The Human Should be Part of the Control Loop?

William D. Nothwang
Sensors and Electron Devices Directorate
Army Research Laboratory
Adelphi, MD 20783
william.d.nothwang.civ@mail.mil

Michael J. McCourt
Department of Aerospace and Mechanical Engineering
University of Florida REEF
Shalimar, FL 32579
mccourt@ufl.edu

Ryan M. Robinson and Samuel A. Burden
Department of Electrical Engineering
University of Washington
Seattle, WA 20783
rymicro@uw.edu; sburden@uw.edu

J. Willard Curtis
Munitions Directorate
Air Force Research Laboratory
Eglin AFB, FL 32542
jess.curtis@us.af.mil

Abstract—**The capabilities of autonomy have grown to encompass new application spaces that until recently were considered exclusive to humans. In the past, automation has focused on applications where it was preferable to completely replace the human. Today, though, we have the opportunity to leverage the complementary strengths of both human and autonomy technologies to maximize performance and limit risk, and the human should therefore remain "in" or "on" the loop. To adequately assess when and how to accomplish this, it requires us to assess not only the capabilities, but the risks and the ethical questions; coupled to this are the issues with degradation of performance in specific instances (for instance, recovery from failure) that may require a human to remain the sole control authority. This paper investigates the contributors to success/failure in current human-autonomy integration frameworks, and proposes guidelines for safe and resilient use of humans and autonomy with regard to performance, consequence, and the stability of human-machine switching. Key to our proposed approach are (i) the relative error rate between the human and autonomy and (ii) the consequence of possible events.**

Keywords—human-autonomy integration, human-machine interaction, shared control, switched control

I. Introduction

Automated systems have outperformed humans for more than a century. For instance, assembly lines [1], autopilots [2], and treatment plant automation [3] have all been employed to improve efficiency and guard against human performance degradation due to fatigue, high workload, complacency, and other factors [4, 5, 6]. However, only in the past few years has *autonomy* begun to outperform humans in tasks assumed to require "human-like" perception and judgement. Such new capabilities have raised new questions regarding what is lost by replacing the human. The use of autonomy in situations where the consequences of failure are significant has been the subject of major debate [7].

A major question for any application that *could* potentially be performed autonomously is: Should the human remain part of the control loop? There are numerous commercial, domestic, and military-relevant tasks/applications where humans cannot or should not be relieved of all responsibility, and must remain either an active participant ("in-the-loop") or take on a supervisory role ("on-the-loop") to ensure safe and effective operation (particularly in the case of failure modes).

The present work addresses the ambiguity in human/ autonomy role designation by investigating the contributors to success and failure in current and proposed human-autonomy control frameworks. We focus on the issues of confidence in performance quality, consequence of control actions, and time-scale issues. First, we examine the scope of applications that could plausibly be performed by autonomy or human-autonomy teams by reviewing the state-of-the-art in autonomy and human-autonomy integration. Second, we categorize human-autonomy systems and review the types of failure modes that may be experienced. We then elaborate on the notion of consequence as it relates to the limits of autonomous decision-making and the need for human involvement. Lastly, we provide guidance for assessing the proper level of human involvement (and consequently, autonomous involvement) across a broad range of potential applications.

II. State-of-the-Art Autonomy and Human-Autonomy Applications

A. The Range Between Human and Autonomy Control

It is important to note that there is a difference between autonomy and automation. Automation is generally the repetition of the same task with no or minimal adaption between repetitions. Autonomy is allowing a system to adapt and make decisions independent from constant human input, and that there are varying degrees of autonomy. It is a continuous spectrum, but for the purposes of this work we will classify the levels of autonomy as: human, human in the loop (HIL), human on the loop (HOL), and complete autonomy (note that level of autonomy can vary between stages of the decision-making process; see [8] for details).

Human (H) control implies that the human is actively involved in all aspects of an agent's task, and this level of autonomy most closely resembles automation. This does not imply that a human does everything themselves, but that the human is involved with all aspects of decision making. For

This effort was partially funded through the Office of the Secretary of Defense Autonomy Research Pilot Initiative (ARPI), MIPR DWAM31168. This effort was also partially funded by NSF CISE CPS CRII Award #1565529.

example, a human may use a calculator or software computation package to perform an analysis, or a human may control all degrees of freedom of a teleoperated robot. In both cases the autonomy is only executing the commands of the human and does not make any but the most rudimentary decisions. In human control: the human has a very high task load for decision making; is the ultimate arbiter of decisions; and is actively involved in all autonomy decisions.

