Future Directions in Human Machine Teaming Workshop

July 16-17, 2019
Arlington, VA

John Laird, *University of Michigan*

Charan Ranganath, *University of California, Davis*

Samuel Gershman, *Harvard University*

Prepared by:
Kate Klemic, *VT-ARC*
Rushyannah Killens-Cade, *AAAS S&T Policy Fellow, OUSD(R&E)*

Future Directions Workshop series
Workshop sponsored by the Basic Research Office, Office of the Under Secretary of Defense for Research & Engineering

VT-ARC
Virginia Tech
Applied Research Corporation

Contents

Innovation is the key to the future, but basic research is the key to future innovation.

–Jerome Isaac Friedman,
Nobel Prize Recipient (1990)

Preface

Over the past century, science and technology has brought remarkable new capabilities to all sectors of the economy; from telecommunications, energy, and electronics to medicine, transportation and defense. Technologies that were fantasy decades ago, such as the internet and mobile devices, now inform the way we live, work, and interact with our environment. Key to this technological progress is the capacity of the global basic research community to create new knowledge and to develop new insights in science, technology, and engineering. Understanding the trajectories of this fundamental research, within the context of global challenges, empowers stakeholders to identify and seize potential opportunities.

The Future Directions Workshop series, sponsored by the Basic Research Directorate of the Office of the Under Secretary of Defense for Research and Engineering, seeks to examine emerging research and engineering areas that are most likely to transform future technology capabilities. These workshops gather distinguished academic researchers from around the globe to engage in an interactive dialogue about the promises and challenges of each emerging basic research area and how they could impact future capabilities. Chaired by leaders in the field, these workshops encourage unfettered considerations of the prospects of fundamental science areas from the most talented minds in the research community.

Reports from the Future Direction Workshop series capture these discussions and therefore play a vital role in the discussion of basic research priorities. In each report, participants are challenged to address the following important questions:

- How will the research impact science and technology capabilities of the future?
- What is the trajectory of scientific achievement over the next few decades?
- What are the most fundamental challenges to progress?

This report is the product of a workshop held July 16-17, 2019 at the Basic Research Innovation Collaboration Center in Arlington, VA on the future of human machine teaming research. It is intended as a resource to the S&T community including the broader federal funding community, federal laboratories, domestic industrial base, and academia.

Executive Summary

Interactions with technologically sophisticated artificial intelligence (AI) agents are now commonplace. We increasingly rely on intelligent systems to extend our human capabilities, from chatbots that provide technical support to virtual assistants like Siri and Alexa. However, today's intelligent machines are essentially tools, not teammates. They require the undivided attention of a human user and lack the communicative or cognitive capabilities needed to interact as trusted teammates. To become true teammates, the intelligent machines will need to be flexible and adaptive to the states of the human teammate, as well the environment. They will need to intelligently anticipate their human teammate's capabilities, intentions, and generalize specific learning experiences to entirely new situations.

On July 16-17, 2019, a Future Directions of Human Machine Teaming workshop was held at the Basic Research Innovation Collaboration Center in Arlington, VA to examine the basic research challenges and opportunities to enable this level of human machine teaming. Hosted by the Basic Research Office in the Office of the Under Secretary of Defense for Research and Engineering, this workshop gathered 21 distinguished researchers from the AI, robotics, cognitive science, psychology, and neuroscience communities across academia, industry, and government. Participants debated and discussed how research in these different areas can inform how humans and intelligent machines can work together.

The workshop participants considered current research areas in AI and cognitive sciences, discussed challenges and knowledge gaps, and promising areas for future research advances. This report is the product of those discussions, summarizing the key research challenges, opportunities, and trajectory for research needed to enable true human machine teaming. Several key themes emerged for human machine teaming that we present as a model framework where each teammate creates mental representations of self, teammate, and the team that guide perception, communication, and joint action. With this framework in mind, the research challenges and opportunity areas are divided into four topics:

Human Capabilities: Natural Intelligence – research to better understand human cognitive capabilities in the context of complex and dynamic situations.

Challenges include understanding the human ability to create mental representations of situations ("mental models"), the goals, intentions, and abilities of other people ("theory of mind"), and shared knowledge with a communication partner ("common ground"). In addition, we need a better understanding of how humans learn from single events, to make predictions and generalizations in new situations, and to build knowledge about situations and events that can guide reasoning and deductive inferences ("common sense").

Research Areas include communication studies to understand natural language in real-world situations that can be used to scale up computational models that integrate real-world complexity and uncertainty; learning studies to understand the human capability to learn from single instances, generalize where appropriate, and show transfer to new situations; curiosity studies that examine the factors that drive active information seeking and exploration to actively generate questions in order to fill in knowledge gaps; task learning and generalization studies that characterize how tasks can be represented in a manner that is compositional, such that information can be reused and recombined across tasks; task control and multi-threading studies to understand the constraints on the human operator and the nature of task representations; integrated cognition studies for complex, dynamic environments to develop integrated cognitive and brain-inspired models in which different processes interact in a manner that can emulate human performance; and systems that decode neural signals for human machine interactions.

Human Models of Machines – research to understand what humans must know and learn about machines and their physical and internal structure in order to effectively and efficiently interact with them, including what is required in human machine teaming to establish and maintain trust.

Challenges include understanding how humans represent and reason about machine mental and physical abilities; defining the level of machine description needed; and the level of transparency required to get humans to trust the machine.

Research Areas include real-world team experiments to learn how different humans react to different mixtures or levels of physical and cognitive capabilities in machine systems; team experiments with different machines behavior to determine which makes the machine legible and predictable so that a human can easily infer a machine's intentions and predict machine behavior; explainable AI team experiments to determine when and what the machine should explain so that the human can maintain a valid model of its teammate; team experiments to determine the factors needed to build and maintain human trust in the machine.

Machine Capabilities: Artificial Intelligence – research to improve intelligent machine capabilities in order to enable effective human machine teams.

Challenges include understanding the additional capabilities that a machine needs to be an effective teammate in the relevant tasks and environments. The capability areas include perception and motor control; communication; modeling the environment and itself; reasoning, problem solving, planning, common sense, and task expertise; learning; and integrated architectures.

Research Areas include perception studies such as activity recognition that builds on new advanced sensor, machine learning algorithms and computing hardware to provide robots

with the ability build internal models of their environments and predict and evaluate future states; communication studies to ground the meaning of a communication to context and environment; studies that use computer game engines of the world to improve and develop machine models of the environment and the machine's own capabilities; teaming studies on perspective taking, joint attention, and cooperation and coordination to develop new reasoning, problem solving, planning and task expertise; human-robot interaction and interactive task learning studies where humans teach AI systems new tasks through demonstration and language; new or extended integrated architectures that provide frameworks for developing and integrating many, if not all of the capabilities required for machine teammates.

Machine Models of Humans – research to understand and realize the internal representations and processing of a machine required for reasoning about human teammates. A machine's model of itself and a model of the team structure are also important, but the focus of workshop discussions was on how machines represent and reason about human teammates.

Challenges include understanding which aspects, and to what fidelity, do machines need to model the physical capabilities and minds of humans given the demands of the tasks; understanding how to construct and dynamically model individually and collectively the many components that make up the human mind, including perceptual, motor, planning, and abilities; and understanding joint attention, theory of mind, and perspective taking.

Research Areas include applied social science and human-robot interaction studies to determine which aspects of human-human teaming are necessary to support effective and robust human machine teaming. These studies will define which aspects of human behavior need to be modeled for different teaming situations and establish the range of human perception and motor control abilities, as well as human reasoning and planning abilities, so that the machine can reason about its human teammates. Lastly, studies to enable dynamic models that personalize individuals for specific tasks and then tracking those individuals through the course of a task.

The participants noted that there is significant interaction among these four topics - progress in one will impact the others, and vice versa, and all impact our ability to develop machines that establish and maintain trust with human teammates. Success will require increased dialogue and collaborations across the fields of computer science, robotics, psychology, and neuroscience. The participants also advocated for the development of open datasets and corpuses to facilitate computational modeling and development of intelligent agents. Lastly, they recommended the development of specific "use cases," that is, specific examples of where human machine teaming would have a big impact. By providing a detailed task analysis, researchers can focus on the particular challenges to development of intelligent teammates in a particular situation. Building along these lines, research competitions and prizes can provide added motivation for teams to tackle particular human machine teaming challenges. The participants outlined a trajectory for research in the near- and far-term for each of these topics areas. In general, the consensus was that in the near-term (5-10 years), intelligent machines will have simplified forms of the desired capabilities. They will have simple natural language communication skills and perform simple task planning and reasoning in controlled environmental conditions. The participants were optimistic that the flexible and adaptive intelligent machines are achievable in the far-term (10-20 years). These intelligent machines will intelligently anticipate the teammate's capabilities, intentions, and generalize specific learning experiences and enable true human machine teaming.

Introduction

Human-to-human teaming involves multiple individuals banding together in pursuit of a common goal (Salas, Dickinson, Converse, & Tannenbaum, 1992; Salas, Cooke, & Rosen, 2008; Cooke et al., 2013). Sports teams, music groups, non-profits, business units, as well as small-scale military organizations, are common examples of situations where multiple individuals work together, not in the service of a single individual, but in the service of the team. A team can take advantage of the skills and the expertise of individuals, sometimes through loose coordination where each member works almost autonomously, or tightly coupled where low-level actions of members are performed in concert with others to achieve goals that would be impossible for a single person on their own. At their best, teams amplify the unique skills and capabilities of their members, relying on members to trust each other and sublimate their individual goals for those of the group.

For most of their history, intelligent machines have been tools and not teammates. We have developed increasingly sophisticated devices using remote operation, but these devices require the undivided attention of a human user. Robotic assistants have been developed in domains and applications ranging from human personal assistants to search and rescue operations (Heard et al., 2019; Gombolay et al., 2017; Lasota et al., 2017; Barlett & Cooke, 2015). We are coming to increasingly rely on powerful tools such as self-driving cars, chatbots that provide technical support, and virtual assistants like Siri and Alexa. At best, these technologies are useful in that they extend human capabilities, but their communicative and cognitive capabilities have been inadequate for being a useful and trusted teammate. For instance, consider this real exchange between a human and the virtual assistant Siri:

User: "Play a good song"

Siri: "Sorry, I couldn't find 'A Good Song' in your music."

This exchange helps to illustrate why, despite considerable technological advances, there are many challenges before we can fully trust machines to autonomously handle high-risk, complex operations (e.g., driving in contexts with unpredictable pedestrian behavior) or function as an autonomous teammate in critical situations (e.g., a member of a military unit). Although there are challenges, there continue to be examples of intelligent machines successfully teaming with humans, although in limited contexts. In military training, autonomous AI systems (virtual agents) have been used to not only populate the battlespace with friendly and enemy units (Jones et al, 1999; Hill et al. 1998), but also as simulated co-pilots (Cooke et al., 2013), members of nautical maintenance staff (Rickel & Johnson, 2003), and as virtual humans (Traum et al., 2003). Most of these teaming situations have involved interactions with a human during execution of a constrained task through restricted natural language.

In this Future Directions workshop, participants discussed how we might transition from using machines as human-controlled tools for accomplishing specific tasks to intelligent machines and virtual agents that cooperate and partner with humans across a variety of domains. In this framework, the machine is flexible and adaptive to the states of the human teammate and the environment, intelligently anticipating the teammate's capabilities, intentions, and generalizing specific learning experiences to entirely new situations. There is considerable potential to be gained in developing intelligent machines that can function in a team with humans. Machines can possess sensory (infra-red, FLIR, sonar, etc.) and motor (flying, fast land travel, precision surgery, etc.) capabilities that humans do not possess. They also can perform tasks over and over again, without becoming bored or fatigued, maintaining a level of vigilance that would be difficult for a human. In some cases, they can communicate and access resources (such as data on the web) faster and more precisely than a human, and they can possess computational capabilities (such as complex mathematical calculations) that are beyond those of humans and data analysis. A machine can also be used in dangerous or extreme environments without risk of loss of life. Development of autonomous machines that use such capabilities to collaborate with human partners can have a transformative effect across many commercial and military applications.

