego-depletion [26]. These problems with concepts have led to the widespread demand for replication [24]. But the demand for replication more or less overlooks the larger problem with the lack of generalizability arising from what amounts to the use of strictly independent data collected from individual agents [57], which we have argued, cannot recreate the social effects being observed or captured.

In contrast, with **Eq. 3**, interdependence theory generalizes to several effects. To illustrate, we briefly discuss authoritarianism; risk perception; mergers; deception; rational; and vulnerability and emotion.

4.2.1 Authoritarianism

Authoritarians attempt to reduce social noise by minimizing structural effects under their control. However, instead of seeking the best teams with trail and error processes, authoritarians and gangs seek the same effect by suppressing alterative views, social strife, social conflict, etc. Consequently, these systems are unable to innovate [74]. Two examples are given by China and Amazon: Enforced cooperation in China increases its systemic vulnerability to risk and its need to steal innovations (e.g., [86,87]). Similarly, monopolies increase their organizational vulnerability to risk and their need to *steal* innovations from their clients (e.g., Amazon, in [88]).

4.2.2 Risk Determination vs. Risk Perception

Applying **Eq. 1** first to the risk determination of an uncertain event and then to the subjective risk perception of the same event, not surprisingly, the two risks may not agree. For example, Slovic et al. [89] found large differences between the risks determined by experience and calculations vs. the perceived risks associated with nuclear wastes. In the case of the tragic drone attack in Afghanistan by the US Air Force that killed an innocent man and several children, the risk assessment was driven by risk perceptions that led to an emotional rush to judgment, leading to a tragic result [71].

Human observers can generate an infinite spectrum of possible interpretations or risk perceptions, including non-sensical and even dangerous ones as experienced by DoD's [71] unchallenged decision to launch what became its very public and tragic drone attack. Humans have developed two solutions to this quandary: suppress all but the desired perception, e.g., with authoritarian leader's or monopolist's rules that preclude action except theirs [90] or battle-test the risk perceptions in a competitive debate between the chosen perception and its competing alternative perceptions, deciding the best with majority rules [91]. DoD [71] attributed its failure to its own suppression of alternative interpretations. After its failed drone attack, the Air Force concluded that it needed to test both risk assessments and risk perceptions before launching new drone

attacks. The Air Force concluded that one way to test these judgments is to debate the decision before the launch of a drone (i.e., with the use of "red teams").

4.2.3 Mergers

Several reviews of mergers and acquisitions over the years have found mostly failures; e.g., Sirower [92] concluded that two-thirds of mergers were ultimately unsuccessful. As an example of failure, America Online (AOL) acquired Time Warner in 2001, a megamerger that almost doubled the size of AOL, but the merged firm began to fail almost immediately. From **Eq. 3**, the entropy generated by its new structure could not be minimized by AOL and in fact grew, leading to an extraordinary drop in performance in 2002, the loss of Time Warner in 2009, and a depleted AOL's acquisition by Verizon in 2015. In comparison, Apple, one of the most successful companies in the world, acquires a company or more every few weeks, usually as a faction of a percent of Apple's size, the new firms quickly absorbed [93]. From this comparison, we conclude that it is not possible to determine how a new teammate will work out, requiring a trial and error process, the best fit characterized by **Eq. 3** as a reduction in entropy signifying the fit, most likely when a target company provides a function not available to a firm but that complements it orthogonally [46].

Numerous other examples exist. UPS plans to spin-off its failing truck business [94]. Fiat Chrysler's merger with PSA to form Stellantis in 2021 was designed to better compete in its market [95]. Facebook's merger now plans to shore up its vulnerability after Apple revised its privacy policy, which adversely affected Facebook's advertisement revenue [96]. Mergers can also be forced by a government, but the outcome may not be salutary (e.g., Didi's ride-hailing business has been forced by the Chinese government to allow its representatives to participate in Didi's major corporate decisions; in [97]).

4.2.4 Deception

Equation 3 tells us that the best deceivers do not stand out, but instead, fit into a structure as if they belong. One of the key means of using deception is to infiltrate into a system, especially in computational or cyber-security systems [98]. If done by not disturbing the structure of a system or team, deception applied correctly will not increase the structural entropy generated by a team or system, allowing a spy to practice its trade undetected. From Tzu [99], to enact deception: "Engage people with what they expect; it is what they are able to discern and confirms their projections. It settles them into predictable patterns of response, occupying their minds while you wait for the extraordinary moment—that which they cannot anticipate."

4.2.5 Rational

There is another way to address "rational." As we alluded earlier, the rational can also be formal knowledge [100]. When the "rational" is formal knowledge, it is associated with the effort for logical, analytical reasoning versus the easy non-analytical path of recognition afforded by intuition, which may be incorrect.