Human-in-the-loop (HIL) control is characterized by the human actively (often continuously) engaging in control decisions. There are many examples of human-autonomy teaming that provide improved performance compared to either humans or autonomy in isolation. Advanced chess, also known as "centaur chess", combines humans actively making decisions that are augmented by autonomous suggestions and simulations, resulting in "extraordinary rises in the level of play at both the tactical and strategic level" [9]. Similarly, recent studies have investigated robotic surgery performed by a surgeon continuously operating a remote interface (i.e. teleoperation) [10, 11] while autonomy is employed for stabilization and virtual fixtures to enhance safety [12]. This is expected to result in shorter hospital stays and smaller incisions. In human-collaborative industrial robot systems, a human worker and robotic manipulator share a physical space [13]. In such systems, safety is a major concern, as collisions with industrial robots can cause injury [14]. In HIL, the human has a large task load for decision making, is the ultimate arbiter of decisions, and is actively involved in most, if not all, agent decisions.

Human-on-the-loop (HOL) control is supervisory control in which the human monitors the operation of autonomy, taking over control only when the autonomy encounters unexpected events or when failure occurs. Facebook M is a virtual assistant that defers to humans when it has low confidence in its ability to perform tasks on its own [15]. Although many existing remotely piloted vehicle systems can be characterized as HIL (with one or more operators per vehicle), HOL architectures are also being implemented so that a single operator can control multiple vehicles that may autonomously re-rout in communications-denied environments, as envisioned in the DARPA CODE program [16]. In HOL, the human has a minimal task load for decision making, is the ultimate arbiter for some, if not all, decisions, and is only actively involved in the most crucial agent decisions.

Complete autonomy (CA) is often considered a sub-task of supervisory control, as it is difficult to envision autonomous agents operating sans any human intervention. This is the "teammate" role, where there may even be a human "captain" calling plays, but the agent understands its role and is allowed to complete it with minimal intervention. In CA, the human has a minimal task load for decision making, is not the ultimate arbiter on decisions, and is only minimally involved in agent decision making.

B. *Human & Autonomy Involvement Considerations*

Remarkable advancements in autonomy have occurred over the past decade. Many of these advancements are a result of novel machine learning techniques that fall under the popular term "deep learning". For instance, image recognition powered by deep convolutional neural networks has surpassed human-level performance [17, 18], and video-game playing using deep reinforcement learning has exceeded human levels on a well-known gaming platform [19]. Google DeepMind recently developed a Go-playing AI (AlphaGo) that trained its value/policy neural networks using both supervised learning from experts and reinforcement learning from self-play. AlphaGo defeated a European Go champion [20], and months later a world Go champion [21], demonstrating marked skill improvement within the short time period. In spite of these historical achievements, there is no evidence that deep learning or other current forms of artificial intelligence will gain human-level general intelligence. Therefore, many systems will include both human(s) and autonomy as part of a symbiotic team.

The level of human vs. autonomy involvement necessary and/or permissible to perform a particular task depends on the nature of the task and any performance goals. In many cases, the strengths of humans and autonomous systems are complementary. In order to select optimal roles, it is important to first consider the individual strengths of human and autonomy as sensors and controllers. Humans are highly flexible, operate well in dynamic environments, and are capable of complex social interaction and moral judgement. Autonomous systems are generally capable of higher computational performance, produce repeatable results, can handle multiple simultaneous tasks, and are not prone to fatigue, stress, or boredom. The 'Fitts list' [22] is a compilation of these general strengths/weaknesses and has traditionally been used to inform (static) function allocation [23] between humans and machines.

However, it is critical to acknowledge the dynamic characteristics of the human operator and the coupled interplay between agents. These factors imply a need for dynamic adjustment between human and machine roles to maintain optimal performance, yet it is not straightforward to decide what to adjust and when to adjust it. Dynamic function allocation [24] frameworks have been proposed to adaptively allocate subtasks or adjust levels of automation involved in these subtasks. The recent paradigm of "coactive design" [25, 26] goes beyond task decomposition and allocation, reasoning that human-machine frameworks should be designed according to mutual interdependence relationships. Understanding how the agents' actions affect one another can help the designer predict the impact of changes to the system. The authors of [26] stress the importance of communicating status/needs, ensuring agent actions are predictable, and ensuring agents can be directed by their counterparts. When neither the human nor the autonomy perform a subtask reliably, flexibility is designed into the system so that both can attempt the subtask.