To illustrate both the possibilities and the challenges in achieving the goal of true human machine teaming, consider an example from the comic book and film "Iron Man." The human protagonist, Tony Stark, has an AI assistant named J.A.R.V.I.S. ("Just A Rather Very Intelligent System") who provides real-time information, completes tasks, operates other machines, and even provides emotional support based on Tony's needs at any given time. Here is one vignette from the film (Arad et al., 2008):

Tony: "Jarvis, you up?"

J.A.R.V.I.S.: "For you, sir, always."

Tony: "I'd like to open a new project file, index as Mark Two. "

J.A.R.V.I.S.: "Shall I store this on the Stark Industries Central Database?"

Tony: "Actually, I don't know who to trust right now. Till further notice, why don't we just keep everything on my private server?"

J.A.R.V.I.S.: "Working on a secret project, are we, sir?

Tony: "I don't want this winding up in the wrong hands."

From this snippet of dialogue, we can appreciate that J.A.R.V.I.S. has extraordinary cognitive capabilities. When asked to open a new project file, J.A.R.V.I.S. immediately anticipates that Tony might—or might not—want to store it in the central database. When asked "why don't we just keep everything on my private server?" J.A.R.V.I.S. does not recognize it as a question, but rather as an implicit command. J.A.R.V.I.S.' response reflects his understanding that Tony probably intends to keep this new project confidential until further notice. Perhaps what is most noteworthy about this example, however, is that Tony's interactions with J.A.R.V.I.S. reflect a remarkable level of comfort and trust. Tony fundamentally believes that J.A.R.V.I.S. can understand his underlying intentions, and that he can be trusted to intelligently collaborate with him towards achieving their shared goals.

Over the last 20 years, research in psychology and neuroscience and technological advances in artificial intelligence and robotics have led to insights into the underlying capabilities needed to support this level of autonomy and teaming. The goal of this workshop was to bring researchers from these communities together to discuss the promise, possibilities, and challenges for the development of intelligent machines that can team with humans to cooperatively solve complex problems in dynamic environments.

Future Human Machine Team Scenarios

The workshop participants worked in small groups to discuss the future of human machine teaming in the context of concrete scenarios of human machine teams of the future. Three potential scenarios were envisioned:

1. Intelligent Assistant Teams

Three small groups discussed related visions of human machine teaming based on an intelligent assistant that autonomously guides human decision-making and learning. 1) An "intel" assistant that is proactive, understanding context and asking questions to fill in gaps of knowledge, and adaptive to the user needs (provides relevant information when the user needs it). 2) A "cyber-security" assistant that monitors security threats, "knows" what information is accurate and relevant, and dynamically suggests solutions. 3) A "cooperative learning" assistant that bootstraps knowledge to improve not only the students' learning but also its own learning.

2. Naval Maintenance Teams

One group discussed future human machine teaming in the context of performing routine naval maintenance tasks. Handed a Maintenance Requirement Card (MRC), the team will determine a division of labor and work cooperatively on their common goal. The intelligent system will understand and predict its human collaborator's actions and perform both general and skilled object manipulations, as needed. It will have rich reasoning capabilities of time, space, causality, and identity so that it can adaptively respond to the environment.

3. Disaster Search, Rescue, and Recovery Teams

One group discussed the potential for human machine teams to work cooperatively in disaster search, rescue, and recovery operations. They envision multiple levels of autonomous systems working cooperatively with humans to search for survivors (autonomous air vehicles), move rubble (autonomous movers), and administer medical treatment (robotic medics), see Figure 1. The intelligent systems of these teams will have improved physical, cognitive, and social capabilities that can dynamically adapt to changes in environment and acquire knowledge through experience and interactions with human teammates.

Figure 1. Human machine teaming scenario for disaster search, rescue, and recovery operations. Autonomous robotic systems work collaboratively with humans to search for survivors (high altitude, fixed wing drone, top; hover drone, right; land scout, left), and transport wounded (walking robot carrying injured, center-right; bed transport, center-left).

Framework for Human Machine Teaming

The key themes from the envisioned human team scenarios can be distilled into the framework illustrated in Figure 2. See Wynne & Lyons (2018) for a more general discussion of teammate concept structure.

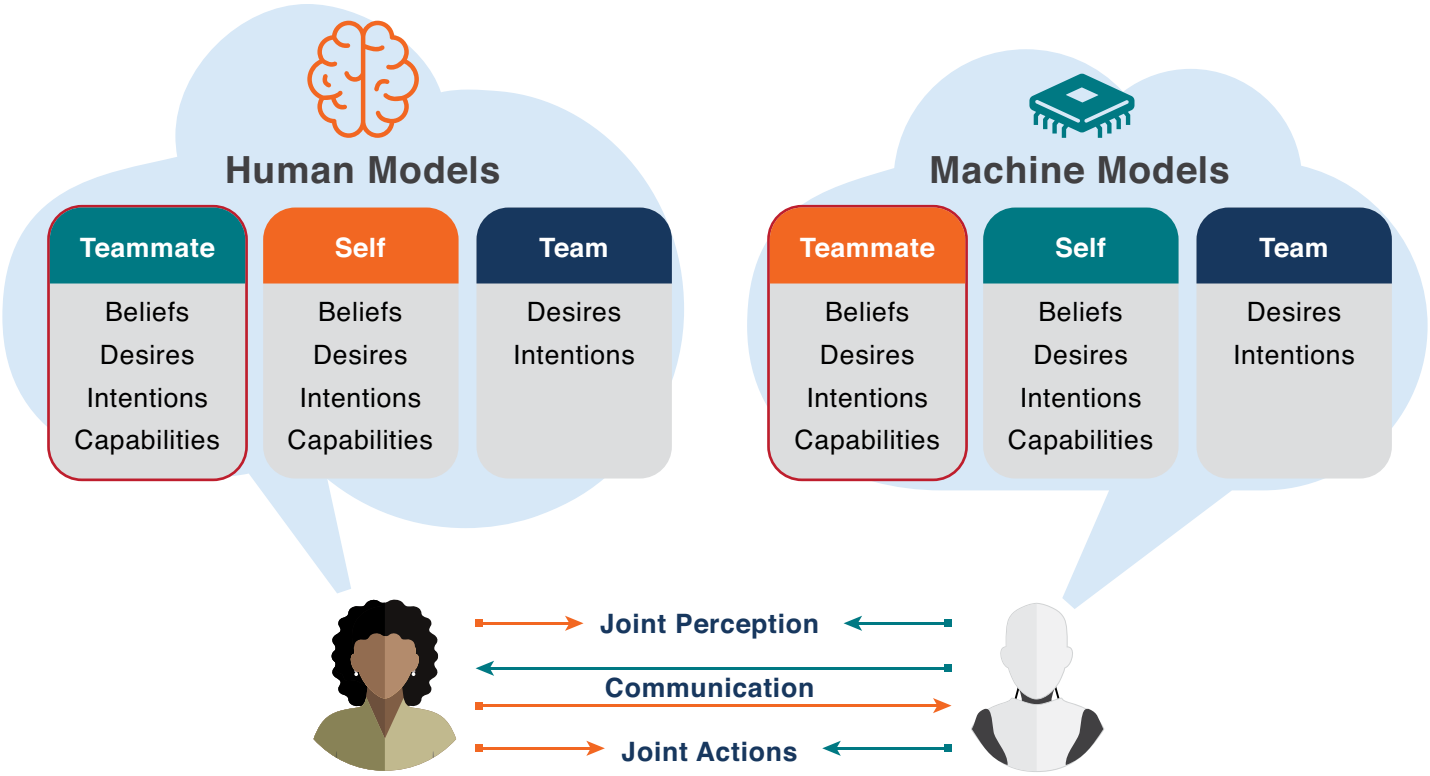


Figure 2. A model-based framework for human machine teaming. Each teammate creates model representations of self, teammate, and the team that guide perception, communication, and joint action The teammate representation (red outline) is critical to successful teaming.

A group of humans and machines (here shown as a single human machine pair) collaborate on a common task, such as preparing a meal. The members of the team use communication to agree on the task for the team (what meal to prepare), to divide the task into subtasks (who cooks which dish), and to coordinate joint activities (emptying a large pot into a serving dish). Communication is also used to inform teammates of individual intentions, such as when a teammate will need a specific utensil or pot. The creation of a joint understanding of the situation is called establishing "common ground."

Each team member maintains a representation of their own beliefs, desires, and intentions (labeled "self" in the diagram), but also uses communication and perception to maintain a representation of the team's goals and plans ("team"), as well as models of teammates ("teammate"), highlighted in red for both the human and the machine. Having a model of another individual's physical and mental capabilities, or "theory of mind" (Premack & Woodruff, 1978), is evident even in young children, and it is central to human communication and interactions (Baron-Cohen, 1995). Being able to model a teammate can dramatically reduce communication, allowing one member to predict the actions of the other, even when those actions are outside those of the team goals, such as when one member must temporarily answer the phone or feed the dog.

Teammates have many responsibilities, which can make human machine teaming challenging. A member needs to inform other members if that member is unable to complete a task, or even if it finishes a task early so that it is available to help out in new ways. They also must track other teammates, so they know when they are expected to help them, possibly even without direct communication. In general, a teammate must be predictable, so other teammates can synchronize their behavior appropriately, with all of these activities having an underlying need to establish and maintain trust throughout the team.

With this framework in mind, the following sections describe the research challenges and opportunities for this vision of human machine teaming to be achieved over the next 20 years.

Research Challenges in Human Machine Teaming

Teams can work toward specific, clearly defined goals, or toward goals that are broad and complex. Given the vast scope of applications for human machine teams, discussions at the workshop centered on identifying topic areas and research questions that could cut across many applications of human machine teaming. Underlying these discussions was a need to better understand human intelligence and to make additional advances in artificial intelligence. What makes effective teamwork especially challenging is that it requires the integration of several components (Groom and Nass, 2007) including joint attention and common ground, motivation toward team versus individual objects, action toward team objects, and trust among team members. This section describes four key research challenges for human machine teaming identified by participants:

Human Capabilities: Natural Intelligence – research to better understand human cognitive capabilities in the context of complex and dynamic situations.

Human Models of Machines – research to understand what humans must know and learn about machines and their internal structure in order to effectively interact with them, including what is required in human machine teaming to establish and maintain trust.

Machine Capabilities: Artificial Intelligence – research to improve intelligent machine capabilities in order to enable effective human machine teams.

Machine Models of Humans – research to establish the internal representations and processing of a machine for reasoning about human teammates. A machine's model of itself and a model of the team structure are also important, but the focus of workshop discussions was on how machines represent and reason about human teammates.

Although we present these challenges individually, there are significant interactions among them. Progress in one will impact the others, or vice versa. Lack of progress in one area can inhibit progress in the others. This positive feedback loop is shown in Figure 3. Advances in our understanding of natural intelligence leads to better models of human capabilities. Together with advances in AI, better machine models of humans can improve the ability of machines to reason about human teammates. The improved behavior of machines leads to better teaming experiments, where we can learn more about how humans interact with intelligent machines, which then leads to improved human models of machines. All of these also impact our ability to develop machines that establish and maintain trust with human teammates.