As iterated before, humans approach naïve intuition or perception by challenging it, leading to a debate, with the best idea surviving the test [84]. But, in addition, and as a generalization from Shannon [44] that we accept, knowledge produces zero entropy [65]. We generalize Conant's concept by applying it to the structure of a perfect team; we have found that, in the limit, the perfect team's structure minimizes the production of entropy by minimizing its degrees of freedom [46]. If we look at the production of entropy as a tradeoff, minimizing the entropy wasted on its own structure allows, but does not guarantee, that a team has more free energy available to maximize its production of entropy (MEP) in the performance of its mission.

4.2.6 Vulnerability, Internal Conflict and Emotion

The effect of conflict in a team illuminates a team's structural vulnerability (e.g. [70]). Internal conflict in a team is essential to identifying vulnerability by a team's opponents during a competition. Regarding the team itself, training is a means to identify and repair (with mergers, etc.) self-weaknesses in a team to prevent it from being exploited by an opponent.

The open conflict between Apple and Facebook provides an excellent example of targeting a structural vulnerability in Facebook by Apple and publicized during the aggressive competition between these two firms. As reported in the *Wall Street Journal* [101],

> Facebook Inc. will suffer damage to its core business when Apple Inc. implements new privacy changes, advertising industry experts say, as it becomes harder for the social-media company to gather user data and prove that ads on its platform work. The core of Facebook's business, its flagship app and Instagram, would be under pressure, too. The Apple change will require mobile apps to seek users' permission before tracking their activity, restricting the flow of data Facebook gets from apps to help build profiles of its users. Those profiles allow Facebook's advertisers to target their ads efficiently. The change will also make it harder for advertisers to measure the return they get for the ads they run on Facebook—how many people see those ads on mobile phones and take actions such as installing an app, for example.

5 DISCUSSION. THE GAPS IN A THEORY OF INTERDEPENDENCE AND AUTONOMY

Interdependence is an unsolved problem that requires more than traditional social science and systems engineering. It is a hard problem. Jones [20] found that a study of interdependence in the laboratory caused "bewildering complexities." Despite his reservations about interdependence, we review our findings and those from the literature that point to the best path going forward to adopt the physics based approach offered by the phenomenon of interdependence in autonomous teams and systems.

Interdependent teams cannot be disaggregated to rationally approach the parts of a team from an individual performer's perspective ([2], p. 11). But we can observe how teams perform with the team as the unit of analysis, we can reduce redundancy to improve interdependence and performance, and we can add better teammates and replace inferior teammates to improve performance [74,102]. This means that a rational approach on paper to building a team is bound to produce poor results. A trial and error method to see what works in the field is the best approach and that can only be determined by a reduction in structural entropy production with, as part of the tradeoff, increases in maximum entropy production measured by the team's overall performance.

From the National Academy of Sciences [2], solutions to autonomous human-machine problems must be found in the field and under the conditions where the autonomy will operate. There, free choices should govern as opposed to the forced choices offered to participants in games. There, teams and organizations must be free to discard or replace the dysfunctional parts of a team, and free to choose the best choice available among the replacement candidates to test whether a good fit occurs. There, vulnerability is also a concern when competing against another team, tested by selling a company's stock short [103–105]. [12] Namely, a vulnerability is characterized by an increase in structural entropy production [84]. There, emotion becomes a factor: as a vulnerability is exposed, emotion increases above a ground state; e.g., the recent financial sting reported by Meta's Facebook from its billions of dollars in losses caused by Apple's new privacy advertisement policy (in [106]).

When minority control by authoritarian leaders impedes the reorganization of structures designed to maximize performance, it is likely to reduce innovation; e.g., by reducing interdependence after adding redundant workers (e.g. [107]); by constraining the choices available to teams and systems [84]; or by reducing the education available to a citizenry (as in the Middle Eastern North African countries plus Israel, where we found that the more education across a free citizenry, the more innovation a country experienced; in [74]). Authoritarian control (by a gang, a monopoly, a country) can suppress the many supported by a group by using forced cooperation to implement its rules, but the more followers that are forced to cooperate, the more that innovation is impeded.

In contrast, Axelrod [105] concludes based on game theory that competition reduced social welfare: "the pursuit of self-interest by each [participant] leads to a poor outcome for all." This outcome can be avoided, Axelrod argued, when sufficient punishment exists to discourage competition. Perc et al. [109] agree that " we must learn how to create organizations, governments, and societies that are more cooperative and more egalitarian . . . " Contradicting Axelrod, Perc and others, we have found the opposite, that the more competitive is a county, the better is its human-development, its productivity,

and its standard of living [74,102,110]. For example, China's forced cooperation across its system of communes promulgated by its Great Leap Forward program was modeled after [111]:

> the Soviet model of industrialization in China [which failed]. . . . The inefficiency of the communes and the large-scale diversion of farm labour into small-scale industry disrupted China's agriculture seriously, and three consecutive years of natural calamities added to what quickly turned into a national disaster; in all, about 20 million people were estimated to have died of starvation between 1959 and 1962.