III. Human-Autonomy Failure and Recovery

Two key factors in gauging the level of human and/or autonomous involvement are (i) when and (ii) how these systems fail. In addition to the potential failures of humans and autonomy in isolation, new failure modes may arise from their (improper) integration. It can be difficult to predict system behavior—and therefore, failure behavior—*a priori* because of the tight coupling between the human and autonomy [27]. In this section, we describe examples of HIL and HOL failure, finding common themes in response time and communication issues.

A. HIL vs. HOL Failure

When the human is an active participant in a shared or switched control loop, the probability of human error causing system failure is relatively high. For instance, the safety features of current high-end motor vehicles such as adaptive cruise control and automatic braking may help avoid accidents [28], but the human is responsible for the overall goal of reaching the destination, and may fail for a variety of reasons.

In application areas with high-bandwidth dynamics and/or high control order [29], the use of hybrid human-autonomy control introduces the possibility of instability in the feedback loop. Instability may occur when communication delay and/or operator reaction time degrade feedback to a plant with high-speed dynamics. It may be difficult to attribute a precise cause in these cases, as it is often a combination of human error and human-machine coupling behavior. For example, the Air France Flight 447 crash was caused by lost situational awareness due to failed sensors and instrumentation [30]. After losing awareness, the pilots panicked and continued to react improperly. Similarly, during testing of the prototype F-22 fighter aircraft, the pilot was not aware of transitions to a high-gain flight mode when landing gear were retracted near ground level [31]. The test pilot's overreaction caused pilot-induced oscillation (PIO) that resulted in a crash.

In contrast to in-the-loop failure, on-the-loop failure occurs when the autonomy fails and performance cannot be recovered by the human supervisor. The causes of autonomous failures are different than typical causes of human failures (fatigue, workload, etc.) and are therefore difficult for the human operator to understand or predict. Short-term loss of situation understanding and long-term loss of expertise are prominent issues in HOL frameworks. As detailed in [8], recovery from failure may suffer because more time is required for the human to gain situational awareness and uncover the root cause of the failure.

Transparency between human and autonomous agents is a major issue for HIL and HOL systems. Accidents like the shooting down of Iran Air Flight 655 by the USS Vincennes [32] were in major part caused by the inability to efficiently communicate contact information, leading to the misidentification of a commercial jet as threat. The growing employment of neural networks in autonomous applications, which are often integrated as "black box" systems, may be difficult for humans to interpret. Obscurity can lead to unnecessarily low operator trust when failure occurs, and may cause the operator to disable the autonomy even if it is beneficial overall. In general, interfaces can be made more transparent to the human operator by conveying confidence in the information presented. This transparency allows the human to make better assessments of when the autonomous control is operating correctly and when human intervention is necessary. Similarly, the interface can convey a notion of expected consequence so that the human operator is aware of potential risks when particular actions are taken [33].

B. Failure and Resilience in Human-Cyber-Physical Systems

From the perspective of resilience literature, failures of networked systems can be split into three categories: cyber failures, physical failures, and human error. Tight coupling of human, cyber, and physical elements creates redundancy that can increase resilience but also introduces unique disadvantages that need to be addressed. We describe some of these issues with regard to a semi-autonomous driving example.

Cyber: A cyber-attack may corrupt the sensors of a semi-autonomous vehicle, as well as the diagnostic systems of the autonomous controller. In severe cases, the controller may induce and/or permit a crash. An alert human on the loop has a short window of time in which to recognize the potential failure and switch to manual control to avert a collision. However, there is evidence that the human quickly disengages with the task when the autonomy's routine performance is high [34, 35].

Physical: In the event of a physical failure (and in the absence of cyber failure), the autonomous controller may recognize the issue and safely stop before a collision takes place; however, there may be a lag between the failure and this recognition, and human intervention may be necessary to avoid a collision.

Human: Finally, human error typically occurs when the human misdiagnoses a potential issue and switches in to take control of the vehicle when it is not needed. In this case, the performance will often degrade and there is a slightly elevated risk of collision, as was recently observed to be the case by the auto insurance industry for lane departure warning systems [36]. However, it is assumed that the human is trained well enough to operate the vehicle that this increased risk is small.