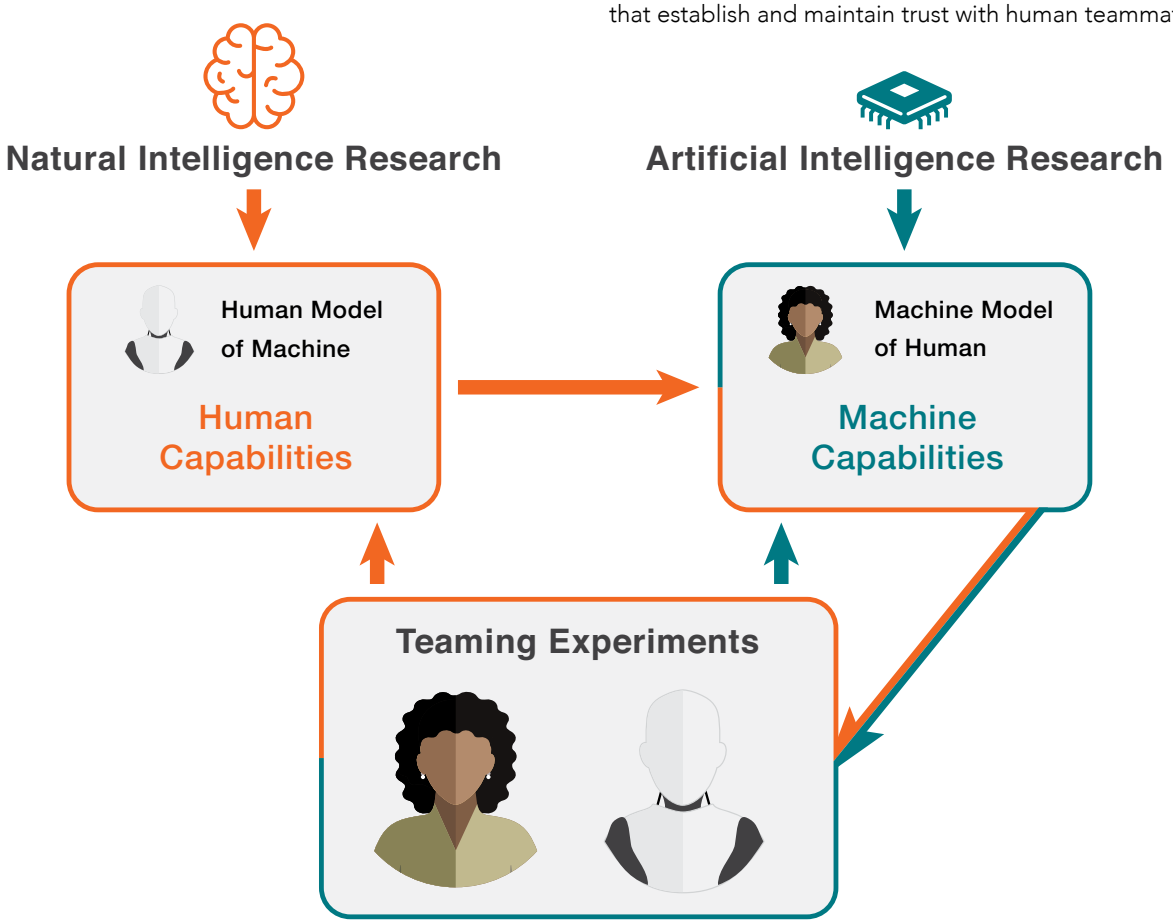


Figure 3. Positive feedback loops drive progress in all aspects of human machine teaming research. Advances in our understanding of human capabilities guide advances in machine capabilities, which leads to better teaming experiments. Results from teaming experiments improve human models of machines and machine models of humans, and the cycle continues.

"Human Capabilities: Natural Intelligence" Challenge

There was a general consensus among workshop participants that we lack an adequate understanding of human high-level cognition, motivation, and social behavior, and that understanding human cognition is critical to the other challenges. Humans excel at learning and problem solving in ways that differ from even sophisticated machines, but as discussants pointed out, beyond coarse descriptors (e.g., "common sense"), the nature of human intelligence remains elusive. Of course, substantial research efforts in cognitive science have been directed at understanding how humans think, learn, and act, but most of this research has been conducted in highly controlled laboratory environments and narrow task domains. In natural environments, the sequence of actions that lead to a goal is not explicitly instructed, and often people must rely on past experiences to make novel inferences and predictions. Basic research on the human brain computations that support perception, cognition, and goal-directed actions in complex situations and extended timescales is lacking. In other words, we do not know enough about human cognition in the kinds of complex and dynamic situations that can benefit the most from human machine teams.

Workshop participants identified several aspects of human intelligence that are generally appreciated but poorly understood. The ability to create mental representations of situations ("mental models"), the goals and intentions of other people ("theory of mind"), and shared knowledge with a communication partner ("common ground") were key themes. Robust learning is another key human capability that is central to human machine teams. Humans can learn information from single events ("episodic memory") to make predictions and generalizations in new situations, and to build knowledge about situations and events that can guide reasoning and deductive inferences ("common sense").

Another topic critical for human machine teams is the nature of goal-oriented behavior. Researchers highlighted the ability to represent goals and tasks at multiple levels of abstraction (e.g., "satisfy hunger" vs. "cook dinner" vs. "make soup"), such that past experiences can be leveraged to rapidly learn new tasks or to flexibly modify approaches to an existing task. Humans might expect teammates to display all of these capabilities, but these capabilities are challenging for machine design. Significant improvements might be possible by developing computational architectures in tandem with novel behavioral and neuroscience research paradigms that investigate cognition in complex, naturalistic environments at extended timescales.

"Human Models of Machines" Challenge

Beyond improving our overall understanding of human cognition, within the context of human machine teaming, we need a much better understanding of how humans represent, reason about, and reason with machine teammates. Machines often have very different physical embodiments and capabilities from humans that impact the roles they can play in a team. On the cognitive side, humans often approach machines with an understanding of their own desires, intentions, and capabilities, and they routinely ascribe complex human-like intentions and beliefs to machines even with minimal stimuli such as moving shapes on a screen (Heider & Simmel, 1944). This can lead to overconfidence in a machine's cognitive capabilities, which, in turn, can lead to frustration and ultimately failure in human machine teams. We also want to understand the impact of having a machine teammate on the human. Although a teammate can help, it can also add an extra workload burden to a human if the human must continually attend to it, checking its behavior, etc. These are fundamentally questions of psychology and neuroscience but require research in intelligence machines in order to address them.

There are many open questions as to how complete a model of a machine's internal state a human requires for effective teaming. Is it sufficient to have an abstract characterization of the machine's goals and intentions, or are more detailed characterization required? What aspects of the machine's internal state and intentions can the human infer from the machine's actions? From the machine design perspective, we will also want to understand how to make machines easier to understand and work with. For those, one open question is whether it is important to create machines that "think like humans?" One theory is that if machines reason and behave like humans, humans can use all of their capabilities for tracking and predicting other humans in teaming with robots. A counter argument is that for many tasks, models that are simpler than full scale human reasoning may be easy to process.

Beyond tracking behavior and reasoning, there is explanation of behavior. Already we have seen the challenge of the opacity of some machine learning techniques where it is difficult for a human to understand why a machine makes a specific decision. A continuing challenge will be to create systems that are both competent and can explain their behavior.

These issues in turn lead to questions about how to train humans so that they have an accurate model of the capabilities of the machines they work with, including a machine's strengths and weaknesses, and especially their unusual behaviors and failure modes.

One point that came up throughout the workshop is that a human must be able to trust a machine teammate. We still have limited understanding of how machines must be designed to earn and maintain that trust. This is especially important in dynamic and complex settings, and open world applications, where it is a certainty that the model will fail at some point to produce an action that the human thought was appropriate, or it produces an action that the human perceives to be unhelpful. Indeed, the history of human machine teaming is rife with examples where people abandoned costly engineered tools because the tool was unable to establish or maintain trust of the user.

"Machine Capabilities: Artificial Intelligence" Challenge

Figure 2 identified in broad strokes some of the representational

"One of our challenges is developing a science of teaming where we gain an understanding of the additional capabilities that a machine needs to be an effective teammate in the relevant tasks and environments."

capabilities required of a machine teammate. The artificial intelligence challenge is to determine the details of all of the capabilities required in an effective machine teammate.

One complexity in this challenge is that the cognitive needs for a machine teammate differ across tasks and teaming arrangements. We see examples of this in existing human teaming arrangements where humans effectively team with other humans, but also with animals such as dogs, horses, etc. We don't need a complete human-level machine teammate for all (and probably most) tasks. In many cases, a machine will have very different embodiments than humans - different sensors and motor systems, and different computational capabilities. One of our challenges is developing a science of teaming where we gain an understanding of the additional capabilities that a machine needs to be an effective teammate in the relevant tasks and environments. These capabilities include:

Perception and Motor Control. To support joint activities, the machine must have perceptual and motor capabilities sufficient to create an internal model of the environment and act on it. In addition, the machine needs to be able to interpret and understand a human's actions (Sukthankar et al., 2014) as they relate to team joint activities, inferring teammate goals and intentions in real time. Often these actions are subtle, such as a change in gaze, facial expression, or intonation, making them difficult for a machine to detect and understand. Moreover, the machine sensing of the environment and teammates is often noisy and incomplete, making it challenging to interpret actions and infer intentions from perception alone.

Communication. An obvious enabler of effective teaming is communication. Communication includes not only language, but gestures, and even interpretation of emotion expression. Even with the rise of personal assistants, and the seeming ubiquity of AI systems that process language, supporting general language understanding and production is still beyond the state of the art. Progress is being made in restricted environments (McNeese et al., 2018), but this will continue to be an open area of research for the foreseeable future. Gesture and emotion expression are active areas of research.

Modeling the Environment and Itself. Beyond perception and communication, for complex tasks, the machine must be able to model the dynamics of the environment and its own capabilities, so that it can predict and evaluate possible futures.

Reasoning, Problem Solving, Planning, Common Sense, Task Expertise. A machine teammate must also have the requisite cognitive capabilities to perform its tasks, and coordinate with the teammates, etc. Although general human-level capabilities in this area are still years away, the discussions in the meeting often centered on existing human capabilities in this area, and how incorporating them in machines is a critical challenge. For example, natural communication between two individuals relies on the ability to generate mental models about situations (van Dijk and Kintsch, 1983; Richmond & Zacks) and integrate assumptions about the communication partner's goals and knowledge ("common ground," cf. Clark, 1992; Brown-Schmidt & Duff, 2016). Humans generate mental models by drawing on knowledge about particular situations (Hard Tversky, & Lang, 2006) and memories of specific events (Schacter, Addis, & Buckner, 2008). Humans are also capable of representing goals and sequences of actions that lead to a goal in a hierarchical and abstract manner (Botvinick, 2008), such that they can use knowledge about previous tasks to rapidly learn related tasks (e.g., knowing how to make a cake can facilitate learning how to make cookies). Furthermore, there are additional capabilities that are needed to support teaming, such as perspective taking and maintaining joint attention. A machine

that exhibits some of these capabilities could more effectively communicate and collaborate with human teammates and fulfill its role on the team without extensive human supervision.

Learning. Although much of a machine teammate's knowledge can be defined offline, a machine teammate may need to dynamically learn from its environment, both to improve its task performance, but also to improve its model of its human teammates to better anticipate their goals, beliefs, actions, and interactions.

