Trusted Autonomy Between Humans and Robots

Toward Human-on-the-Loop in Robotics and Autonomous Systems

by Saeid Nahavandi

Systems that can change their behavior in response to unexpected conditions and events during operation are known as *autonomous* [1]. Autonomy refers to the capability of a machine to perform a task, or part of it, with no—or substantially reduced—human intervention. Over the years, autonomous systems have appeared and sometimes dominated various aspects of human daily activities, such as in robot-controlled operations.

The levels of autonomy range from teleoperation to fully autonomous systems [2]. Autonomy can be categorized into two broad classes: human-in-the-loop (HITL) and human-on-the-loop (HOTL). Machines that carry out a task for a time period, then stop and wait for human commands before continuing are known as *HITL systems*, while machines that can execute a task completely and independently but have a human in a monitoring or supervisory role, with the ability to interfere if the machine fails, are known as *HOTL systems* [3]. HOTL systems can also be fully autonomous if human supervisors allow them to perform a function entirely on their own.

Safety, accuracy, and security have always been major concerns pertaining to the adoption of autonomous systems. HOTL systems, in particular, require a high level of trust to be accepted in our daily lives. Trust translates as a human operator's willingness to rely on the actions performed by an autonomous system [4]. Trust depends on the specific system and circumstances as well as on the environment the system is working in and the operator's background and beliefs [5]. Achieving the trust that can lead to realizing particular objectives individually and accurately in uncertain conditions over a long period of time without human involvement has always been the main goal with autonomous systems.

Digital Object Identifier 10.1109/MSMC.2016.2623867
Date of publication: 17 January 2017

2333-942X/17©2017IEEE

Robots have become popular in many industries, and it is anticipated that robotics technologies will dominate most aspects of our lives in the coming decades [6], [7]. The field of robotics and autonomous systems (RAS) constitutes one of the emerging breakthrough areas in science and technology in the 21st century, enabling innovations for our businesses and society. Humans have benefited and will continue to benefit from RAS technologies capable of carrying out so-called *4D tasks* (dangerous, dirty, difficult, and dull) in a variety of sectors. The RAS field is assuming a progressively more important role in such diverse areas as addressing national defense and security challenges, enhancing our health-care system, assisting with our aging population, making our roads safe, and enabling greater productivity in manufacturing. Examples include working in the ocean depths, in nuclear power plants, at repetitive manufacturing operations, and on delicate surgeries in remote locations. In addition, RAS applications can minimize soldiers' risks by removing them from the battlefield, collecting information about the battle environment, and traveling between waypoints without human assistance. In general, the RAS area consists of interconnected and interactive systems that can perceive their surroundings, reason about events, make or modify plans, and control their activities.

A report by the McKinsey Global Institute [8] has identified advanced robotics and autonomous and near-autonomous vehicles as two of a dozen emerging disruptive technologies that could have a transformational impact on our lives, business activities, and the global economy. In this regard, it is estimated that advanced robotics alone will generate annual economic activity of US$1.9 trillion to US$6.4 trillion by 2025 [8], [9]. For safe, reliable, and effective RAS deployment, close collaboration and communication between humans and robots is necessary. Human–robot interaction has become vital, as it allows operators to quickly comprehend the state of the system under control and efficiently supervise its activities to achieve a new, preferred state.

While RAS technologies have started to emerge from research laboratories into industry and society, comprehensive efforts are required to demonstrate, test, and de-risk these technologies to make them trustworthy. How this can be achieved and realized needs to be carefully considered and investigated by subject-matter experts.

Different Forms of Autonomy

Owing to significant technological advances, RAS has already become part of our lives and will continue to do so. Different forms of autonomy have been used in RAS in recent decades. As stated previously, the two main forms of autonomy in RAS are HITL and HOTL. HITL systems or semiautonomous RAS are robots that perform a task autonomously for a time period, then halt and wait for a human operator's commands before continuing [3]. For example, HITL autonomous weapons use autonomy to search for, detect, and evaluate threats and then select and engage separate targets under human control (that is, humans decide the targets to be selected and engaged). The Raytheon Patriot antimissile system is an HITL system. The Patriot system can select a target according to human-defined rules, but will not engage the target until receiving confirmation from the human operator [10].