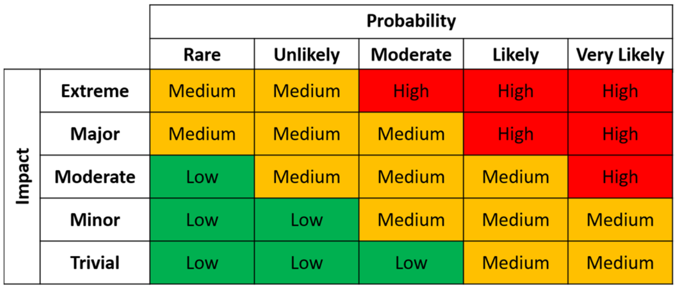
IV. Impact and Consequence

When a system fails and nothing happens, is it truly a failure? Failure implies that there is an unintended outcome with negative consequence. Consequence is inherently subjective, and in many cases inherently uncertain. However, for cases involving autonomy, that impact of the consequence is borne by humans. While humans are often poor at estimating consequence [37], machines cannot judge consequence without bounded pre-programming (e.g. using a utility function). Learning techniques are not practical as machines do not understand the "why" of consequence (e.g. "why is collateral damage bad") and therefore must be validated. In this way, evaluating consequence requires imposing ethics: "what actions are right or wrong in particular circumstances." These decisions are beyond what autonomy technologies are capable of making, and more importantly, we, as a society, are not willing to allow an autonomy to make.

V. Discussion: Methods for Assessing Level of Involvement

A. *Risk Management*

In most applications, some level of risk is unavoidable. Risk has two primary dimensions: (i) "impact" (i.e. consequence) of possible events and (ii) the probability of such events occurring. Fig. 1 displays a general probability-impact matrix [38], denoting the risk associated with various levels of these two variables. For instance, if the probability of a major-impact event is "likely", then there is high risk associated with that event. It is the responsibility of the designer to minimize the probability of high-impact events. Most safety mitigation strategies try to push the composite probability-impact score down and to the left. This approach holds for managing the risk associated with autonomy. However, with regard to applications potentially involving both humans and autonomy, it may be useful to consider the relative probability of human error vs. autonomy error; we discuss this in the following section.



		Probability				
		Rare	Unlikely	Moderate	Likely	Very Likely
Impact	Extreme	Medium	Medium	High	High	High
	Major	Medium	Medium	Medium	High	High
	Moderate	Low	Medium	Medium	Medium	High
	Minor	Low	Low	Medium	Medium	Medium
	Trivial	Low	Low	Low	Medium	Medium

Fig. 1. Probability-impact matrix typically employed in safety protocols.

B. *Guidelines for Human and/or Autonomy Control*

Taking into account the unique properties of humans and autonomy, as well as effects of their coupling, we can begin to develop guidelines for human autonomy collaboration. Fig. 2 illustrates a representative sequence of questions/answers that can be used to inform the use of human-only, autonomy-only, or human-autonomy joint control.

First, any decision to employ autonomy must insist that the autonomy be built to the highest professional and ethical standards [39]. Assuming this is true, one must assess the overall benefits of using autonomy; if the autonomy is not able to perform a useful function such as increase productivity, decrease cost, and/or reduce workload then the application may as well be performed solely by the human.

If autonomy does improve routine performance, the next step is to examine consequence/impact. If system failure or poor decision-making is guaranteed to have low impact (e.g. no potential for loss of life, property, etc.), then autonomy should be worthy of a role. Alternatively, if high-consequence events are possible, one should first examine the nature of these consequences. It is important to distinguish between applications where harm can come to humans and those where it cannot, given ethical standards and the need for accountability. An autonomy cannot readily be held responsible for (or punished for) its actions; therefore, it is critical to minimize the number of "ethical dilemmas" an autonomy must solve. If, in the event of an impending failure or ethical dilemma, transitions from autonomous control to human control can be performed in time to prevent catastrophic failure and allow the human to make an informed decision, then it is recommended that autonomy be part of the control loop.

Unfortunately, there exist applications where (i) the time scale for successful intervention is so short that the human cannot possibly regain timely control, or (ii) the human may have become complacent and require excessive time to regain situational awareness [6, 40]. In these cases, the designer must focus on the benefits of autonomy during this high-consequence event, rather than the benefits to routine performance. If the probability of a high-consequence event is reduced because of the use of autonomy, then it may be permissible to employ autonomy. The example of self-driving vehicles is apt: if autonomous driving can be proven to reduce accidents/deaths (perhaps by an order of magnitude or more), then the required autonomous decision-making may gain social acceptance.

If it can be shown that the chance of a major accident is significantly lower due to autonomy, then the final question is: Is the human a necessary/useful component of the control loop? If the human serves no additional function (or perhaps inhibits optimal performance), then generally they should not be part of the control loop. The level of utility that the human can provide at different stages in the decision-making process [8] will help dictate whether the human should be placed "in" or "on" the control loop.