Integrated Architectures. Beyond individual components, an ongoing research challenge is how these components work together to create coherent, effective behavior for complex problems that involve many different types of reasoning and problem solving. Many approaches to AI provide solutions to specific types of problems, where in many teaming situations, the machine must use a variety of approaches to solve many different types of problems.

"Machine Models of Humans" Challenge

One specific aspect of machine capabilities that elicited much discussion was machine modeling and reasoning about human teammates. As shown in Figure 2, a machine needs some kind of "mental model" of human partners, in terms of their physical and mental capabilities (e.g., memory, attention, and reasoning) and their goals and motivations. Although understanding human physical capabilities is important, the emphasis in the discussions was on mental capabilities.

The first challenge is understanding which aspects, and to what fidelity, do machines need to model the minds of humans given the demands of the tasks. One can imagine that in some situations, the model might be implicit—baked into the machine's design based on the designer's assumptions about the human teammate's goals, intentions, and capabilities. In other cases, the machine may require an explicit model of only the human's goals, while in others, the machine would be most effective if it could predict the human's motor actions with sub-second accuracy. As illustrated in the J.A.R.V.I.S. example, natural interactions and communication might even require a machine to generate a relatively rich model of the human based on the context and situation. Furthermore, the machine design might be improved by modeling the limitations of human cognition and brain function. For instance, machines could tailor delivery of information and decision options based on inferences about a human teammate's attentional capacity and emotional state.

There are additional challenges as to how to construct (Hayes and Scassellati, 2016) and dynamically model individually and collectively the many components that make up the human mind, including perceptual, motor, planning, and abilities. Teamwork adds extra components, such as understanding joint attention, theory of mind, and perspective taking. Although these are active areas in AI for building machine intelligence, the vast majority of work on modeling human abilities comes from the cognitive sciences: cognitive psychology, cognitive neuroscience, linguistics, etc.

Research Opportunities in Human Machine Teaming

Research in AI and robotics, psychology, and neuroscience, have laid the foundation for future advances in human machine teaming. Below are promising research opportunities that can address the four research challenges described above.

"Human Capabilities: Natural Intelligence" Research

Advances in the understanding of human capabilities are necessary to understand how humans might optimally interact with machines and to inform design of flexible, intelligent systems for human machine teaming. Below is a description of research areas that show promise in advancing our understanding of the cognitive and neural foundations for human intelligence.

Communication. Research in artificial intelligence has made progress on automated speech recognition, comprehension, and generation. Recent computational models have focused on interpretation of subtle aspects of language such as indirect, non-literal, and pedagogical speech (Goodman & Frank, 2016). These models assume that the speaker is optimizing speech to account for the listener's beliefs and desires, and the listener, in turn, uses this principle to guide speech comprehension, see Figure 4. Nonetheless, unrestricted natural language comprehension continues to be beyond current AI capabilities. Human language comprehension relies on context, prior knowledge about events and situations (Van Dijk & Kintsch, 1983), memories of specific events (Duff & Brown-Schmidt, 2012), beliefs about the speaker's identity (Eckert, 2012), and principles that communicators are expected to follow (Grice, 1975; Goodman & Frank, 2016). Experimental research on pragmatic language understanding has mostly been restricted to very simple laboratory experiments, and in the near term, psychologists and neuroscientists need to develop experimental paradigms to understand natural language in real-world situations. In the long-term, it will be necessary to translate empirical findings from these studies into scaled up computational implementations that can cope with real-world complexity and uncertainty.

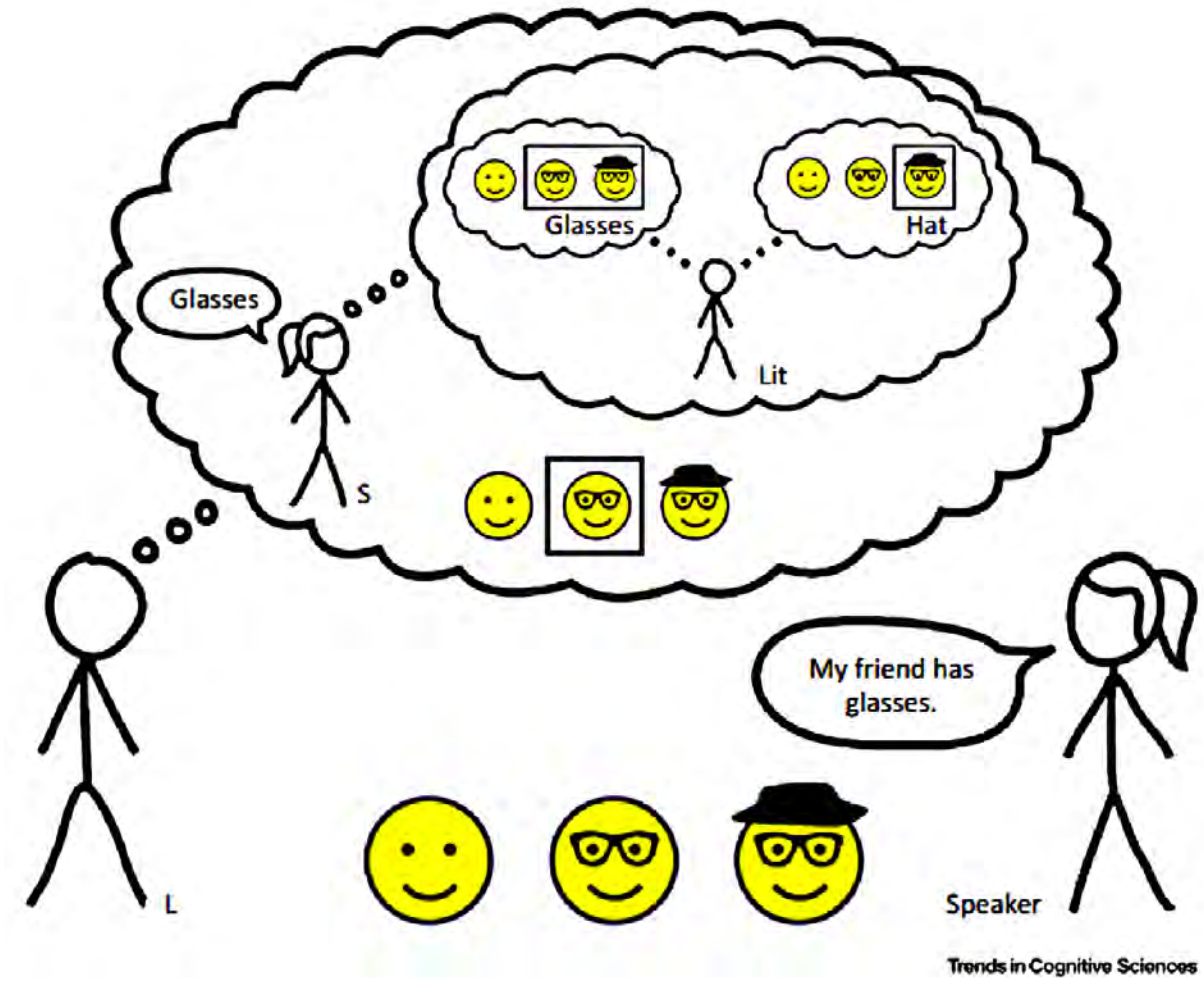


Figure 4 Rational Speech Act (RSA)-style reasoning applied the signaling game. The three faces along the bottom show the signaling game context. Agents are depicted as reasoning recursively about one another's beliefs: listener L reasons about an internal representation of a speaker S, who in turn is modeled as reasoning about a simplified literal listener. Lit. Boxes around targets in the reference game denote interpretations available to a particular agent. [Credit: Goodman and Frank, 2016]

Robust Unsupervised Learning. Humans have the capability to learn from single instances, generalize where appropriate, and show transfer of knowledge to new situations—capabilities that have proven to be a challenge for current AI approaches. Recent work suggests that, as an alternative to supervised reinforcement learning or Hebbian learning, unsupervised learning may be accelerated by using biologically-inspired predictive learning (O'Reilly et al., 2014), along with constraints that humans typically use to regularize learning. Another promising approach may be to flexibly control learning through attentional mechanisms that determine when and how much to learn, as in Neural Turing Machines (Graves et al., 2014).

Another approach to improving learning is to adopt multiple, complementary representations that interact with one another (O'Reilly et al., 2014). Humans can learn and remember specific experiences (episodic memory), but they also can build cognitive maps and schemas that capture knowledge about the structure of events and the environment (Bellmund et al., 2018; Ekstrom & Ranganath, 2018; Stachenfeld et al., 2017). Research in this area will be especially important to addressing the challenge of lifelong learning—that is, continuous acquisition of knowledge over extended timescales. If the human brain has a generic representational format for many different kinds of knowledge, then computer scientists may be able to develop similar systems in artificial agents.

Understanding how these mental representations are learned will allow us to build more human-like artificial memory systems, and also improve the ability of machines to reason about the memory abilities and limitations of human partners. Studies in animal models indicate that representations of specific experiences may be reactivated during sleep or rest, such that the brain can discover new knowledge by integrating across events (Lewis, Knoblich, & Poe, 2018). Brain-inspired architectures that leverage similar replay mechanisms show promise in improving the efficacy of reinforcement learning (Gershman & Daw, 2017; Botvinick et al., 2019) and the ability to learn new tasks without showing catastrophic forgetting of old tasks (Parisi et al., 2018). In the coming years, it will be important to know when and how neural replay occurs, and how the brain determines the memories that will be reactivated. This knowledge will be critical for understanding which aspects of replay are optimal for improving machine intelligence.

Active Learning and Curiosity. Research in AI is approaching the question of whether flexible problem solving can be improved by reinforcing acquisition of information in the face of uncertainty (Hutson, 2017). In humans, the intrinsic motivation to acquire knowledge is called curiosity, and curiosity has been shown to enhance learning and retention in humans (Kang et al., 2009; Gruber et al., 2014; Gruber & Ranganath, 2019; Stare et al., 2018) and machines (Hester & Stone, 2017). Recent research in cognitive neuroscience and psychology has begun to address this topic, but much remains to be learned about the factors that elicit curiosity and exploration, and the ways in which curiosity can affect learning (Gruber & Ranganath, 2019) and decision making (Bennett et al., 2016 ;Gershman, 2019). Incorporating curiosity can enhance the efficiency of learning in AI systems, by focusing exploration at points of high prediction error (Ecoffet et al., 2019). Studies of curiosity and exploration in infants have led to the development of robots that autonomously generate their own learning curricula, as shown in Figure 5 (Oudeyer and Smith, 2016). Going forward, it will be particularly important to research the factors that drive active information seeking and exploration in humans and, in the long term, to build machines that know how to actively generate questions in order to fill knowledge gaps about the external world and about the internal mental states of their partners.