HITL is increasingly being replaced by HOTL in RAS. HOTL or human-supervised autonomous robots can perform a function completely without human help, but with a human in a monitoring or supervisory role who has the ability to interfere with and override the robot's decision if the robot should fail or if there is any error [3]. HOTL RAS can be also fully autonomous if the human supervisor allows them to carry out a task completely on their own. Keeping the human on the loop adds a needed human–robot interaction and human–machine interface. The degree of autonomy is determined based on the RAS's relationship to the human supervisor. HOTL RAS receive, evaluate, decide, and begin execution of an operation, but the human supervisor can veto or stop it when necessary. For example, HOTL RAS weapons are systems that use autonomy to choose and engage targets. While the human supervisor does not decide the selected targets to be engaged, the human can monitor the RAS weapon system's intention and performance and can interfere to stop its operations if required.

Currently, HOTL weapon systems are increasingly being used for defense applications, which include air and missile defense systems. The Phalanx is a defensive, close-in naval weapon system (a fast-fire, computer-controlled, radar-guided gun system) created to shoot down antiship missiles and surface threats. Once activated, it searches, detects, and evaluates threats and then tracks and engages the threat. An abort button is available for the human supervisor to reject the system's decision. The main question is whether these HOTL weapon systems are trustworthy. That is, can they distinguish blue and red teams on the battlefield? And are they ethical? To have a usable and useful HOTL RAS, trusted autonomy becomes essential.

Trusted Autonomy

Trust is a firm belief in the reliability, truth, or ability of someone or something. An autonomous system requires the reliability of and trust in its technology. With RAS, trust is defined as the level of confidence a human has in an autonomous system based on the person's observations, perceptions, and expectations of the system's performance and on other information regarded as evidence of competence [11]. Trust in HOTL RAS is also defined as the ability of HOTL RAS to successfully perform an activity, at a specific time, and under conditions characterized by vulnerability and uncertainty [12].

If the actions of HOTL RAS lead to harmful consequences to humans or belongings (e.g., from an unmanned aircraft or a driverless vehicle), the human supervisor needs to reestablish and maintain trust in the operations of HOTL RAS [13]. Humans need to be confident that HOTL RAS will perceive conditions properly in all situations, make the right decisions, and perform its tasks accurately and efficiently. For building trust in HOTL RAS, the systems have to be certified. Certification is a formal means by which a regulator confirms the expected efficiency and performance of various components of an autonomous system.

A general HOTL RAS includes a number of essential units [14]: 1) a sensing and perception unit, which involves the abilities to sense, interpret, detect, and evaluate objects in different environments; 2) a control and decision-making unit, which involves the ability to make accurate decisions in an uncertain and unpredictable environment; and 3) an execution unit, which involves the ability to perform tasks provided by the control and decision-making unit. Intelligent systems play the main roles in the control and decision-making unit, which include learning, adaptation, and cognition. Learning is the acquisition of knowledge, skills, or abilities through experience, as observed by the attainment of increasing success (enriched behavior). Adaptation is a change and modification in behavior when the environment is changed. Cognition includes learning development, adaptation, and natural interaction through intelligent behavior in response to complex objectives in a complex environment. Cognitive mechanisms are required for decision making in RAS. John Boyd distilled the decision-making process into an observe, orient, decide, act scenario, which is known as the *OODA loop* [15]–[17]. Boyd realized that it is necessary for military pilots to make decisions faster and more accurately than their opponents, and he used the OODA loop concept for the combat and military operations process [18]. The structure of the OODA loop is shown in Figure 1.

Trusted autonomy is the greatest technical barrier that needs to be overcome in HOTL RAS. As technology evolves, HOTL RAS can become intelligent in learning, adaptation, and decision making without direct human engagement. To measure trust, a number of quantitative factors related to system behavior must be checked: 1) performance, which includes competence, accuracy, reliability, and robustness; 2) transparency of the control and decision-making unit; and 3) security vulnerabilities [19]. To quantify the degree of trust, these factors need to be measured in certain and uncertain environments. In addition, efficient human–robot interaction positively affects humans' trust in HOTL RAS. Trust can be established through an efficient way of communicating and disseminating information. The human supervisor needs correct information from the system to establish suitable reliance and prevent system misuse and malfunction. Understanding the reasons for system failures can increase trust. In addition, providing essential information and making the autonomous system a team player inevitably increase trust.