By rejecting Axelrod's and China's use of punishment to enforce its minority control, when an interdependence between culture and technology is allowed to freely exist, free expression "reflects interdependent processes of brain-culture co-evolution" [64].

Our study is not exhaustive (e.g., due to the limitations of space, we left out: factorable tensors, implying no interdependence; orthogonality, precluding individuals from being able to multitask; competition generates bistable information; perturbations collapse teams with redundancy or otherwise poorly structured and operated teams; etc.; we also had plans to apply our physics in **Eq. 3** to the U.S. Army's Multi Domains Operations, or MDO, to show that, based on interdependence theory, MDO would be an inferior application for autonomous human-machine teams because its agents are independent, precluding synergy or power from team arrangements). Thus, we have much to study in the future.

6 CONCLUSION

We conclude that the cause of the Uber self-driving car accident was the lack of interdependence between the human operator and the machine. The case studies and theory indicate that no synergy arises when teammates remain independent of each other (**Eqs 1**, **3**). Internal conflict causes a vulnerability in a team that can be exploited by an opponent. In contrast, for autonomy to occur, an HMS must have shared experiences by training, operating, and communicating together to control each other. When that happens, when a structure of a team is stable and producing minimum entropy in a state of interdependence, synergy occurs (a mutually beneficial symbiosis). Self-awareness of each other and of the team must be built during training and continued during operations. Interdependence requires situation awareness of the environment of each other and of the team's performance; trust; sustainable attention; mutual understanding; and communication devices all come in to play (for a review of bidirectional trust, see [8]).

Traditional social science is weakest when it has little to say to improve states of interdependence, strongest when it contributes to its advancement. By not sidestepping the physics of what is occurring in the physical reality of a team, we conclude that a state of interdependence cannot be disaggregated into elements that can then be summed by states of independence to recreate the

---

[12] https://www.investopedia.com/terms/s/shortsale.asp.

interdependent event being witnessed [2]. This happens because interdependence reduces the degrees of freedom for a whole, precluding a simple aggregation of the parts for the whole. A human-machine awareness of each other in a team, and of the team as a team, a human-machine team sharing coordination among its teammates, a human-machine team collaborating together as a team, all are necessary.

In closing, social systems based on closed models used to study independent agents are unable to contribute to the evolution posed by autonomous human-machine systems. The data derived from these models are subjective, whether based on game theory (e.g. [108]), rational choices (e.g. [40]), or superforecasts (e.g. [30]). To be of value, subjective interpretations must be tested, challenged or debated. As we have portrayed in this review, the disruption to social theory posed by human-machine systems is more likely to be revolutionary, a dramatic change that autonomy scientists working with human-machine teams and systems must be prepared to contribute, to manage and to live with.

6.1 Conclusion. The Contribution of the Manuscript to the Literature

This manuscript contributes to the literature by applying basic concepts from physics to an autonomous HMS. **Equation 3** is a metric for the tradeoffs between an autonomous human-machine team's structure and its performance. We also recognize that interdependence is a phenomenon in nature that can be modeled with physics like any other natural phenomenon. We recognize that an autonomous HMS cannot occur with independent agents. And we have postulated that one of the contributions by future machines in an autonomous HMS is by monitoring the emotional states among its human teammates with alerts about distorted situational awareness, by providing an open awareness of their emotional states, and by impeding their haste to decide.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

1. NJ Cooke ML Hilton, editors. *Enhancing the Effectiveness of Team Science. National Research Council.* Washington (DC): National Academies Press (2015).

2. Endsley MR. Human-AI Teaming: State-Of-The-Art and Research Needs. In: *The National Academies of Sciences-Engineering-Medicine.* Washington, DC: National Academies Press (2021). Accessed 12/27/2021 from https://www.nap.edu/catalog/26355/human-ai-teaming-state-of-the-art-and-research-needs.

3. Epstein B. In: EN Zalta, editor. *Social Ontology, the Stanford Encyclopedia of Philosophy.* Stanford, CA: The Metaphysics Research Lab (2021). Accessed 4/1/2022 from https://plato.stanford.edu/entries/social-ontology/.