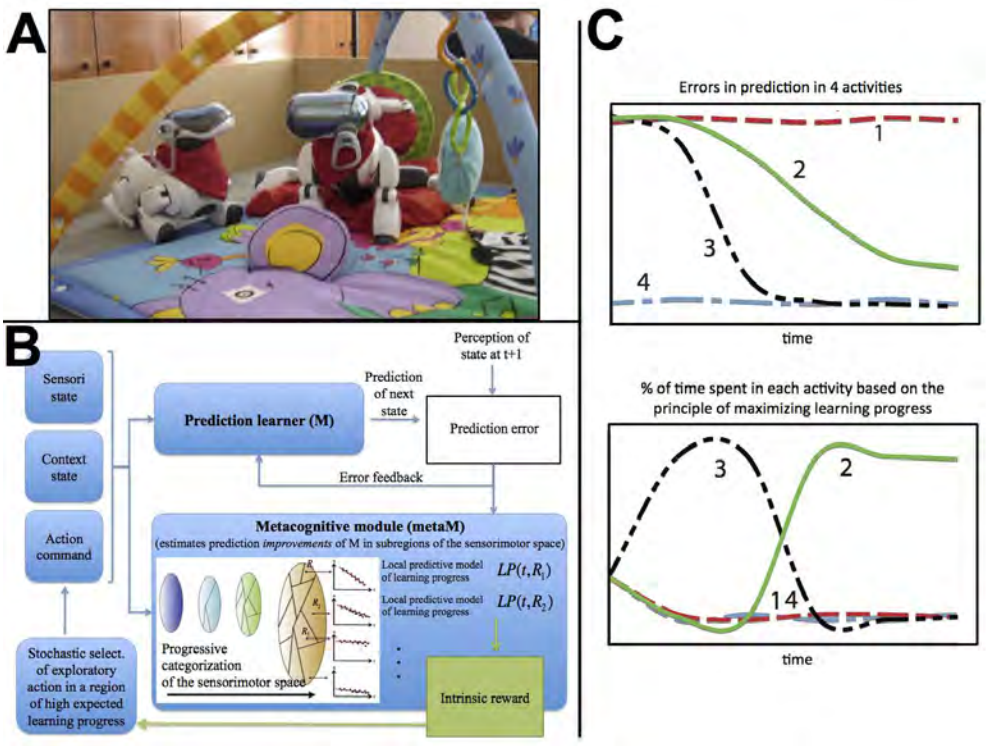


Figure 5. Curiosity-driven learning and intrinsically-driven exploration experiments with developmental robots. [Credit: Oudeyer and Smith, 2016]

Task Learning and Generalization: What allows humans to quickly learn tasks and generalize this learning to other tasks? One possibility is that humans "learn to learn" by acquiring knowledge at multiple levels of abstraction (Tenenbaum et al., 2011; Lake et al., 2017), and variations on this idea have begun to reap benefits in machine learning (Botvinick et al., 2019). Other work is investigating how learning of extended sequences of actions are represented in a manner that can support rapid learning of new tasks that involve overlapping actions (Desrochers et al., 2016 Frontiers Syst Neuro, Yang et al., 2019). Neuroscience research can guide future development in this area by characterizing how tasks can be represented in a manner that is compositional, such that information can be reused and recombined across tasks.

Task Control and Multi-Threading. Humans often must manage multiple goals at once and at different levels of abstraction, but we often fail at multitasking (Salvucci and Taatgen, 2008). The potential for human machine teaming could be strengthened if machines can aid humans in multitasking situations. Yet without an understanding of the constraints on the human operator and the nature of task representations that impact multitasking, it would be difficult to take advantage of this teaming relationship. Current work suggests that the specific way in which a task is represented can constrain how we multitask (Fusi et al., 2016; Musslick et al., 2016). Representing tasks in a simplified, abstract manner might improve generalization and novel task performance, but such representations are vulnerable to interference. In contrast, detailed, high dimensional representations have low interference but are difficult to generalize. This dichotomy might be solved by the adoption of multiple task representations by different brain networks. A near-term goal for research in this field will be to gain data on neural task representations and to develop models to understand who the strengths and weaknesses of these representations in multitasking situations.

Integrated Cognition in Complex, Dynamic Environments. Most of the research discussed above has been conducted in highly controlled laboratory environments using well-defined, temporally-delimited tasks. These laboratory paradigms do not necessarily capture behavior in environments where uncertainty is high, or in tasks that require multiple steps towards a high-level goal. Recent cognitive models have been directed towards real-world applications, such as consumer online choice behavior (Schulz et al., 2019), and comprehension of large-scale natural image concepts (Griffiths et al., 2016). Nonetheless, these models tend to focus on specific cognitive domains (e.g., memory, attention, language, decision making, etc.) in isolation, and there remains a large gap between these applications and the kinds of problems that require human machine teaming. It will be important to develop controlled, theory-driven research paradigms to understand temporally-extended task performance in naturalistic environments (Reggente et al., 2018). These data can be used to develop integrated cognitive and brain-inspired models in which different processes interact in a manner that can emulate human performance.

Decoding Neural Signals for Human Machine Interactions. Models and algorithms to decode neural signals show a high degree of promise for application in human machine teaming situations (Millan, 2019). Investigations of neural activity related to naturalistic perception, action, and cognition, along with analyses with computational models and machine learning tools can be used to develop interfaces and sensors for interactions with machines. For instance, Figure 6 illustrates how sensors can be used by machines to integrate neural activity patterns and behavioral observations, in order to infer the human's intentions and cognitive states (e.g., Chavarriaga et al., 2018). Although current approaches rely on data-driven decoding of neural signals, the robustness of brain machine interfaces could be improved by theory-driven models that incorporate an understanding of the neural representations that support high-level cognition (e.g., semantics, memory, cognitive control, etc.).

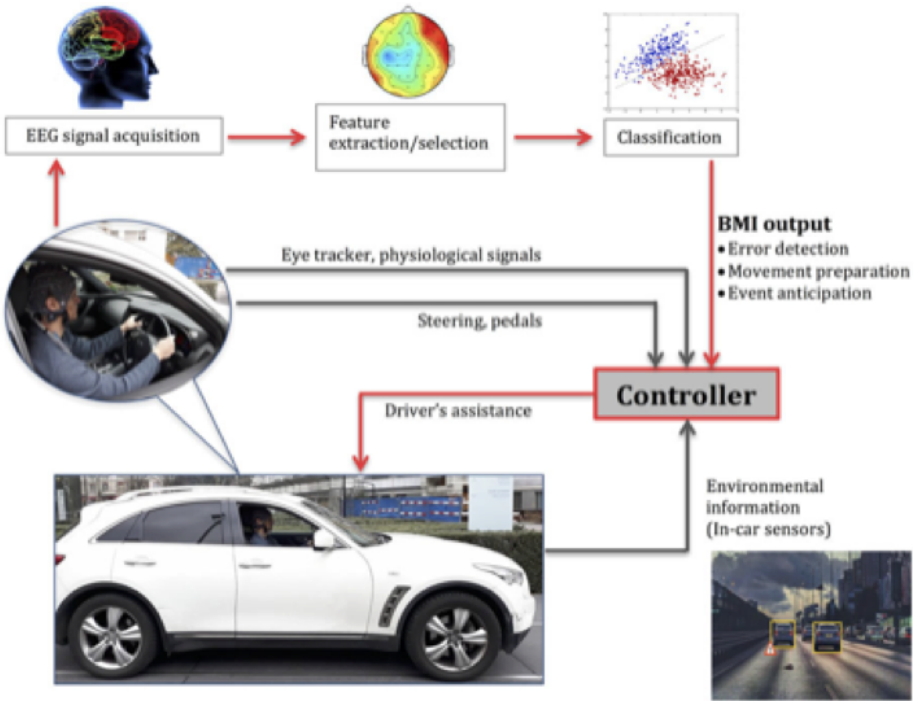


Figure 6. Brain-Machine Interface (BMI) for enhanced interaction: symbiotic car driving. The controller of the intelligent car takes into account environmental information, driver's actions and physiological signals, as well as their cognitive states inferred by the BMI from EEG (red arrows) to decide on the type and level of assistance it provides [Credit: Chavarriaga et al., 2018].

Research Trajectory for "Human Capabilities: Natural Intelligence"
We summarize the research trajectory for these advances in human capabilities: natural intelligence as:

Near-term (5-10 years)

- New experimental paradigms to understand natural language in real-world situations.
- Studies to understand human mechanisms for regulating learning rates and memory retrieval.
- Studies to understand the factors that elicit human curious states, the effects of curiosity on learning rates, and the tradeoffs between information seeking and goal-directed learning.
- Studies of neural representations of overlapping tasks and generalization to novel tasks.
- Research examining generalizability of models based on laboratory findings to investigations of perception, action, and cognition across extended timescales in dynamic environments.
- Development of minimally invasive sensors to collect detailed real-time measures of neural activity.

Long-term (10-20 years)

- Develop natural language models that represent real-world complexity and uncertainty.
- Identify the factors that determine memory replay and the content and timing of replay events; understand real-time interactions between systems that support episodic memory and systems that form general knowledge about events and situations.
- Develop computational models that can actively seek information through directed questioning of human teammates.
- Understand the fundamental mechanisms that allow humans to learn from single instances, generalize where appropriate, and show transfer to new situations
- Develop computational models of situations where multiple goals must be satisfied by sequencing and combining task representations.
- Theoretical frameworks (e.g., integrated cognitive architectures) to explain how multiple cognitive processes interact across extended timescales.
- Development of machine interfaces that use theory-driven computational neuroscience models to decode brain activity.

"Human Models of Machines" Research

Real-world Team Experiments. Advances in psychology and neuroscience are enhancing our understanding of human reasoning; however, many of the challenges in developing human models of machines identified earlier require research on the empirical side of human machine teaming research. Figure 3 showed a basic cycle of how improvements in different areas can facilitate research in others. Central to this is having experimental research in real-world, mixed human machine teams. Through these experiments we need to learn how different humans react to different mixtures or levels of cognitive capabilities in machine systems. What is the impact of a machine having adaptive models of the human, versus a fixed model, or even no model, and how important is it that the machine's reason mimics human reasoning?

Legible and Predictable Machine Behavior. An active area of research is on how to make the machines behavior "legible" and predictable so that a human can easily infer a machine's intentions and predict ahead of time what to expect of the robot (Dragan et al., 2015). To enhance legibility, a robot might exaggerate aspects of a motion in order to make it clear what its intentions are. This work has mainly focused on manipulation and movement tasks; however, these concepts could conceivably be applied to tasks that involve more abstract goals and tasks, using many different modalities. Research on social robots suggests that humans do respond to emotion expressions in machines for simple tasks and interactions, but further research is needed for complex tasks.

Explainable AI. Research on explainable AI has expanded, but mostly in areas related to machine learning. Research is needed on more dynamic explanations where the machine is reasoning about when and what it should explain so that it can help the human maintain a valid model of its teammate.

Trust. Building and maintenance of trust are rich scientific questions, requiring study in humans, machines, their models of each other, and their interactions, building on the research areas described above. What does a human need to know about the inner workings of a machine beyond its intentions in order to trust it? What is the impact of different machine transparency mechanisms and when is more information about a machine's inner working actually detrimental? To what degree are our models of others (human models of machines, and machine models of humans) shaped by the need to establish and maintain trust? How do we address and repair potential violations of trust? There is some existing research on conditions that establish human-human trust, particularly with respect to knowledge acquisition and problem-solving (Landrum et al., 2015); and this is an area where sophisticated computational models of cognition have been developed, but only for highly controlled laboratory tasks. Research is need on whether these results apply to complex real-world tasks.

Research Trajectory for "Human Models of Machines"
We summarize the research trajectory for these advances in human models of machines as:

Near-term (5-10 years)

- Design real-world experiments to examine how humans react to a variety of human machine teaming arrangements and variations in machine capabilities.
- Research how to make machine behavior legible and predictable beyond manipulation and motion.
- Develop machines that can dynamically explain their behavior for simple human machine teaming tasks.
- Studies in simple human machine teaming domains to examine the effect of variations in machine explanation, legibility, and related capabilities on human trust.

Far-term (10-20 years)

- Develop theories and corresponding implementations of legible and predictable machine behaviors.
- Design machines that can produce explanations as appropriate for the needs of their teammate for both simple and complex human machine teaming tasks.
- Develop new theories to describe the impact of variations and combinations of explanation, legibility, and related capabilities on human trust for both simple and complex human machine teaming domains.