A Framework for Formulating Trusted Autonomy

It is argued that there are three main factors in fostering trusted autonomy between humans and RAS: 1) the humans who are interacting with RAS; 2) RAS; and 3) the environment in which RAS are intended to function, as shown in Figure 2.

The Humans

Since HOTL RAS are designed to be managed and supervised by humans aiming to benefit from them, humans and their desires are the most important factor. Different humans have different viewpoints pertaining to RAS, according to their role. A human's perception is based on

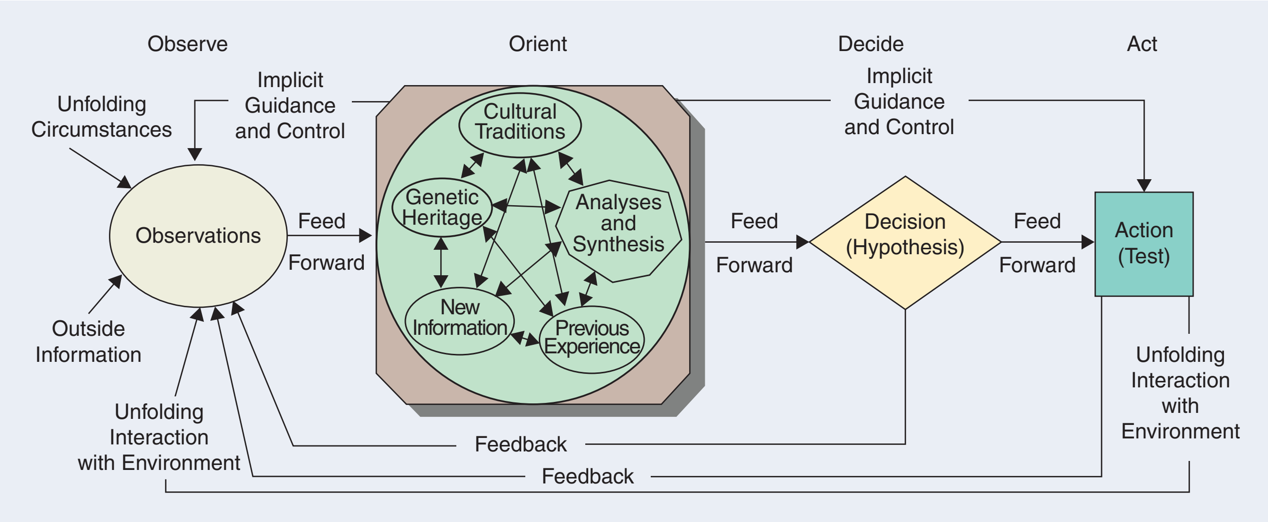

Figure 1. The OODA loop structure [15].

personally observed or received evidence and is affected by personal experience and culture. Culture and background are multifaceted, such as ethical, religious, and professional affiliation and age. For example, the younger generation has more confidence in autonomous systems than the elderly because different new technologies such as smartphones and the Internet are second nature to young people. Therefore, cultural and demographic differences play a significant role in fostering trusted autonomy between humans and HOTL RAS [20]. Trust, in addition to being influenced by culture and demographics, can vary according to experience and circumstances [21]. In other words, the level of trust differs from one individual to the next.

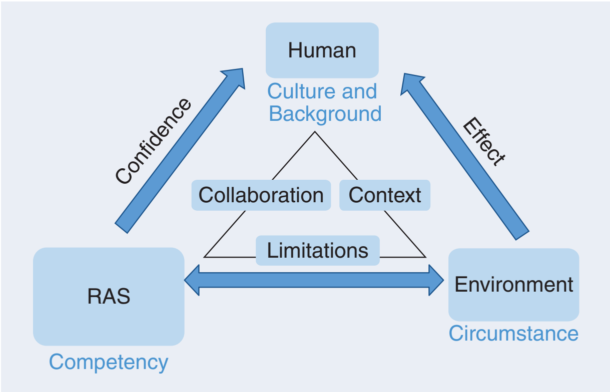

Figure 2. The framework for discussing trust in autonomy.

RAS

A certain level of competency, which leads to trustworthiness, is necessary for all designed RAS. The notion of competency should be precisely detected, measured, and evaluated as a feature of RAS. The competency depends on the RAS's structure and performance in different scenarios. The competency of RAS can be measured and evaluated based on the outcome of different tests. During the test phase, if known inputs lead to expected and satisfactory results, the human can be confident that RAS are functional as expected.