Currently there are no accepted thresholds that define metrics such as "significant" improvement over human-only performance, a "high-impact" event, or the significantly high probability of such an event. Nevertheless, it is possible to outline a strategy for controller selection by adapting the probability-impact matrix described above to consider the relative probability of human error vs. autonomy error. Fig. 3 depicts such a matrix, partitioning the area between CA, HOL, HIL, and H control frameworks. This example would allow complete autonomy control (CA) over all tasks with trivial consequences, but only consider CA in higher-impact categories if the probability of human error was significantly higher than the probability of autonomy error. On the other hand, the human should be given complete control of high-impact tasks if the human errs much less than autonomy. In intermediate zones (moderate impact, closer relative error rate), a combination of human and autonomy may be ideal, especially if the systems are complementary. Within this range, HOL should be adopted with lower-impact, lower autonomy-to-human error rates, and HIL would be preferred with higher-impact, higher autonomy-to-human error rates.

We show in Fig. 3, how an impact-probability matrix can be utilized to determine under what conditions each agent, or combinations of agents should be given authority. This framework is agnostic to the specific failure mode; it only assumes that the failure can be classified with an impact and probability. Note that the decision tree shown in Fig. 2 represents discrete decisions, though we acknowledge that it is unlikely that these decisions will be discrete. Understanding how to arbitrate between human and autonomy decisions, or

HIL or HOL, can be informed by a thorough understanding of the impact-probability score across all combinations of agents.

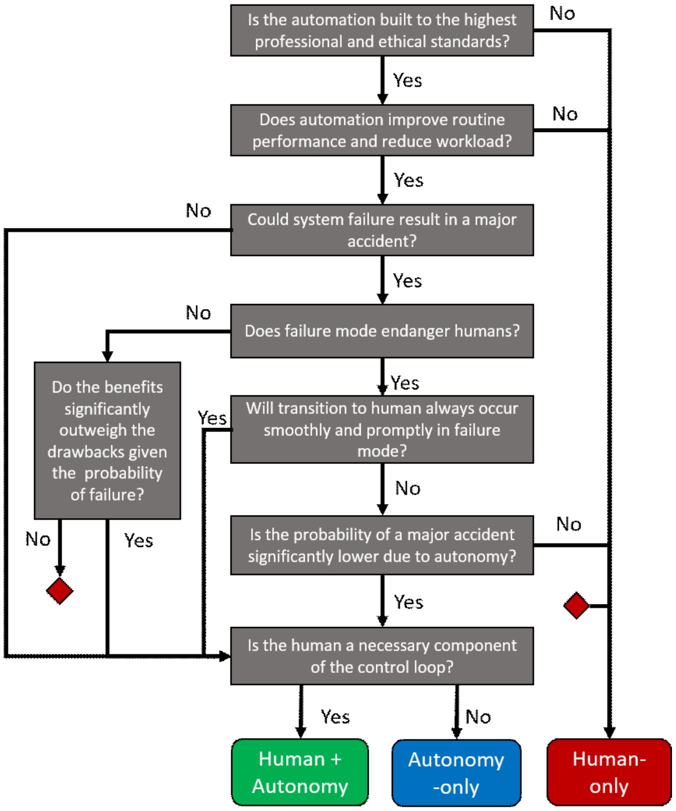

Is the automation built to the highest professional and ethical standards?
No
Yes
Does automation improve routine performance and reduce workload?
No
Yes
No
Could system failure result in a major accident?
Yes
No
Does failure mode endanger humans?
Yes
Do the benefits significantly outweigh the drawbacks given the probability of failure?
Yes
Will transition to human always occur smoothly and promptly in failure mode?
No
No
Is the probability of a major accident significantly lower due to autonomy?
No
Yes
Yes
Is the human a necessary component of the control loop?
Yes
No
Human + Autonomy
Autonomy -only
Human- only

Fig. 2. Decision tree recommending controller type (human, autonomy, or joint control) based on application qualities.

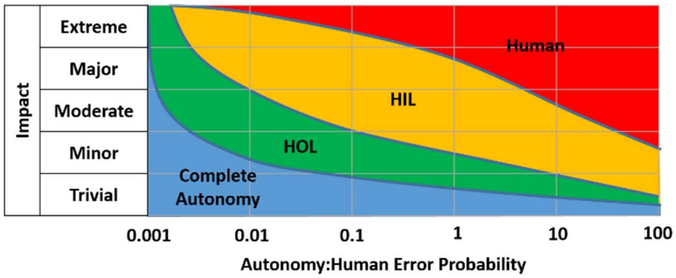


Fig. 3. Adapted probability-impact matrix considering the ratio between autonomous error and human error. Extreme impact (high-consequence) events are biased toward human operation, even if the human is more likely to fail, because of ethical concerns related to autonomous decision-making.