"Machine Capabilities: Artificial Intelligence" Research

Research in AI, and specifically machine learning, has exploded in recent years, and as outlined in "The National Artificial Intelligence R&D Strategic Plan: 2019 Update" (Select Committee on Artificial Intelligence, 2019), it has become a national priority. However, we are still far from machine teammates that have the capabilities needed to fully support joint tasks at the same level as human teams. There are many opportunities for future advancements that can fill in these gaps across essentially all areas of AI. The key AI research areas relevant to human machine teaming include:

Perception. We are seeing continual advances across the spectrum of accessible data from multiple sensors, machine learning algorithms, and computing hardware that promise to provide improvements in robot perception. Research in activity recognition that builds on those capabilities promises to provide robots with the ability to build internal models of their environments and to predict and evaluate future states. Furthermore, research is needed on how cognitive processing can provide top-down influence to aid in perceptual processing.

Communication. Some of the necessary next steps in communication require "precision semantics", not just getting the gist of the communication, but grounding the meaning of a communication to the specific environment, and in the specific context (prior interactions) of the utterance. This requires progress in ambiguity resolution ("The council denied the protesters a permit because "they" feared/advocated violence); contextual threshold resolution ("tall", "dangerous" relative to what comparison class?); and speech act recognition (request, command, etc.).

Modeling the Environment and Itself. One somewhat unintuitive direction for future research is to build on the continual enhancement and growth of detailed computational models of the world that are developed for large scale computer games. By taking advantage of the underlying game engines and the accompanying physics models, machine agents could have real-time models that can be used to reason about, and even predict the world. This approach has been explored in robot architectures and even cognitive science research to model human spatial reasoning (Battaglia, et al., 2013).

Reasoning, Problem Solving, Planning, Task Expertise. Significant progress has been made over the last several decades to create effective AI systems, and integrating them with other components. However, additional research is need outside of standard AI research areas to support teamwork, such as perspective-taking, joint attention, and advancements in computational theories of cooperation and coordination (e.g., Kleiman-Weiner et al., 2016).

Learning. Promising research is developing on machines that learn and adapt their behavior directly from human instruction, including imitation, demonstration, and language. Much of this work is currently pursued by researchers in Human-Robot Interaction, as well as Interactive Task Learning, where humans teach AI systems new tasks through demonstration and language (Laird et al., 2017a; Gluck & Laird, 2019).

Integrated Architectures. Integrated cognitive architectures have the potential to provide frameworks for developing and integrating many, if not all of the capabilities required for machine teammates. Over the last thirty years, there has been continued development (Kotseruba and Tsotsos, 2018) and some of these, such as ACT-R and Soar, have been used for both modeling human behavior as well as controlling robotic systems (Mininger and Laird, 2018). Thus, they have the potential for not only being used for the reasoning of the machine, but also the machine models of humans it is teaming with. Recently there has been a drive for consensus on the overall abstract functional structure of the mind, leading to the Common Model of Cognition (Laird et al., 2017b), which provides an opportunity for integrating research across cognitive science, AI, and cognitive neuroscience.

Research Trajectory for "Machine Capabilities: Artificial Intelligence"

We summarize the research trajectory for these advances in machine capabilities: artificial intelligence as:

Near-term (5-10 years)

- Research on how cognitive processing can aid perceptual processing and enable machines that track humans by accurately recognizing movement and navigation behavior throughout different types of tasks.
- Communication studies to better understand ambiguity resolution, contextual threshold resolution, and speech act recognition will lead to machines that carry on extended dialogs about tasks, not limited to single interactions.
- Research on human-robot interaction and interactive task learning where humans teach AI systems new tasks through demonstration and language and correct the machine's behavior.
- Research and development of computational models for perspective-taking, joint attention, and theories of cooperation and coordination will be mapped onto a common model of cognition.

- New cognitive architectures that integrate knowledge from cognitive science, AI, and cognitive neuroscience. Machine teammates will use these cognitive architectures both for reasoning and learning, but also to model their human teammates across multiple tasks.

Far-term (10-20 years)

- Design machines with the ability to build internal models of their environments and predict future states.
- Continued research on cognitive and perceptual processing to build machines that can communicate with contextual understanding of the specific environment and the specific context of the utterance. This will lead to machines that can dynamically track humans throughout different types of tasks that require different types of interactions with the world.
- Design machine agents with real-time models that can be used to reason about, and even predict, the world around it.
- Develop the theoretical framework to enable machine reasoning, problem solving, planning, task expertise needed for human machine teaming.
- Build machines that can learn and adapt their behavior directly from human instruction, including imitation, demonstration, and language.
- Develop integrated cognitive architectures that create the reasoning of the machine and also provide the machine models of human teammates.

"Machine Models of Humans" Research

As illustrated in Figure 2, a critical component of a machine teammate is its ability to reason about its human teammates. This is a rich area for future research in these five areas:

Understanding Which Aspects of Human Behavior Need to be Modeled. Continual progress is being made in human modeling and prediction, including reasoning about (and predicting) the beliefs, desires, and intentions of other agents. Ongoing applied social science and human-robot interaction research attempts to determine which aspects of human-human teaming are necessary to support effective and robust human machine teaming. Research in this area can clarify whether we need high-fidelity models of human behavior, or whether abstract, approximate models are sufficient for a machine to effectively team with a human. The expectation is that there might not be a single, one-size fits all approach to modeling, and extensive research is needed to fill out our understanding of what types of models are needed in different teaming situations.

Understanding Human Perceptual Abilities. The limited abilities of human perception are well understood (see for example Frisby & Stone, 2010). Incorporating an understanding of these abilities will be important for engineering effective machine partners. For example, a machine that is referring to some object that is difficult to see or locate for a human partner must take this into account and possibly provide additional information.

Understanding Human Motor Control Abilities. Human motor control abilities are in many ways much more advanced than many robots. Humans are more dexterous, stable and often faster. On the other hand, humans may be weaker than some robots, and may fatigue more quickly. Knowledge of these abilities should be an important component of the human models represented by machines.

Understanding Human Reasoning and Planning Abilities. In order to effectively team with humans in complex sequential decision problems, machines must have some understanding of how humans represent and solve such problems. Progress has been made on answering this question for well-constrained problems in cognitive neuroscience, which has pointed towards multiple algorithmic strategies, possibly implemented as separate neural systems (Kool, Cushman & Gershman, 2018). For example, people seem to use model-based "goal-directed" and model-free "habitual" action selection under different circumstances. People even understand that other individuals will be more goal-directed or habitual under different circumstances (Gershman et al., 2016). Models of human sequential decision making have been developed with sufficient computational formalization that they could in principle be incorporated into a machine's model of a human decision maker. One limiting factor is that these models have primarily been studied in small-scale laboratory settings, so their generalizability to complex problems is still untested. For more complex tasks, cognitive modeling based on cognitive architectures such as ACT-R have been used, including tasks involving teaming.

Building Dynamic Models. Even if a machine has an initial, somewhat generic model of a human, it must customize it to the specific teammate and continually update and revise during performance. Research in intelligent tutors has been successful in creating personalized models of individuals for specific tasks and then tracking those individuals (Leyzberg et al., 2018). Research is needed in taking these and other techniques and scaling them up to human machine teaming applications.

Research Trajectory for Machine Models of Humans

We summarize the research trajectory for these advances in machine models of humans as:

Short-term (5-10 years)

- Design real-world teaming experiments for specific tasks to test which levels and types of human behavior, human perception, human motor control, and human reasoning and planning abilities are needed for effective human machine teaming.
- Research in building personalized models of human teammates for specific tasks that tracks and updates during the task.

Far-term (10-20 years)

- Develop theories for what levels and types of human modeling are needed for effective teaming.
- Develop theory for what levels of human perception are needed on teaming tasks.
- Build dynamic models of human teammates that extend across a range of tasks.

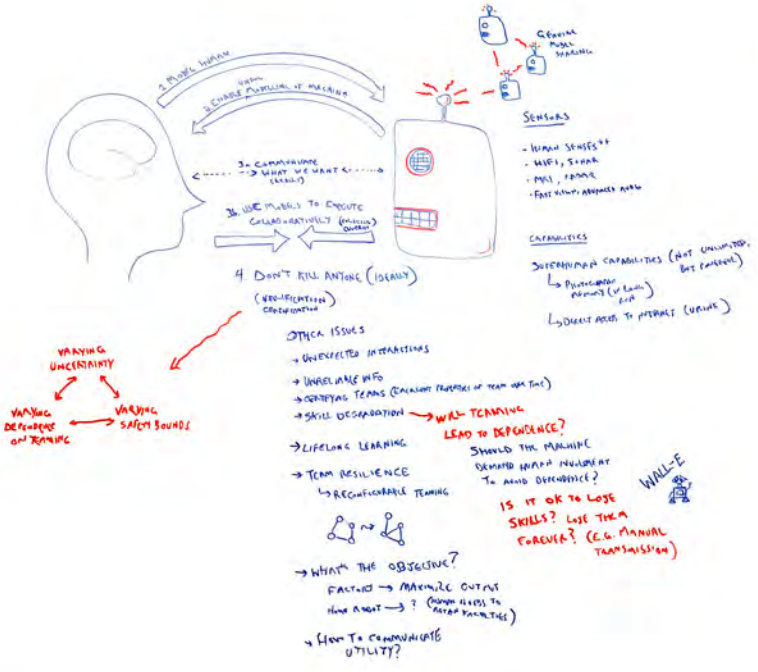
Accelerating Progress in Human Machine Teaming

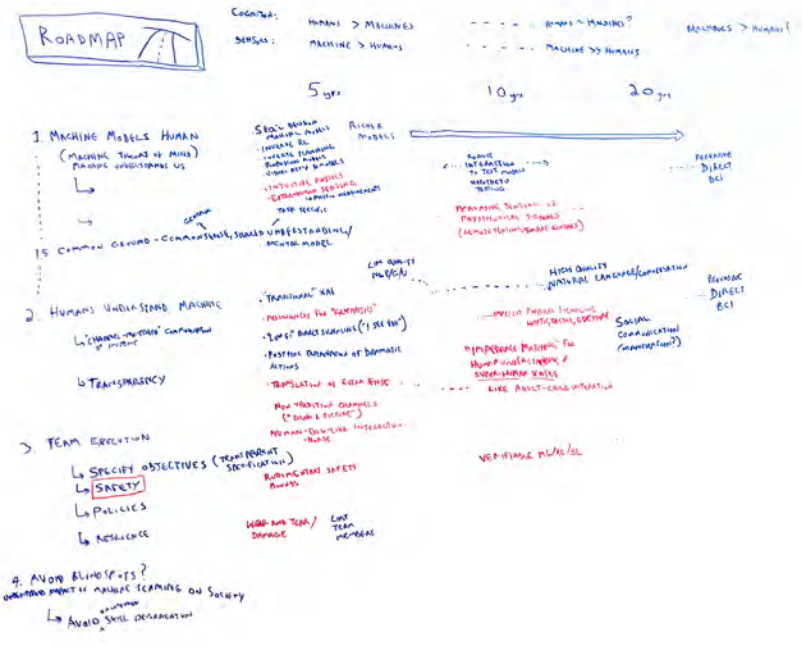
In addition to discussing specific opportunities and challenges, workgroup discussions also touched on more general factors that are essential to progress in human machine teaming. One key theme to emerge from the workshop is the meta-challenge of integrating research approaches from across disciplines like AI, cognitive psychology, cognitive neuroscience, robotics, and human-robot interaction to solve the big research challenges of human machine teaming described above. These different disciplines have their own terminology, their own measures of progress and success, and invariably live in different departments with independent meetings and communities. Overcoming these barriers will be important to bring about progress in a multidisciplinary research field like human machine teaming. A combined approach leveraging basic behavioral and neuroscience research, computational architectures and brain-inspired computational models, and implementations in autonomous machines could yield breakthroughs in both basic science and applications of human machine teaming.