4. Davies P. Does New Physics Lurk inside Living Matter? *Phys Today* (2021) 73(8). doi:10.1063/PT.3.4546

5. Lawless WF. The Interdependence of Autonomous Human-Machine Teams: The Entropy of Teams, but Not Individuals, Advances Science. *Entropy* (2019) 21(12):1195. doi:10.3390/e21121195

6. Peterson JC, Bourgin DD, Agrawal M, ReichmanGriffiths DTL, Griffiths TL. Using Large-Scale Experiments and Machine Learning to Discover Theories of Human Decision-Making. *Science* (2021) 372(6547):1209–14. doi:10.1126/science.abe2629

7. Cooke NJ, Lawless WF. Effective Human-Artificial Intelligence Teaming. In: WF Lawless, R Mittu, DA Sofge, T Shortell, TA McDermott, editors. *Engineering Science and Artificial Intelligence.* Springer (2021). doi:10.1007/978-3-030-77283-3_4

8. WF Lawless, R Mittu, D Sofge, S Russell, editors. *Autonomy and Artificial Intelligence: A Threat or Savior?* New York: Springer (2017).

9. Sofge D, Lawless W, Mittu R. AI Bookie. *AIMag* (2019) 40(3):79–84. doi:10.1609/aimag.v40i3.5196

10. Cummings J. *Team Science Successes and Challenges.* Bethesda MD: National Science Foundation Sponsored Workshop on Fundamentals of Team Science and Science of Team Science (2015).

11. DD Walden, GJ Roedler, KJ Forsberg, RD Hamelin, TM Shortell, editors. *Systems Engineering Handbook. Prepared by International Council on System Engineering (INCOSE-TP-2003-002-04).* 4th ed. Hoboken, NJ: John Wiley & Sons (2015).

12. NTSB. *Collision between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, AZ, March 18, 2018. National Transportation Safety Board (NTSB), Accident Report*. Washington, DC: NTSB/HAR-19/03. PB2019-101402 (2019).

13. Bisbey TM, Reyes DL, Traylor AM, Salas E. Teams of Psychologists Helping Teams: The Evolution of the Science of Team Training. *Am Psychol* (2019) 74(3):278–89. doi:10.1037/amp0000419

14. Lewin K. In: D Cartwright, editor. *Field Theory of Social Science*. New York: Harper and Brothers (1951).

15. Lawless WF, Mittu R, Sofge D, Hiatt L. Artificial Intelligence, Autonomy, and Human-Machine Teams - Interdependence, Context, and Explainable AI. *AIMag* (2019) 40(3):5–13. doi:10.1609/aimag.v40i3.2866

16. Liu Y-Y, Barabási A-L. Control Principles of Complex Systems. *Rev Mod Phys* (2016) 88(3):035006. doi:10.1103/RevModPhys.88.035006

17. Amadae SM. *Rational Choice Theory. Political Science and Economics*. London, UK: Encyclopaedia Britannica (2017). from https://www.britannica.com/topic/rational-choice-theory. Accessed 12/15/2018.

18. Arora S, Doshi P. *A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress*. Ithaca, NY. arXiv: 1806.06877v3 (2020).

19. Gray NS, MacCulloch MJ, Smith J, Morris M, Snowden RJ. Violence Viewed by Psychopathic Murderers. *Nature* (2003) 423:497–8. doi:10.1038/423497a

20. Jones EE. Major Developments in Five Decades of Social Psychology. In: DT Gilbert, ST Fiske, G Lindzey, editors. *The Handbook of Social Psychology, Vol. I*. Boston: McGraw-Hill (1998). p. 3–57.

21. Thagard P. Cognitive Science. In: EN Zalta, editor. *The Stanford Encyclopedia of Philosophy*. Department of Philosophy, Stanford University (2019). Accessed 11/15/2020 from https://plato.stanford.edu/archives/spr2019/entries/cognitive-science.

22. Diener E. Subjective Well-Being. *Psychol Bull* (1984) 95(3):542–75. doi:10.1037/0033-2909.95.3.542

23. Baumeister RF, Campbell JD, Krueger JI, Vohs KD. Exploding the Self-Esteem Myth. *Sci Am* (2005) 292(1):84–91. doi:10.1038/scientificamerican0105-84

24. Nosek B. Estimating the Reproducibility of Psychological Science. *Science* (2015) 349(6251):943.

25. Baumeister RF, Vohs KD. Self-Regulation, Ego Depletion, and Motivation. *Social Personal Psychol Compass* (2007) 1. doi:10.1111/j.1751-9004.2007.00001.x

26. Hagger MS, Chatzisarantis NLD, Alberts H, Anggono CO, Batailler C, Birt AR, et al. A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspect Psychol Sci* (2016) 11(4):546–73. doi:10.1177/1745691616652873

27. Greenwald AG, McGhee DE, Schwartz JLK. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *J Personal Soc Psychol* (1998) 74(6):1464–80. doi:10.1037/0022-3514.74.6.1464

28. Blanton H, Jaccard J, Klick J, Mellers B, Mitchell G, Tetlock PE. Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT. *J Appl Psychol* (2009) 94(3):567–82. doi:10.1037/a0014665

29. Berenbaum MR. Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End. *PNAS* (2021) 118(38):e2115397118. doi:10.1073/pnas.2115397118