C. Passivity

Many systems will be designed to have varying levels of human-autonomy integration depending on circumstances. Design with respect to the concept of passivity provides a straightforward methodology for assessing stability in a (potentially modular) system with one or more humans and one or more autonomous agents (or autonomous modes within the same agent) that switch between exclusive control of a plant. Passivity of the entire system is guaranteed when each subsystem is passive and the switching between human/machine controllers does not violate the notion of passivity, and stability follows from this passivity guarantee. This type of design is most applicable to systems where the human is "in the loop" performing a tracking/regulation task. In our prior work, we demonstrated stable switching between a simulated passive human and passive autonomy on an Euler-Lagrangian dynamic system [41]. Note that in some applications, ethical questions of autonomous decision-making coupled with the inability to relinquish control to the human with sufficient time to make informed control decisions may render the human unable to intervene appropriately, and as such these needs must be taken into consideration when designing the system.

D. Regulatory and Societal Burdens

It is our conclusion that the technological issues facing ubiquitous autonomy adoption are imminently solvable, and solvable in the near term. We also believe that for the foreseeable future, the human will retain at least some role in that decision loop. So, what is inhibiting adoption of currently capable technologies today? While autonomy technology will create some advantages to individuals, we will see the largest impacts at the societal level. However, autonomy technology will occasionally fail; those failures will have consequences; and, those consequences will be felt most acutely at the individual level. This is the societal good vs individual risk paradox colloquially known as NIMBY (not in my back yard). This is a monumental hurdle, especially in today's litigious society that has inhibited many companies/organizations from pursuing autonomous technologies [42]. The solutions perhaps can be found by examining another controversial technology that has now found near-ubiquitous adoption. In 1988 the US Congress passed the National Childhood Vaccine Injury Act established a framework with clear methods for arbitrating between the rights of consumers and companies [43]. This framework has enabled companies to bring products to market that provide significant societal benefits despite initial risk or uncertainty regarding individual outcomes. A similar approach, as recently proposed by Congress [42] that limits corporate risk while also protecting consumers for autonomy technologies may see many more companies willing to enter the market with much more impactful technologies.

E. Future of Autonomy and Humans

In many cases, as autonomy increases in capability (makes fewer errors), humans will be moved from in the loop to on the loop, (moving down and to the left on the impact-probability matrix) and potentially out of the loop. Semi-automated driving [34, 44] will eventually give way to fully self-driving cars where the human is completely out of the loop (but not out of danger). However, in other cases, as autonomous platforms become more capable, the need for human interaction often *increases* because the autonomy's new responsibilities are part of an interdependent, "team" relationship [39]. Autonomous systems can only responsibly fill these roles if their social capability is adequate.

Many systems will contain coupled and/or cascading control loops that can be characterized as HIL, HOL, or both. For instance, the transportation network company Uber employs (semi-)autonomous algorithms to produce recommendations and incentives for contracted drivers to work at a particular time or in a particular area [45]. The choices/actions of the drivers provide feedback to the autonomy. Uber drivers may be considered "in-the-loop" while Uber employees may be "on-the-loop" supervising both the autonomy and the drivers. Another example are military decision support systems [46], in which the human typically supervises during information collection and integration but is an active participant during decision-making. A recently-developed framework for human-machine image triage comprises human and computer-vision-based target detectors working in parallel; the individual "scores" can be dynamically weighted based on the confidence in individual sensor modalities, and high-consequence detections may be passed to other humans "on the loop" for further analysis and decision-making [47].

There is a special case of when we allow the autonomy to take control from a human in based on an evaluation of human mental/physiological state. As a society we are becoming more comfortable with an autonomy taking control in dire circumstances, such as anti-collision systems, anti-lock braking, and breathalyzers to prevent the starting of a car. In the near term, as autonomy becomes more ubiquitous, we will need to address under what conditions we will allow autonomy to over-rule our decisions. Many banks have systems that warn you when you have violated your budget criteria. Should we empower autonomy to keep us frugal? Doctors' offices will send automated reminders about upcoming appointments we have made or should make. Should we empower autonomy to keep us well?

Conclusions

Despite the growth of autonomy in many modern industries/fields, legal and ethical concerns ensure that humans will occupy a critical role in many systems for the foreseeable future. It is therefore in our best interest to capitalize on the abilities of both agents. Solutions will vary based on task type, failure modes/consequences, and autonomy-to-human error ratio.