A person's confidence can be enhanced by checking not only what operation RAS should perform and when, but also the reason and justification for RAS acting so [13], [22]. Therefore, human–machine interaction (and collaboration) is vital for the human operator to understand the reason for each RAS operation in an attempt to build confidence with respect to the RAS's capability.

The Environment

The environment is an important influence, since RAS need to respond and act based on environmental circumstances. The sensor inputs form the RAS's perception of the environment. Therefore, perceiving the environment properly is a significant part of the RAS's ability to respond to it and make decisions accurately. The human supervisors might impose some constraints on the system's behavior as part of safety measures aiming to reduce any potential negative or harmful complications of RAS. If the degree of confidence increases, the applied constraints can be relaxed over time. But constraint relaxation needs to be matched with the RAS's verified trustworthiness.

Since ensuring that RAS can be trusted and remain competent is a major concern today, HOTL RAS need new approaches to establish what we might call *certifiable trust*. For establishing certifiable trust and enabling the capabilities of HOTL RAS, a number of factors and challenges need to be addressed.

How to Establish Trusted Autonomy in HOTL RAS

While advanced technologies are emerging, it is still a great challenge to enable HOTL RAS to sense and understand the surrounding environment, comprehend and perceive objectives, make informed decisions, and complete a mission successfully without hurting people. Advances in machine intelligence with respect to learning, adaptation, decision making, and perception contribute toward establishing trusted autonomy in HOTL RAS. As such, a variety of technologies must be developed, tested, enhanced, or even invented before the potential of HOTL RAS can be fully realized. While the applications and benefits of HOTL RAS are endless, proliferation of these systems in our lives creates real concerns and significant risks, especially in safety-critical tasks in areas such as warfare, health care, and transportation.

Better Haptic Sensors and Accurate Data Collection

Future RAS capabilities require more accurate three-dimensional (3-D) measurements of the environment for automatically detecting terrain or environmental properties to enable better movement and performance. Currently, state-of-the-art sensor technologies are relatively slow, large in size, and more power-demanding than desired, requiring humans to review, plan, and develop certain operations. Therefore, reducing the size, weight, and power consumption of 3-D sensing technology is also important. In addition, RAS sensors and data acquisition are highly susceptible to degradation under poor weather conditions, adding to untrustworthiness. Future 3-D sensing technology should consider improvements in a number of factors: 1) frame rate, 2) maximum range, and 3) spatial and range resolution aiming to increase the exploration range of the operating environment, ensure safe operation in hazardous terrain, and provide safe object manipulation when humans are closely involved. It is also essential to have accurate estimation of the relative position or velocity of RAS on a centimeter-per-second scale.

Furthermore, current touch sensors are delicate and expensive and tend to drift in response to temperature changes, as they are sensitive to environmental conditions. Therefore, haptic sensors that are small, low cost, and reliably high resolution and that have good functional bandwidth are essential to enable RAS to handle delicate objects as capably as humans. RAS also need to learn to ignore the haptic effects of their own motions so they can better identify the sensations caused by contacts with the outside world.

The availability of accurate data is another challenge in designing and developing trusted autonomy for HOTL RAS. Sensor data in RAS should be collected accurately, without losing any information during the process of producing the correct control output. The operation of the majority of RAS, such as driverless vehicles, depends on the Global Positioning System (GPS). This high dependency puts RAS at risk in situations where the GPS signals are lost or spoofed. RAS could hurtle in entirely the wrong direction when receiving counterfeit GPS signals. Gaining a better understanding of the broader implications of GPS spoofing is essential for safe operation of RAS.

Machine Intelligence, Learning, Reasoning, and Perception

A Better Intelligent System

To further realize trusted autonomy in RAS, real-time access to environmental information and comprehensive analysis must come together efficiently. Intelligent mapping algorithms that construct 3-D models of the environment using state estimates and sensor measurements are necessary for efficient RAS operation. A great demand in this domain is to generate geometric maps that aid in autonomous navigation and object manipulation. In addition, new technologies for static as well as dynamic object recognition are important to allow RAS to operate accurately.