30. Tetlock PE, Gardner D. *Superforecasting: The Art and Science of Prediction*. New York City, NY: Crown (2015).

31. Shu LL, Mazar N, Gino F, Ariely D, Bazerman MH. Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End. *Proc Natl Acad Sci U.S.A* (2012) 109:15197–200. doi:10.1073/pnas.1209746109

32. Lawless W F. Risk Determination versus Risk Perception: A New Model of Reality for Human-Machine Autonomy. *Informat* (2022) 9(2):30. doi:10.3390/informatics9020030

33. Rudd JB. *Why Do We Think that Inflation Expectations Matter for Inflation?* Washington, D.C.: Federal Reserve Board (2021). doi:10.17016/FEDS.2021.062

34. Krugman P. *When Do We Need New Economic Theories?* New York City, NY: New York Times (2022). Accessed 2/8/2022 from https://www.nytimes.com/2022/02/08/opinion/economic-theory-monetary-policy.html.

35. Shiller RJ. *Inflation Is Not a Simple Story about Greedy Corporations*. New York City, NY: New York times (2022). Accessed 2/8/2022 from https://www.nytimes.com/2022/02/08/opinion/dont-blame-greed-for-inflation.html.

36. Smialek J. *Why Washington Can't Quit Listening to Larry Summers*. New York City, NY: New York Times (2021). Accessed 2/8/2022 from https://www.nytimes.com/2021/06/25/business/economy/larry-summers-washington.html.

37. Smialek J. *Is This what Winning Looks like? Modern Monetary Theory*. New York City, NY: New York Times (2022). Accessed 2/8/2022 from https://www.nytimes.com/2022/02/06/business/economy/modern-monetary-theory-stephanie-kelton.html.

38. Leonard C. *The Lords of Easy Money*. New York City, NY: Simon & Schuster (2022).

39. Augier M, Barrett SFX. *General Anthony Zinni (Ret.) on Wargaming Iraq, Millennium Challenge, and Competition*. Washington, DC: Center for International Maritime Security (2021). Accessed 10/21/2021 from https://cimsec.org/general-anthony-zinni-ret-on-wargaming-iraq-millennium-challenge-and-competition/.

40. Mann RP. Collective Decision Making by Rational Individuals. *Proc Natl Acad Sci U S A* (2018) 115(44):E10387–E10396. doi:10.1073/pnas.1811964115

41. Pearl J. Reasoning with Cause and Effect. *AI Mag* (2002) 23(1):95. doi:10.1609/aimag.v23i1.1612

42. Pearl J, Mackenzie D. *AI Can't Reason Why*. New York City: Wall Street Journal (2018). Accessed Apr. 27, 2020. from https://www.wsj.com/articles/ai-cant-reason-why-1526657442.

43. NTSB. *Preliminary Report Released for Crash Involving Pedestrian*. Washington, DC: Uber Technologies, Inc. (2019). Accessed 5/1/2019 from https://www.ntsb.gov/news/press-releases/Pages/NR20180524.aspx.

44. Shannon CE. A Mathematical Theory of Communication. *Bell Syst Tech J* (1948) 27:379–423. , 623–656. doi:10.1002/j.1538-7305.1948.tb01338.x

45. Eagleman DM. Visual Illusions and Neurobiology. *Nat Rev Neurosci* (2001) 2(12):920–6. doi:10.1038/35104092

46. Lawless WF. Quantum-Like Interdependence Theory Advances Autonomous Human-Machine Teams (A-HMTs). *Entropy* (2020) 22(11):1227. doi:10.3390/e22111227

47. Weinberg S. Steven Weinberg and the Puzzle of Quantum Mechanics. In: N David Mermin, J Bernstein, M Nauenberg, J Bricmont, S Goldstein, editors. *In Response to: The Trouble with Quantum Mechanics from the January 19, 2017 Issue*. New York City: The New York Review of Books (2017).

48. Hansen LP. Nobel Lecture: Uncertainty outside and inside Economic Models. *J Polit Economy* (2014) 122(5):945–87. doi:10.1086/678456

49. Walch M, Mühl K, Kraus J, Stoll T, Baumann M, Weber M. From Car-Driver-Handovers to Cooperative Interfaces: Visions for Driver-Vehicle Interaction in Automated Driving. In: G Meixner C Müller, editors. *Automotive User Interfaces*. Springer International Publishing (2017). p. 273–94. doi:10.1007/978-3-319-49448-7_10

50. Christoffersen K, Woods D. 1. How to Make Automated Systems Team Players. *Adv Hum Perform Cogn Eng Res* (2002) 2:1–12. doi:10.1016/s1479-3601(02)02003-9

51. Winner H, Hakuli S. Conduct-by-wire–following a New Paradigm for Driving into the Future. *Proc FISITA World Automotive Congress* (2006) 22:27.