We conclude with a prescription for deploying autonomy: If (i) the action-consequence relationship can be accurately encoded in an impact-probability matrix (by a human professional, based on social and ethical guidelines), (ii) system, environmental, and human states can be estimated reliably (by autonomous sensors), and (iii) the probability of negative consequences (e.g. potential for human injury or death) satisfies acceptable bounds, then it is acceptable to deploy autonomy.

References

[1] D. E. Nye, America's Assembly Line, MIT Press, 2013.

[2] F. W. Meredith, "The modern autopilot," Aeronautical Journal, 1949.

[3] A. Wolman, "Industrial water supply from processed sewage treatment plant effluent at Baltimore, Md." Sewage Works Journal, 20(1), 1948, pp. 15-21.

[4] Y. Harrison and J. A. Horne, "The impact of sleep deprivation on decision making: a review," Journal of Experimental Psychology, 6(3), 2000, pp.236-249.

[5] B. Donmez, C. Neheme, and M. L. Cummings, "Modeling workload impact in multiple unnmanned vehicle supervisory control," IEEE Transactions on Systems, Man, and Cybernetics, 40(6), 2010, pp. 1180-1190.

[6] R. Parasuraman and D. Manzey, "Complacency and bias in human use of automation: an attentional integration," Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 2010, pp. 381-410.

[7] P. Scharre and M. Horowitz, "An introduction to autonomy in weapon systems," Center for New American Security, 2015.

[8] L. Onnasch, C. D. Wickens, H. Li, and D. Manzey, "Human performance consequences of stages and levels of automation: an integrated meta-analysis," Human Factors: The Journal of the Human Factors and Ergonomics Society, 56(3), May 2014, pp. 476-488.

[9] M. Pilling, "Issues regarding the future application of autonomous systems to command and control (C2)," Defense Science and Technology Organisation, Australian Department of Defense, DSTO-TR-3112, 2015.

[10] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, and W. C. Meyers, "Robotic surgery, a current perspective," Annals of surgery, 239(1), 2004, pp. 14–21.

[11] M. A. Talamini, S. Chapman, S. Horgan, and W. S. Melvin, "A prospective analysis of 211 robotic-assisted surgical procedures," Surg Endosc, 17, 2003, pp. 1521-1524.

[12] A. M. Okamura, "Methods for haptic feedback in teleoperated robot-assisted surgery," Industrial Robot, 31(6), 2004, pp. 499-508.

[13] H. Ding, J. Heyn, B. Matthias, and H. Staab, "Structured collaborative behavior of industrial robots in mixed human-robot environments," IEEE International Conference on Automation Science, 2013, pp. 1101-1106.

[14] A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: metrics and control," IEEE Trans. on Automation Science and Engineeriing, 13(2), April 2016, pp. 882-893.

[15] K. Wagner, "Facebook's virtual assistant 'M' is super smart. It's also probably a human," Recode, Nov. 3, 2015, http://www.recode.net/2015/11/3/11620286/facebooks-virtual-assistant-m-is-super-smart-its-also-probably-a-human.

[16] DARPA, "Collaborative Operations in Denied Environment (CODE)," http://www.darpa.mil/program/collaborative-operations-in-denied-environment, accessed July 4, 2016.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," IEEE International Conference on Computer Vision, Feb. 2015, pp. 1026–1034.

[18] S. Ioffe, C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," International Conference on Machine Learning, 2015, pp. 448–456.

[19] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves et al., "Human-level control through deep reinforcement learning," Nature, 518(7540), 2015, pp. 529-533.

[20] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser et al., "Mastering the game of Go with deep neural networks and tree search," Nature 529(7587), 2016, pp. 484-489.

[21] C. Moyer, "How Google's AlphaGo beat a Go world champion," The Atlantic, March 28, 2016, http://www.theatlantic.com/technology/archive/2016/03/theinvisibleopponent/475611/

[22] P. M. Fitts, "Human Engineering for an effective air-navigation and traffic-control system," Washington, DC: National Research Council, 1951.

[23] B. Kantowitz and R. Sorkin, "Allocation of functions," Handbook of Human Factors, New York: Wiley, pp. 365-369.

[24] J.-M. Hoc, "Towards a cognitive approach to human-machine cooperation in dynamic situations," Int. J. Human-Computer Studies, 54, 2001, pp. 509-540.

[25] J. M. Bradshaw, V. Dignum, C. Jonker, and M. Sierhuis, "Human-agent-robot teamwork," IEEE Intelligent Systems, 27(2), 2012, pp. 8-13.