One way forward is to facilitate interdisciplinary communication. Workshops and small conferences can be an ideal venue to encourage researchers to identify connections with researchers in other disciplines. Such workshops can be followed up with research investments to support multidisciplinary collaborations on human machine teaming research. It might be useful to consider seed funding for initiating multidisciplinary collaborations, as well as large-scale multi-investigator grants to support coordinated efforts to tackle ambitious questions.

Some workshop discussants also advocated for the development of open datasets and corpuses to facilitate computational modeling and development of intelligent agents. Open datasets are widely used in many areas of computer science and engineering, such as computer vision and audio processing. Discussants pointed out that the currently available datasets do not capture the complexity of real-world situations that pose challenges to human machine teams. Additionally, an open knowledge database would be useful in equipping intelligent agents with the kind of world knowledge that humans expect from teammates. For instance, in their "Artificial Intelligence Roadmap" (Gill and Selman, 2019), the Computing Community Consortium recommended the creation of "An Open Knowledge Network (OKN) with large amounts of knowledge about entities, concepts, and relationships in the world." The OKN would be similar to proprietary databases created by large corporations like Google and Microsoft, enabling artificial systems to acquire links between objects, entities, relationships, contexts, and situations. Developing an OKN would require considerable data mining and basic research, and it would require resources for maintenance and updating, as described in the Open Knowledge Network Report (Executive Office of the President of the United States, Committee on Science and Technology, 2018). This kind of database could be of significant value for stimulating new basic research (e.g., development of computational models of natural intelligence, research on neural representation of complex world knowledge, etc.) and for engineering of intelligent systems (e.g., improving natural language, context-sensitive behavior, and "common sense" inferences).

Finally, human machine teaming might also be facilitated by specifying "use cases," that is, specific examples of where human machine teaming would have a big impact. Use cases can be helpful because a great deal of research on human machine teaming is based on scientific assumptions that may be overly broad or out of touch with what is actually required in particular situations. Given the very broad range of applications for human machine teams, approaches that are sufficient for one application may be inappropriate for others. By providing a detailed task analysis, researchers can focus on the particular challenges to development of intelligent teammates in a particular situation. Building along these lines, research competitions and prizes can provide added motivation for teams to tackle particular human machine teaming challenges.





Conclusion

This report summarizes a bold vision for the potential of intelligent systems to become teammates that can communicate with both human and machine partners, coordinate activities, signal intent to support share goals, and represent teammates' goals and situations. The workshop discussants outlined a framework for the next generation of systems and outlined research challenges and opportunities. Although the goals that emerged from the workshop are ambitious, the participants were optimistic of the potential for human machine teaming to become reality over the next twenty years.

References

Arad, A and Feige, K. (Producer), and Favreau, J. (Director). (2008). Iron Man [Motion picture]. *United States: Marvel Studios.*

Baron-Cohen, S. (1995). Mindblindness: An Essay on Autism and Theory of Mind.

Bartlett, C., & Cooke, N. (2015). Human-Robot Teaming in Urban Search and Rescue. *59th International Annual Meeting of the Human Factors and Ergonomics Society, HFES 2014,* 59(1), 250-254.

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America,* 110(45), 18327-18332.

Bellmund, J., Gärdenfors, P., Moser, E., & Doeller, C. (2018). Navigating cognition: Spatial codes for human thinking. *Science,* 362(6415).

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic Valuation of Information in Decision Making under Uncertainty. *PLOS Computational Biology,* 12(7).

Botvinick, M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences,* 12(5), 201-208.

Botvinick, M., Ritter, S., Wang, J., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences,* 23(5), 408-422.

Brown-Schmidt, S., & Duff, M. (2016). Memory and Common Ground Processes in Language Use. *Topics in Cognitive Science,* 8(4), 722-736.

Chavarriaga, R., Uscumlic, M., Zhang, H., Khaliliardali, Z., Aydarkhanov, R., Saeedi, S., . . . Millan, J. (2018). Decoding Neural Correlates of Cognitive States to Enhance Driving Experience. *IEEE Transactions on Emerging Topics in Computational Intelligence,* 2(4), 288-297.

Clark, H. H. (1992). Arenas of language use. Stanford, CA, US: Center for the Study of Language & Information; Chicago, IL, US: University of Chicago Press.

Cooke, N., Gorman, J., Myers, C., & Duran, J. (2013). Interactive team cognition. *Cognitive Science,* 37(2), 255-285.

Executive Office of the President of the United States. (2018). Open Knowledge Network. Report by the Big Data Interagency Working Group, Washington DC.

Executive Office of the President of the United States. (June 2019). The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update. Prepared by the Select Committee on Artificial Intelligence Washington DC.

Desrochers, T., Burk, D., Badre, D., & Sheinberg, D. (2016). The monitoring and control of task sequences in human and non-human primates. *Frontiers in Systems Neuroscience,* 9, 185-185.

Djukic, M., Adams, J., Fulmer, T., Szyld, D., Lee, S., Oh, S., & Triola, M. (2015). E-Learning with virtual teammates: A novel approach to interprofessional education. *Journal of Interprofessional Care,* 29(5), 476-482.

Dragan, A., Bauman, S., Forlizzi, J., & Srinivasa, S. (2015). Effects of Robot Motion on Human-Robot Collaboration. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction,* (pp. 51-58).

Duff, M., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience,* 6, 69-69.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology,* 41(1), 87-100.

Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K., & Clune, J. (2019). Go-Explore: a New Approach for Hard-Exploration Problems. *arXiv preprint arXiv:1901.10995.*

Ekstrom, A., & Ranganath, C. (2018). Space, Time and Episodic Memory: the Hippocampus

is all over the Cognitive Map. *Hippocampus,* 28(9), 680-687.

Frisby, J., & Stone, J. (2010). Seeing, Second Edition: The Computational Approach to Biological Vision.

Fusi, S., Miller, E., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66-74.

Gershman, S. (2019). Uncertainty and exploration. *Decision,* 6(3), 277-286.

Gershman, S., & Daw, N. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology,* 68(1), 101-128.

Gershman, S., Gerstenberg, T., Baker, C., & Cushman, F. (2016). Plans, Habits, and Theory of Mind. *PLOS ONE,* 11(9).

Gil, Y., & Selman, B. (2019). A 20-Year Community Roadmap for Artificial Intelligence Research in the US. *arXiv preprint arXiv:1908.02624*.

Gluck, K. A., & Laird, J. E. (2019). *Interactive Task Learning: Humans, Robots, and Agents Acquiring New Tasks through Natural Interactions.* Boston, MA: MIT Press.

Gombolay, M., Bair, A., Huang, C., & Shah, J. (2017). Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences. *The International Journal of Robotics Research,* 36, 597-617.

Goodman, N., & Frank, M. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences,* 20(11), 818-829.

Graves, A., Wayne, G., & Danihelka, I. (2014). Neural Turing Machines. *arXiv preprint arXiv:1410.5401*.

Grice, H. (1975). Logic and conversation. *Syntax and Semantics,* 3, 41-58.

Griffiths, T., Abbott, J., & Hsu, A. (2016). Exploring Human Cognition Using Large Image Databases. *Topics in Cognitive Science,* 8(3), 569-588.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies,* 8(3), 483-500.

Gruber, M., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory. *in press.*

Gruber, M., Gelman, B., & Ranganath, C. (2014). States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit. *Neuron,* 84(2), 486-496.

Hard, B., Tversky, B., & Lang, D. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition,* 34(6), 1221-1235.

Hayes, B., & Scassellati, B. (2016). Autonomously constructing hierarchical task networks for planning and human-robot collaboration. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, (pp. 5469-5476).

Heard, J., Heald, R., Harriott, C., & Adams, J. (2019). A Diagnostic Human Workload Assessment Algorithm for Collaborative and Supervisory Human—Robot Teams. *ACM Transactions on Human-Robot Interaction*, 8(2), 1-30.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology,* 57(2), 243-259.

Hester, T., & Stone, P. (2017). Intrinsically motivated model learning for developing curious robots. *Artificial Intelligence*, 247, 170-186.

Hill, J., Chen, J., Gratch, J., Rosenbloom, P., & Tambe, M. (1998). Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft.

Hutson, M. (2017, May 31). *Scientists imbue robots with curiosity. doi:10.1126/science.aan6916*

Jones, R., Laird, J., Nielsen, P., Coulter, K., Kenny, P., & Koss, F. (1999). Automated Intelligent Pilots for Combat Flight Simulation. *Ai Magazine*, 20(1), 27-41.

Kang, M., Hsu, M., Krajbich, I., Loewenstein, G., McClure, S., Wang, J.-y., & Camerer, C. (2009). The Wick in the Candle of Learning Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science*, 20(8), 963-973.

Kleiman-Weiner, M., Ho, M., Austerweil, J., Littman, M., & Tenenbaum, J. (2016). Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. *Cognitive Science*.

Kool, W., Cushman, F., & Gershman, S. (2018). Competition and Cooperation Between Multiple Reinforcement Learning Systems. 153-178.

Kotseruba, I., & Tsotsos, J. (2018). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 1-78.

Laird, J. (2012). *The Soar Cognitive Architecture*.

Laird, J., Gluck, K., Anderson, J., Forbus, K., Jenkins, O., Lebiere, C., . . . Kirk, J. (2017). Interactive Task Learning. *IEEE Intelligent Systems*, 32(4), 6-21.

Laird, J., Lebiere, C., & Rosenbloom, P. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, 38(4), 13-26.

Lake, B., Ullman, T., Tenenbaum, J., & Gershman, S. (2017). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, 40, 1-101.

Landrum, A., Eaves, B., & Shafto, P. (2015). Learning to trust and trusting to learn: a theoretical framework. *Trends in Cognitive Sciences*, 19(3), 109-111.

Lasota, P., Fong, T., & Shah, J. (2017). A Survey of Methods for Safe Human-Robot Interaction. *Foundations and Trends in Robotics*, 5(4), 261-349.

Lewis, P., Knoblich, G., & Poe, G. (2018). How Memory Replay in Sleep Boosts Creative Problem-Solving. *Trends in Cognitive Sciences*, 22(6), 491-503.

Leyzberg, D., Ramachandran, A., & Scassellati, B. (2018). The Effect of Personalization in Longer-Term Robot Tutoring. *ACM Transactions on Human-Robot Interaction (THRI) archive*, 7(3), 19.

McNeese, N., Demir, M., Cooke, N., & Myers, C. (2018). Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming:. *Human Factors*, 60(2), 262-273.

Millán, J. (2018). The human-computer connection: an overview of brain-computer interfaces. Mètode Science Studies Journal: *Annual Review(9)*, 134-141.

Mininger, A., & Laird, J. (2018). Interactively Learning a Blend of Goal-Based and Procedural Tasks. *AAAI-18 AAAI Conference on Artificial Intelligence*, (pp. 1487-1494).

Musslick, S., Dey, B., Özcimder, K., Patwary, M., Willke, T., & Cohen, J. (2016). Controlled vs. Automatic Processing: A Graph-Theoretic Approach to the Analysis of Serial vs. Parallel Processing in Neural Network Architectures. *Cognitive Science*.

O'Reilly, R., Bhattacharyya, R., Howard, M., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science*, 38(6), 1229-1248.

O'Reilly, R., Wyatte, D., & Rohrlich, J. (2014). Learning Through Time in the Thalamocortical Loops. *arXiv preprint arXiv:1407.3432*.