There is currently a lack of intelligent algorithms able to perform complex activities in complex environments. Developing autonomous capabilities to operate effectively and precisely in unpredictable environments is still a huge challenge. Autonomous robots should be able to perform their tasks intelligently and adapt to different environments without direct human supervision for each task. In an unpredictable environment, RAS need to observe their environment, understand it, and adjust their performance accordingly. In situations when humans and RAS work in close proximity, it is important for RAS to recognize humans and their activities. In this case, for the sake of human safety it is imperative to ensure that RAS perception algorithms function accurately. In particular, the future autonomous robot has to carry out its tasks in a dynamic environment that is subject to changes and without a human dedicated to directly controlling the tasks.

Full autonomy in RAS can be very challenging because of ambiguity resulting from noisy sensors, the lack of safety guarantees, unpredictable environments, inefficient perception algorithms, and unmodeled operator intentions. Therefore, isolated decision-making systems are shifted to those that share control, with important autonomy devolved to RAS, leaving the human supervisor to monitor the decisions. Significant innovation is expected to occur in the future from sharing control between RAS and humans, with HITL stimulus and motion mapping enabling robots to learn how to predict and adapt according to real-world conditions and perform safely in the HOTL mode.

Better Decision Making

The chief difference between a human and an RAS is decision making. RAS decision making depends on the received information and the way the machine perceives it. Decision making in HOTL RAS should be performed even if the information is limited or ambiguous. The human brain is very adaptive and able to distinguish apparent decisions from circumstances that need more thought. However, decision-making algorithms are gradually being improved using more flexible and robust algorithms that can focus on the quality of decision making by reasoning in a way similar to the human brain. Nevertheless, the maturity of this technology is still a long way off. As computers and decision-making algorithms have become more advanced, it has become necessary that they be continually tested and assessed in unpredictable environments. Therefore, while humans have abilities such as manipulation, dexterity, decision making, and perception, current RAS technologies are very far away from this. These problems need to be addressed in further research related to HOTL RAS.

Nowadays, researchers are trying to give RAS the ability to make their own decisions in wide-ranging circumstances. While today's most advanced RAS can handle diverse situations, they still have some problems pertaining to ambiguity resulting from inaccurate sensors, inefficient learning and decision-making algorithms, lack of safety, and unpredictable environments. The learning and decision-making algorithms should be more developed and improved to make the future RAS more capable, reliable, independent, and accurate, with safety and efficiency in performance guaranteed. Therefore, through advances in artificial intelligence, RAS can acquire more active behaviors and plan their actions in complex, unpredictable, and unfamiliar environments. Human safety (the safest physical interaction) is the first priority, and efficiency is the second.

In the future, efforts should be dedicated toward producing autonomous robots that are lighter, more accurate in facing different environments, and more cost effective and that have better quality in terms of performance and safety as compared with those currently available. Future robots should be easily handled by users with different skill levels through designing a user-friendly interface to minimize programming and controlling costs. Some novel techniques and robust control algorithms for making correct and ethical decisions under conditions of uncertainty are essential.

Better Teaming, Human–Autonomous System Interaction, and Collaboration

Human–robot interaction has become more and more important in recent years because of the increasing number of complex RAS as well as our exposure to such technologies in our lives [23]–[25]. Human–robot interaction is a research area dealing with understanding, designing, and assessing robotic systems to be used by or with humans. The human–robot interaction challenge is to understand and create interactions between one or more robots and humans and to determine how these interactions can be influenced and improved to a certain level while safety, ease of communication, and precision and efficiency of performance are guaranteed. Human–robot interaction covers the classical domain of remote robotics and telerobotics (for example, haptics), human–system interfaces, and improved reality as well as new subjects that contain human–system integration, human safety, human–robot teams, and supervision with time delays in remote areas.

Research on RAS, intimate human collaboration in manipulation tasks, human control and behavior of humanoid and anthropomorphic robots for hazardous environments, and social interaction with RAS is at the initial stage at the moment, and there are many problems that need to be addressed in the future [23], [26], [27]. Advanced technologies in human–system interaction need to be explored for enhancing humans' situational awareness, developing RAS to receive a human's intent, and allowing RAS to perform safely. Well-developed human–machine interfaces can improve human–robot interaction capabilities, making a system understandable and reducing human errors while operators are on the loop. Advanced human interfaces are able to communicate with clarity about their objectives, abilities, strategies, and accomplishments; cooperate to solve issues, particularly when the circumstances shift away from autonomous abilities; and communicate through multiple modalities (speech and gestures). Therefore, a more efficient way of communicating and disseminating information can enhance trust, as human supervisors need the appropriate information from RAS to build confidence and prevent system failures and malfunctions.