52. Flemisch FO, Adams CA, Conway SR, Goodrich KH, Palmer MT, Schutte PC. *The H-Met-Aphor as a Guideline for Vehicle Automation and Interaction*. Washington, DC: NASA/TM-2003-212672 (2003).

53. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press (1953). (originally published in 1944).

54. Thibaut JW, Kelley HH. *The Social Psychology of Groups*. New York: John Wiley & Sons (1959).

55. Kelley HH, Thibaut JW. *Interpersonal Relations. A Theory of Interdependence*. New York: Wiley (1978).

56. Kelley HH. Lewin, Situations, and Interdependence. *J Soc Issues* (1991) 47:211–33. doi:10.1111/j.1540-4560.1991.tb00297.x

57. Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. *Ithaca, NY: Towards Causal Representation Learning*. arXiv, Accessed 7/6/2021 (2021).

58. Kenny DA, Kashy DA, Bolger N, Gilbert DT, Fiske ST, Lindzey G. Data Analyses in Social Psychology. In: *Handbook of Social Psychology*. 4th ed. Boston, MA: McGraw-Hill (1998). p. 233–65.

59. NTSB. *Vehicle Automation Report*. Washington, DC: National Transportation Safety Board (2019). Accessed 12/3/2020 from https://dms.ntsb.gov/pubdms/search/document.cfm?docID=477717anddocketID=62978andmkey=96894.

60. NTSB. *Inadequate Safety Culture' Contributed to Uber Automated Test Vehicle Crash*. Washington, DC: National Transportation Safety Board (2019). Accessed 4/11/2020 from https://www.ntsb.gov/news/press-releases/Pages/NR20191119c.aspx.

61. Hawkins AJ. *Tesla's Autopilot Was Engaged when Model 3 Crashed into Truck, Report States*. New York City: The Verge (2019). Accessed 3/27/2022 from https://www.theverge.com/2019/5/16/18627766/tesla-autopilot-fatal-crash-delray-florida-ntsb-model-3.

62. Boudette NE. *'It Happened So Fast': Inside a Fatal Tesla Autopilot Accident*. New York City, NY: New York Times (2021). Accessed 3/27/2022 from https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html.

63. NPR. *A Tesla Driver Is Charged in a Crash Involving Autopilot that Killed 2 People*. Washington, DC: NPR (2022). Accessed 3/27/2022 from https://www.npr.org/2022/01/18/1073857310/tesla-autopilot-crash-charges?t=1647945485364.

64. Ponce de León MS, Bienvenu T, Marom A, Engel S, Tafforeau P, Alatorre Warren JL, et al. The Primitive Brain of Early Homo. *Science* (2021) 372(6538):165–71. doi:10.1126/science.aaz0032

65. Conant RC. Laws of Information Which Govern Systems. *IEEE Trans Syst Man Cybern* (1976) SMC-6:240–55. doi:10.1109/tsmc.1976.5408775

66. Leach CW. Editorial. *J Personal Soc Psychol* (2021) 120(1):30–2. doi:10.1037/pspi0000226

67. Douglass BP. *"Introduction to Model-Based Engineering?" Senior Principal Agile Systems Engineer*. Mitre (2021). Accessed 2/13/2022 from https://www.incose.org/docs/default-source/michigan/what-does-a-good-model-smell-like.pdf?sfvrsn=5c9564c7_4.

68. Bertalanffy L. *General System Theory: Foundations, Development, Applications, Rev*. New York: Braziller (1968).

69. Checkland P. *Systems Thinking, Systems Practice*. New York: Wiley (1999).

70. Vanderhaegen F. Heuristic-based Method for Conflict Discovery of Shared Control between Humans and Autonomous Systems - A Driving Automation Case Study. *Robotics Autonomous Syst* (2021) 146:103867. doi:10.1016/j.robot.2021.103867

71. DoD. Pentagon Press Secretary John F. Kirby and Air Force Lt. Gen. Sami D. Said Hold a Press Briefing (2021). Accessed 11/3/2021 from https://www.defense.gov/News/Transcripts/Transcript/Article/2832634/pentagon-press-secretary-john-f-kirby-and-air-force-lt-gen-sami-d-said-hold-a-p/.

72. Martyushev L. Entropy and Entropy Production: Old Misconceptions and New Breakthroughs. *Entropy* (2013) 15:1152–70. doi:10.3390/e15041152

73. Wissner-Gross AD, Freer CE. Causal Entropic Forces. *Phys Rev Lett* (2013) 110(168702):168702–5. doi:10.1103/PhysRevLett.110.168702

74. Lawless WF. Towards an Epistemology of Interdependence Among the Orthogonal Roles in Human-Machine Teams. *Found Sci* (2019) 26:129–42. doi:10.1007/s10699-019-09632-5

75. Baras JS. *Panel. New Inspirations for Intelligent Autonomy*. Washington, DC: ONR Science of Autonomy Annual Meeting (2020).