[26] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. van Riemsdijk, and M. Sierhuis, “Coactive design: designing support for interdependence in joint activity,” Journal of Human-Robot Interaction, 3(1), 2014, pp. 43-69.

[27] R. M. Robinson, D. Scobee, S. A. Burden, and S. S. Sastry, “Dynamic inverse models in human-cyber-physical systems,” SPIE Conference on Defense and Commercial Sensing, April 2016.

[28] M. Korber, W. Schneider, and M. Zimmerman, “Vigilance, boredom proneness and detection time of a malfunction in partially automated driving,” IEEE Collaboration Technologies and Systems, 2015, pp. 70–76.

[29] R. J. Jagacinski, J. M. Flach, Control Theory for Humans: Quantiitative Approaches to Modeling Performance, CRC Press, 2003.

[30] J. Wise, “What really happened aboard Air France 447,” Popular Mechanics, Dec. 6, 2011, http://www.popularmechanics.com/ flight/a3115/ what-really-happened-aboard-air-france-447-6611877/.

[31] J. Harris and G. T. Black, “F-22 control law development and flying qualities,” AIAA Atmospheric Flight Mechanics Conference, 1996, pp. 155–168.

[32] W. Fogarty, “Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988,” DoD Report, 53 pp., August 1988.

[33] R. M. Robinson, M. J. McCourt, W. D. Nothwang, and J. W. Curtis, “Levels and stages of automation in decision support systems for command and control,” IEEE International Conference on Systems, Man, and Cybernetics, submitted April 2016.

[34] N. Strand, J. Nilsson, I. C. M. Karlsson, and L. Nilsson, “Semi-autonomous versus highly automated driving in critical situations by automation failures,” Transportation Research Part F, 27, 2014, pp. 218-228.

[35] K. Kircher, A. Larsson, and J. A. Hultgren, “Tactical driving behavior with different levels of automation,” IEEE Trans. on Intelligent Transportation Systems, 15(1), Feb. 2014, pp. 158–167.

[36] M. Moore and D. Zuby, “Collision avoidance features: initial results,” Proc. Int. Conf. on the Enhanced Safety of Vehicles 13(0126), 2013.

[37] A. R. Marathe, B. J. Lance, K. McDowell, W. D. Nothwang, and J. S. Metcalfe, “Confidence metrics improve human-autonomy integration.” Proc. ACM/IEEE International Conference on Human-Robot Interaction, 2014, pp. 240-241.

[38] V. Dumbrava and V.-S. Iacob, “Using probability-impact matrix in anaylsis and risk assessment projects,” Journal of Knowledge Management, Economics, and Information Technology, Special Issue, Dec. 2013, pp. 76-96.

[39] R. R. Murphy and D. D. Woods, “Beyond Asimov: the three laws of responsible robotics,” IEEE Intelligent Systems, July 2009, pp. 2-8.

[40] R. Parasuraman, K. A. Cosenzo, and E. De Visser, “Adaptive automation for human supervision of multiple uninhabited vehicles: effects of change detection, situation awareness, and mental workload,” Military Psychology, 21, 2009, pp. 270-297.

[41] M. J. McCourt, R. M. Robinson, W. D. Nothwang, E. A. Doucette, and J. W. Curtis, “Passive switched system analysis of semi-autonomous systems,” IEEE International Conference on Systems, Man, and Cybernetics, submitted April 2016.

[42] http://www.wsj.com/articles/obama-administration-proposes-spending-4-billion-on-driverless-car-guidelines-1452798787 Jan 2016.

[43] R. F. Edlich, D. M. Olson, B. M. Olson, et al. (2007). "Update on the National Vaccine Injury Compensation Program". J Emerg Med 33(2), pp. 199–211. *doi:10.1016/j.jemermed.2007.01.001. PMID 17692778*.

[44] Tesla, “Model S Software Version 7.0,” Tesla Motors, accessed May 16, 2016, https://www.teslamotors.com/presskit/autopilot.

[45] A. Rosenblat and L. Stark, “Uber’s drivers: information asymmetries and control in dynamic work,” Data & Society Research Institute, October 2015, 17 pp.

[46] R. M. Robinson, H. T. Lee, M. J. McCourt, A. R. Marathe, C. Ton, and W. D. Nothwang, “Human-autonomy sensor fusion for rapid object detection,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sept. 2015, pp. 205-312.

[47] G. M. Gremillion, J. S. Metcalfe, A. R. Marathe, V. Paul, J. C. Christensen, K. Drnec, B. Haynes, and C. Atwater, “Analysis of trust in autonomy for convoy operations,” SPIE Conference on Defense and Commercial Sensing, April 2016.