In human–robot teaming, RAS should be able to respond to a dynamic environment as fast as possible, decreasing the load of human teammates needed to explicitly manage and direct their activities. Other essential qualities for the HOTL RAS are dynamic adaptation, capability to self-organize and dynamically restructure, robustness to the addition and loss of agents, and agent-to-agent collaboration.

Supplementary information about the human teammates' current cognitive and emotional state needs to be delivered to an RAS teammate for making decisions that could enhance human–robot team performance. By including electroencephalogram, electrocardiogram, and pupil diameter measures (as shown in Figure 3), which could be utilized to evaluate the workload in near real time, advanced online frameworks can be developed to create a closed-loop system to enhance the RAS behavior.

In addition to the aforementioned challenges, a better transparency and trust-based bias in the human–machine context is essential. Calibrated trust could be obtained through an understanding of and confidence in the RAS's behavior and decision-making capabilities as well as transparency to enable humans to understand what RAS are doing and for what reasons. Developing a common understanding and shared perception is another challenge in HOTL RAS. For humans and robots to have shared understanding, perception, and situational awareness, information should be shared in common and transmittable in logical formats and time scales.

Better Evaluation

Developing HOTL RAS for some critical applications such as armed autonomous systems requires making accurate distinctions between combatants and civilians. This is particularly challenging as nowadays combatants often intentionally disguise themselves as civilians to avoid detection. Even if HOTL RAS had 99% accuracy, it would

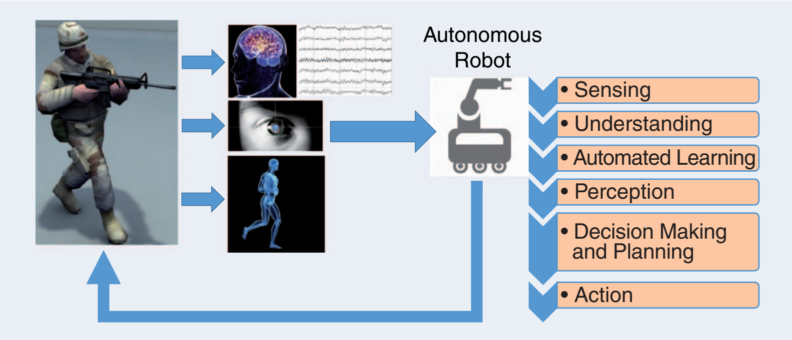

Figure 3. The teaming of future autonomous systems. [Images courtesy of cgtrader (soldier), Pintrest/caudate-nucleus (brain), and stock-clip.com (skeleton).]

still be imperative not to accidentally sacrifice even one human life. Therefore, it is not probable that fully autonomous RAS will be sent into the battlefield when civilian lives might be at risk, at least not until a satisfactory result could be provided in terms of absolutely trustworthy detection, evaluation, decision making, learning, and performance. New methodologies and metrics must be developed to evaluate distributive control and situational awareness. In addition, developing test protocols to support on-board management system evaluation and autonomous shutdown is also required for future HOTL RAS.

Furthermore, developing realistic test-bed components and environments to support HOTL RAS evaluation is necessary. In this respect, driving simulators (e.g., the Universal Motion Simulator [28], as shown in Figure 4) are safer and more cost-effective tools to evaluate new autonomous vehicle designs as well as human behavior in a virtual environment [31]–[33]. The degree of human trust in a designed autonomous vehicle can be measured through driving simulators before an actual test drive, as can the efficiency of human–robot interaction when the driver is supervising the driving scenarios.

Better Protection

The majority of HOTL RAS, such as driverless cars or unmanned aerial vehicles, greatly rely on sensors, onboard computers, and networks to collect, store, process, and communicate data. Therefore, as software-intensive HOTL RAS become more prevalent, the threat of attacks from hackers grows. Cyberattacks on these systems through the digital network (hacking) could lead to catastrophes. Vulnerabilities exploited by cyberattackers could be in the software of RAS, in security policies, or in communication technologies. Poorly designed and protected RAS become vulnerable targets for criminals or extremists to remotely take partial or full control of these systems and misuse them to achieve their objectives. An attacker can manipulate all safety-critical systems through accessing a vehicle's electronic control unit [29].