76. Nass C, Fogg BJ, Moon Y. Can Computers Be Teammates? *Int J Human-Computer Stud* (1996) 45(6):669–78. doi:10.1006/ijhc.1996.0073

77. Nass C, Moon Y, Fogg BJ, Reeves B, Dryer DC. Can Computer Personalities Be Human Personalities? *Int J Human-Computer Stud* (1995) 43(2):223–39. doi:10.1006/ijhc.1995.1042

78. Nass C, Steuer J, Tauber ER. Computers Are Social Actors. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York City: ACM (1994). p. 72–8. doi:10.1145/191666.191703

79. forthcoming Blasch E, Schuck T, Gagne OB. Trusted Entropy-Based Information Maneuverability for AI Information Systems Engineering. In: WF Lawless, J Linas, DA Sofge, R Mittu, editors. *Engineering Artificially Intelligent Systems*. Springer's Lecture Notes in Computer Science (2021).

80. Moskowitz IS, Brown NL, Goldstein Z. A Fractional Brownian Motion Approach to Psychological and Team Diffusion Problems. In: WF Lawless, R Mittu, DA Sofge, T Shortell, TA McDermott, editors. *Systems Engineering and Artificial Intelligence*. Springer (2021). Chapter 11. doi:10.1007/978-3-030-77283-3_11

81. Herrera S, Chin K. *Amazon, Berkshire Hathaway, JPMorgan End Health-Care Venture Haven*. New York City: Wall Street Journal (2021). Accessed 1/5/2021 from https://www.wsj.com/articles/amazon-berkshire-hathaway-jpmorgan-end-health-care-venture-haven-11609784367.

82. McDowell JJ. On the Theoretical and Empirical Status of the Matching Law and Matching Theory. *Psychol Bull* (2013) 139(5):1000–28. doi:10.1037/a0029924

83. Woide M, Stiegemeier D, Pfattheicher S, Baumann M. Measuring Driver-Vehicle Cooperation. *Transportation Res F: Traffic Psychol Behav* (2021) 83:773 424–39. doi:10.1016/j.trf.2021.11.003

84. Lawless WF, Sofge D. Interdependence: A Mathematical Approach to the Autonomy of Human-Machine Systems. In: *Proceedings for Applied Human Factors and Ergonomics 2022*. Springer (2022).

85. Prigogine I. Ilya Prigogine. Facts. The Nobel Prize in Chemistry 1977 (1977). Accessed 2/10/2022 from https://www.nobelprize.org/prizes/chemistry/1977/prigogine/facts/.

86. Baker G. *Michael Hayden Says U.S. Is Easy Prey for Hackers. Former CIA and NSA Chief Says 'shame on Us' for Not Protecting Critical Information Better*. New York City: Wall Street Journal's editor in chief (2015). from http://www.wsj.com/articles/michael-hayden-says-u-s-is-easy-prey-for-hackers-1434924058. Accessed 6/22/2015

87. Ratcliffe J. *U.S. Director of National Intelligence: "China Is National Security Threat No. 1"*. New York City: Wall Street Journal (2020). Accessed 4/4/2022 from https://www.wsj.com/articles/china-is-national-security-threat-no-1-11607019599.

88. Mattioli D. *Amazon Scooped up Data from its Own Sellers to Launch Competing Products*. New York City: Wall Street Journal (2020). Accessed 10/18/2021 from https://www.wsj.com/articles/amazon-scooped-up-data-from-its-own-sellers-to-launch-competing-products-11587650015?mod=article_inline.

89. Slovic P, Flynn JH, Layman M. Perceived Risk, Trust, and the Politics of Nuclear Waste. *Science* (1991) 254:1603–7. doi:10.1126/science.254.5038.1603

90. Lawless WF, Bergman M, Feltovich N. Consensus-Seeking Versus Truth-Seeking. *ASCE Practice Periodical Hazard Toxic Radioact Waste Manage* (2005) 9(1):59-70.

91. Lawless WF, Akiyoshi M, Angjellari-Dajcic F, Whitton J. Public Consent for the Geologic Disposal of Highly Radioactive Wastes and Spent Nuclear Fuel. *Int J Environ Stud* (2014) 71(1):41-62.

92. Sirower ML. *The Synergy Trap*. New York: Free Press (2007).

93. Leswing K. *How Apple Does MandA*. Englewood Cliffs, NJ: CNBC (2021). Accessed 4/4/2022 from https://www.cnbc.com/2021/05/01/how-apple-does-ma-small-and-quiet-with-no-bankers.html.

94. Smith J, Ziobro P. *UPS to Sell Freight Trucking Business to TFI*. New York City: Wall Street Journal (2021). Accessed 1/26/2021 from https://www.wsj.com/articles/ups-to-sell-freight-trucking-business-to-tfi-for-800-million-11611592797.