There is no question or doubt that RAS such as driverless cars could be hijacked. A car hijacked by a malicious party becomes an immediate physical threat. While car manufacturers focus on safety engineering, the security aspect of software engineering needs to be enhanced.

Better Legislation and Ethical Regulation

How and when RAS can make a correct ethical decision is a hot issue. For example, should a driverless car with full autonomy sacrifice its passengers to avoid crashing into a bus or building full of people? Another example is RAS used in combat. The use of HOTL RAS in the military raises many moral and ethical concerns. For example, is it permissible for HOTL RAS to sacrifice one innocent person to help save tens of others? Existing RAS technology complies with the laws of war and rules of engagement. However, the artificial intelligence and robotics societies face an important ethical decision: whether to support or oppose the development of lethal autonomous weapons systems (LAWS) [30]. *LAWS* is defined as the third revolution in warfare, after gunpowder and nuclear arms.

Programming and giving license to HOTL RAS to harm or kill under certain conditions are something many experts have warned about. How to ensure that humans and RAS coexist safely is a great challenge. Another is how HOTL RAS could distinguish between red and blue teams or civilians. If HOTL RAS are not able to tell the difference, many innocent people could be hurt or killed. Some environments, including the battlefield, have more complex social rules than those in factories or hospitals. Therefore, HOTL RAS should be able to correctly perceive, create, and understand dynamic models of their surroundings. RAS should accurately categorize objects around them and recognize and detect humans as well as their emotions before making any decision. How humans and the RAS could complement each other is also yet to be fully investigated. Due to these obstacles and the current technology

Figure 4. The Universal Motion Simulator [28].

shortcomings, equipping HOTL RAS with full autonomy has become the Holy Grail.

Legal and regulatory frameworks need to be revised to keep up with the technological advances. The current lack of clarity in the legislation and laws in handling incidents pertaining to HOTL RAS technologies poses an important issue that needs to be resolved. Nevertheless, there is no doubt that humans and RAS should work in harmony, maximizing the strength of both. Indeed, the real frontier is collaboration between RAS and humans, which includes trusted autonomy. However, different levels of autonomy in different RAS technologies should be imposed under the control of a human supervisor. It is imperative that the human component be at the heart of RAS capabilities.

As it is inconvenient and in some cases unsafe to use untested RAS close to humans or where human lives are involved, RAS must be deployed with restricted functionality as assistants to humans until a testing procedure is completed in the future. Therefore, newly produced HOTL RAS should operate only under supervision by humans for all the critical tasks where humans could be at risk, and they should be given more duties only when their ethical decision making, evaluation, and learning algorithms become more effective and efficient.

About the Author

Saeid Nahavandi (saeid.nahavandi@deakin.edu.au) is with the Institute for Intelligent Systems Research and Innovation, Deakin University, Victoria, Australia.

References

[1] D. P. Watson and D. H. Scheidt, "Autonomous systems," *Johns Hopkins APL Tech. Dig.*, vol. 26, no. 4, pp. 368–376, 2005.

[2] J. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *J. Human-Robot Interaction*, vol. 3, no. 2, pp. 74–99, 2014.

[3] P. Scharre and M. Horowitz, "An introduction to autonomy in weapon systems" (working paper), Center for a New American Security, Washington, D.C., Feb. 2015.

[4] D. H. McKnight and N. L. Chervany, "What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology," *Int. J. Electron. Commerce*, vol. 6, no. 2, pp. 35–59, Dec. 2001.

[5] M. Aitken, N. Ahmed, D. Lawrence, B. Argrow, and E. Frew, "Assurances and machine self-confidence for enhanced trust in autonomous systems," in *Proc. Robotics: Science and Systems Workshop on Social Trust in Autonomous Systems*, Ann Arbor, MI, 2016.

[6] A. Smith and J. Anderson. (2014, Aug.). AI, robotics, and the future of jobs. Pew Research Center. Washington, D.C. [Online]. Available: http://www.pewinternet.org/files/2014/08/Future-of-AI-Robotics-and-Jobs.pdf

[7] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to Autonomous Mobile Robots*. Cambridge, MA: MIT Press, 2011.