95. Naughton N. *Fiat Chrysler, PSA Aim to Complete Trans-Atlantic Merger in Mid-january*. New York City: Wall Street Journal (2021). Accessed 1/5/2021 from https://www.wsj.com/articles/fiat-chrysler-psa-shareholders-to-vote-on-trans-atlantic-tie-up-11609764732.

96. Tweh B. *Facebook Says Apple's Privacy Changes Hurt Digital Ad Measurement*. New York City: Wall Street Journal (2021). Accessed 2/13/2022 from https://www.wsj.com/articles/facebook-says-apples-privacy-changes-hurt-digital-ad-measurement-11632341624.

97. Webb Q, Wei L. *Chinese Ride-Hailing Giant Did I Could Get State Investment*. New York City: Wall Street Journal (2021). Accessed 9/4/2021 from https://www.wsj.com/articles/city-of-beijing-leads-plan-for-state-investment-in-didi-the-embattled-ride-hailing-giant-1163068135.

98. Lawless WF, Mittu R, Moskowitz IS, Sofge D, Russell S. Cyber-(in)security, Revisited. In: P Dasgupta, JB Collins, R Mittu, editors. *Adversary Aware Learning Techniques and Trends in Cyber Security*. Switzerland: Springer Nature (2020).

99. Tzu S. In: RD Sawyer, editor. *The Art of War*. New York: Basic Books (1994). First published in the 5th Century, BC.

100. Kryjevskaia M, Heron PRL, Heckler AF. Intuitive or Rational? Students and Experts Need to Be Both. *Phys Today* (2022) 74(8):28–34. doi:10.1063/PT.3. 4813

101. Haggin P, Hagey K, Schechner S. *Apple's Privacy Change Will Hit Facebook's Core Ad Business*. New York City: Wall Street Journal (2021). Accessed 1/30/ 2021 https://www.wsj.com/articles/apples-privacy-change-will-hit-facebooks-core-ad-business-heres-how-11611938750.

102. Lawless WF. The Physics of Teams: Interdependence, Measurable Entropy, and Computational Emotion. *Front Phys* (2017) 5:30. doi:10.3389/fphy.2017. 00030

103. Platt HD, Fuller A. Theory of Short Selling. *J Asset Manag* (2002) 5(1). Accessed 2/18/2022 from. doi:10.2139/ssrn.301321

104. Farooq F. *10 Most Successful Short Sellers of All Time*. Sunnyvale, CA: Yahoo.com (2021). Accessed 2/12/2022 from https://www.yahoo.com/ video/10-most-successful-short-sellers-141904957.html.

105. Kennon J. The Basics of Shorting Stock, the Balance (2021). Accessed 2/ 18,2022 from https://www.thebalance.com/the-basics-of-shorting-stock-356327.

106. Bobrowsky M. *Facebook Feels $10 Billion Sting from Apple's Privacy Push*. New York City: Wall Street Journal (2022). Accessed 2/6/2022 from https:// www.wsj.com/articles/facebook-feels-10-billion-sting-from-apples-privacy-push-11643898139.

107. Lawless WF, Sofge DA. The Intersection of Robust Intelligence and Trust. In: WFR LawlessMittu, D Sofge, S Russell, editors. *Autonomy and Artificial Intelligence: A Threat or Savior?* New York: Springer (2017). p. 255–70.

108. Axelrod R. *The Evolution of Cooperation*. New York: Basic (1984).

109. Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. Statistical Physics of Human Cooperation. *Phys Rep* (2017) 687:1–51. doi:10.1016/j. physrep.2017.05.004

110. Friedman M. *Capitalism and freedom*. Chicago: New York City: University of Chicago Press (1982).

111. Britannica. The Editors of Encyclopaedia. In: *Great Leap Forward*. London, UK: Encyclopedia Britannica (2020). https://www.britannica.com/event/ Great-Leap-Forward (Accessed February 18, 2022).

112. Llinas J. Motivations for and Initiatives on AI Engineering. In: WF Lawless, R Mittu, DA Sofge, T Shortell, TA McDermott, editors. *Systems Engineering and Artificial Intelligence*. Springer (2021). doi:10.1007/978-3-030-89385-9_1

113. Somerville H. *Uber Sells Self-Driving-Car Unit to Autonomous-Driving Startup*. Wall Street Journal (2020). Accessed 12/8/2020 from https:// www.wsj.com/articles/uber-sells-self-driving-car-unit-to-autonomous-driving-startup-11607380167.

114. Shelbourne M. *HII Purchases Autonomy Company to Bolster Unmanned Surface Business*. Annapolis, MD: USNI News (2021). Accessed 1/5/2021 from https:// news.usni.org/2021/01/04/hii-purchases-autonomy-company-to-bolster-unmanned-surface-business.