[8] J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs. (2013, May). Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Global Institute. San Francisco, CA. [Online]. Available: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies

[9] D. Lane and R. Buckingham. (2014, July). RAS 2020: Robotics and autonomous systems. Robotics and Autonomous Systems Special Interest Group, The Knowledge Transfer Group. Horsham, U.K. [Online]. Available: https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20Strategy?version=1.0

[10] S. Welsh. (2016, Jan. 28). We need to keep humans in the loop when robots fight wars. *The Conversation.* [Online]. Available: http://theconversation.com/we-need-to-keep-humans-in-the-loop-when-robots-fight-wars-53641

[11] D. G. Johnson and M. E. Noorman, "Responsibility practices in robotic warfare," *Military Rev.*, vol. 94, no. 3, pp. 12–21, May–June 2014.

[12] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors: J. Human Factors and Ergonom. Soc.*, vol. 46, no. 1, pp. 50–80, Mar. 2004.

[13] A. Lacher, R. Grabowski, and S. Cook, "Autonomy, trust, and transportation," in *Proc. 2014 Assoc. Advancement Artificial Intelligence Spring Symp. Series*, pp. 42–49.

[14] S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," *Inform. Software Technol.*, vol. 73, pp. 136–150, May 2016.

[15] J. R. Boyd, "The essence of winning and losing," unpublished lecture notes, 1996.

[16] W. C. Marra and S. K. McNeil. (2012, May). "Understanding 'the loop': Autonomy, system decision-making, and the next generation of war machines." Lawfare Research Paper Series, Research Paper No. 1-2012. [Online]. Available: https://lawfare.s3-us-west-2.amazonaws.com/staging/s3fs-public/uploads/2012/05/SSRN-id2051596.pdf

[17] B. T. Thomas, "Autonomous weapon systems: The anatomy of autonomy and the legality of lethality," *Houston J. Int. Law*, vol. 37, no. 1, Jan. 2015.

[18] S. E. McIntosh, "The wingman-philosopher of MiG Alley: John Boyd and the OODA loop," *Air Power Hist.*, vol. 58, no. 4, pp. 24, Dec. 2011.

[19] M. Endsley, "Autonomous horizons: System autonomy in the Air Force–A path to the future (Volume I: Human autonomy teaming)," Dept. Chief Scientist, U.S. Air Force, Washington, D.C., Rep. No. AF/ST TR 15-01, June 2015.

[20] T. B. Sheridan, T. Vámos, and S. Aida, "Adapting automation to man, culture and society," *Automatica*, vol. 19, no. 6, pp. 605–612, Nov. 1983.

[21] K. A. Hoff and M. Bashir, "Trust in automation integrating empirical evidence on factors that influence trust," *Human Factors: J. Human Factors and Ergonom. Soc.*, vol. 57, no. 3, pp. 407–434, May 2014.

[22] M. Fisher, L. Dennis, and M. Webster, "Verifying autonomous systems," *Commun. ACM*, vol. 56, no. 9, pp. 84–93, Sep. 2013.

[23] T. B. Sheridan, "Human–robot interaction: Status and challenges," *Human Factors: J. Human Factors and Ergonom. Soc.*, vol. 58, no. 4, pp. 525–532, June 2016.

[24] M. Barnes and F. Jentsch, Eds., *Human-Robot Interactions in Future Military Operations*. Boca Raton, FL: CRC, 2010.

[25] D. Feil-Seifer and M. J. Matarić, "Human–robot interaction," in *Encyclopedia of Complexity and Systems Science*. New York: Springer-Verlag, 2009, pp. 4643–4659.

[26] B. Durán and S. Thill, "Rob's robot: Current and future challenges for humanoid robots," in *The Future of Humanoid Robots: Research and Applications*, Riadh Zaier, Ed. Rijeka, Croatia: INTECH Open, 2012.

[27] I. J. Cox and G. Wilfong, Eds., *Autonomous Robot Vehicles*. New York: Springer-Verlag, 1990.

[28] S. Nahavandi, "Robot-based motion simulators using washout filtering: Dynamic, immersive land, air, and sea vehicle training, vehicle virtual prototyping, and testing," *IEEE Syst., Man, Cybern. Mag.*, vol. 2, no. 3, pp. 6–10, July 2016.

[29] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *Proc. 2010 IEEE Symp. Security and Privacy*, 2010, pp. 447–462.

[30] S. Russell. (2015, May 27). Ethics of artificial intelligence. *Nature*. [Online]. Available: losnuevosguerreros.org/pluginfile.php/16/mod_glossary/attachment/1279/Ethics%20of%20artificial%20intelligence.pdf

SMC