Oudeyer, P.-Y., & Smith, L. (2016). How Evolution May Work Through Curiosity-Driven Developmental Process. *Topics in Cognitive Science*, 8(2), 492-502.

Parisi, G., Tani, J., Weber, C., & Wermter, S. (2018). Lifelong Learning of Spatiotemporal Representations With Dual-Memory Recurrent Self-Organization. *Frontiers in Neurorobotics*, 12.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, 1(4), 515-526.

Reggente, N., Essoe, J.-Y., Aghajan, Z., Tavakoli, A., McGuire, J., Suthana, N., & Rissman, J. (2018). Enhancing the Ecological Validity of fMRI Memory Research Using Virtual Reality. *Frontiers in Neuroscience*, 12, 408.

Richmond, L., & Zacks, J. (2017). Constructing Experience: Event Models from Perception to Action. *Trends in Cognitive Sciences*, 21(12), 962-980.

Rickel, J., & Johnson, W. (1998). STEVE: A Pedagogical Agent for Virtual Reality. *Autonomous Agents and Multi-Agent Systems*, 332-333.

Salas, E., Cooke, N., & Rosen, M. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors*, 50(3), 540-547.

Salas, E., Dickinson, T., Converse, S., & Tannenbaum, S. (1992). Toward an understanding of team performance and training.

Salvucci, D., & Taatgen, N. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101-130.

Schulz, E., Bhui, R., Love, B., Brier, B., Todd, M., & Gershman, S. (2019). Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences of the United States of America*, 116(28), 13903-13908.

Stachenfeld, K., Botvinick, M., & Gershman, S. (2017). The hippocampus as a predictive map. *Nature Neuroscience,* 20(11), 1643-1653.

Stare, C., Gruber, M., Nadel, L., Ranganath, C., & Gómez, R. (2018). Curiosity-driven memory enhancement persists over time but does not benefit from post-learning sleep. *Cognitive Neuroscience*, 9, 100-115.

Sukthankar, G., Geib, C., Bui, H., Pynadath, D., & Goldman, R. (2014). Plan, Activity, and Intent Recognition: Theory and Practice. Plan, Activity, and Intent Recognition: *Theory and Practice 1st* , 424-424.

Tambe, M. (1997). Agent architectures for flexible, practical teamwork. *AAAI'97/IAAI'97 Proceedings of the fourteenth national conference on artificial intelligence and ninth conference on Innovative applications of artificial intelligence*, (pp. 22-28).

Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, 331(6022), 1279-1285.

Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003). Negotiation over tasks in hybrid human-agent teams for simulation-based training. *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, (pp. 441-448).

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. *New York, NY: Academic Press.*

Wynne, K., & Lyons, J. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19(3), 353-374.

Yang, G., Joglekar, M., Song, H., Newsome, W., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22(2), 297-306.

Appendix I – Workshop Attendees

Co-Chairs

John Laird	University of Michigan
Charan Ranganath	University of California, Davis
Sam Gershman	Harvard University

Academic/Industry Participants

Julie A. Adams	Oregon State University
David Badre	Brown University
Joyce Chai	University of Michigan
David Cox	IBM Research
Judith Degen	Stanford University
Jose del R. Millan	Ecole Polytech Fed Lausanne
Josh Greene	Harvard University
Christian Lebiere	Carnegie Mellon University
Bradley Love	University College of London
Yuko Munakata	University of California, Davis
Aude Oliva	MIT
Randy O'Reilly	University of California, Davis
Brian Scassellati	Yale University
Matthias Scheutz	Tufts University
Patrick Shafto	Rutgers University

DoD Participants

Paul Bello	Naval Research Laboratory
Joseph B. Lyons	US Air Force Res Laboratory
Greg Trafton	Naval Research Laboratory

Observers

Tamara Chelette	US Air Force Research Laboratory
Zola Donovan	US Air Force Research Laboratory
Theresa Fensch	MITRE
Mary Anne Fields	US Army Research Laboratory
Hal Greenwald	Air Force Office of Scientific Research
Laura Kallal	OUSD(R&E) Human Systems Directorate
Nicholas Kasdaglis	MITRE
Rushyannah Killens-Cade	AAAS S&T Policy Fellow OUSD(R&E)
Michael La Fiandra	US Army Research Laboratory
Bindu Nair	OUSD(R&E)
Dan Osborn	OUSD(R&E)
Edward Palazzolo	US Army Research Laboratory
Shanni Silberberg	OUSD(R&E)
Marc Steinberg	Office of Naval Research Science & Technology
David Stout	OUSD(R&E)
Cliff Wang	US Army Research Laboratory

Appendix II – Workshop Participant Biographies

Julie A. Adams, Professor
Oregon State University
http://research.engr.oregonstate.edu/hmtl/
julie.a.adams@oregonstate.edu
Julie A. Adams is Professor of Computer Science in the School of Electrical Engineering and Computer Science and the Associate Director of Research in the Collaborative Robotics and Intelligent Systems Institute (CoRIS) at Oregon State University. Dr. Adams was the founder of the Human machine Teaming Laboratory at Vanderbilt University, prior to moving the laboratory to Oregon State. Adams has worked in the area of human machine teaming for almost thirty years. Throughout her career she has focused on human interaction with unmanned systems, but also focused on manned civilian and military aircraft at Honeywell, Inc. and commercial, consumer and industrial systems at the Eastman Kodak Company. Her research, which is grounded in robotics applications for domains such as first response, archaeology, oceanography, the national airspace and the U.S. military, focuses on distributed artificial intelligence, swarms, robotics and human machine teaming. Adams received her M.S. and Ph.D. degrees in Computer and Information Sciences from the University of Pennsylvania and her B.S. in Computer Science and B.B.E. in Accounting from Siena College.

David Badre, Professor
Brown University
https://sites.brown.edu/badrelab/
david_badre@brown.edu
David Badre is Professor in the Department of Cognitive, Linguistic and Psychological Sciences at Brown University. Dr. Badre received his B.S. from the University of Michigan in 2000, his Ph.D. from the Department of Brain and Cognitive Sciences at MIT in 2005 and held a postdoctoral fellowship at the UC, Berkeley. He is also an affiliate of the Carney Institute for Brain Science at Brown University. His research focuses on the cognitive neuroscience of executive function.

Dr. Badre serves on the editorial boards of Psychological Science, Cognitive Science, and Behavioral Neuroscience. He served as Section Editor covering "Executive Function and Cognitive Control" for Neuropsychologia until 2017. Presently, he serves on the Board of Reviewing Editors for the journal eLife, and he is a standing member of the Cognition and Perception study section of NIH. His research is supported by NINDS and NIMH at the NIH, and through the Office of Naval Research. His work has been recognized by several awards, including an Alfred P. Sloan Foundation Fellowship in Neuroscience, a James S. McDonnell Scholar Award in Understanding Human Cognition, and the Cognitive Neuroscience Society Young Investigator Award.

Paul Bello, Section Head
Naval Research Laboratory
https://www.nrl.navy.mil/itd/aic/InteractiveSystems
paul.bello@nrl.navy.mil
Paul Bello is the Section Head of the Integrated Cognitive Systems Section at the US Naval Research Laboratory. Dr. Bello leads a multidisciplinary team of seven researchers exploring a variety of topics at the intersection of philosophy, psychology, and artificial intelligence. Along with Dr. Will Bridewell, he developed and now co-directs the ARCADIA research program: an ambitious attempt to computationally explore the connections between attention, consciousness, and agency. Bello received dual Bachelor of Science degrees in Computer and Systems Engineering and Philosophy from Rensselaer Polytechnic Institute in 1999. He stayed on and earned a Master of Science degree in Computer Science in 2001 and earned his Ph.D. in Cognitive Science under the supervision of Professor Selmer Bringsjord. His graduate work was among the earliest explorations of what has now come to be called "AI Ethics" (or Machine Ethics), and he remains active in pursing questions of this nature within the ARCADIA framework.

Joyce Chai, Professor
University of Michigan
http://web.eecs.umich.edu/~chaijy/
chaijy@umich.edu
Joyce Chai is a Professor in the Department of Electrical Engineering and Computer Science at the University of Michigan. Previously, Dr. Chai held the position of Professor at Michigan State University and was a Research Staff Member at IBM T. J. Watson Research Center. Dr. Chai's research interests include natural language processing, situated dialogue agents, human-robot communication, and artificial intelligence. Her recent work is focused on grounded language processing to facilitate natural communication with robots and other artificial agents. She received a National Science Foundation CAREER Award in 2004, the Best Long Paper Award from the Annual Meeting of Association of Computational Linguistics (ACL) in 2010, and the William Beal Outstanding Faculty Award from MSU in 2018. She holds a Ph.D. in Computer Science from Duke University.

David Cox, Director
IBM Research, MIT-IBM Watson AI Lab
https://researcher.watson.ibm.com/researcher/view.php?person=ibm-David.D.Cox
David.D.Cox@ibm.com
David Cox is the IBM Director of the MIT-IBM Watson AI Lab, a first of its kind industry-academic collaboration between IBM and MIT, focused on fundamental research in artificial intelligence. The Lab was founded with a $240m, 10 year commitment from IBM and brings together researchers at IBM with faculty at MIT to tackle hard problems at the vanguard of AI.

Prior to joining IBM, Dr. Cox was the John L. Loeb Associate Professor of the Natural Sciences and of Engineering and Applied Sciences at Harvard University, with appointments in Computer Science, the Department of Molecular and Cellular Biology and the Center for Brain Science. David's ongoing research is primarily focused on bringing insights from neuroscience into machine learning and computer vision research. His work has spanned a variety of disciplines, from imaging and electrophysiology experiments in living brains, to the development of machine learning and computer vision methods, to applied machine learning and high performance computing methods.

Judith Degen, Assistant Professor
Stanford University
https://sites.google.com/site/judithdegen/
jdegen@stanford.edu
Judith Degen is Assistant Professor of Linguistics at Stanford University. Trained as a cognitive scientist at the University of Rochester and Stanford University, Judith is interested in the inference processes involved in language production and comprehension – how do speakers choose an utterance to convey an intended meaning? How do listeners arrive at interpretations that are often much richer and more detailed than the literal meaning provided by a sentence? She employs a combination of linguistic analysis, behavioral methods, corpus methods, and computational models to develop explicit theories of these processes and test them against behavioral data.

Jose del R. Millán, Professor
Ecole Polytech Fed Lausanne/ UT Austin
https://cnp.epfl.ch/laB.S./millanlab/
jose.millan@epfl.ch
José del R. Millán is the Defitech Chair at the École Polytechnique Fédérale de Lausanne (EPFL) and directs the Brain-Machine Interface Laboratory. Dr. del R. Millán joined École Polytechnique Fédérale de Lausanne (EPFL) in 2009 to help establish the Center for Neuroprosthetics. In September 2019 he will start a new laboratory at the University of Texas at Austin as the Motorola Regent Chair #2 in the Department of Electrical & Computer Engineering and in the Department of Neurology.

Dr. del R. Millán received a Ph.D. in computer science from the Technical University of Catalonia, Barcelona, in 1992. He has made several seminal contributions to the field of brain-machine interfaces (BMI), especially based on electroencephalogram (EEG) signals. Most of his achievements revolve around the design of brain-controlled robots. He has received several recognitions for these seminal and pioneering achievements, notably the IEEE-SMC Nobert Wiener Award in 2011 and elevation to IEEE Fellow in 2017. During the last years Dr. Millán is prioritizing the translation of BMI to end-users suffering from motor disabilities. As an example of this endeavour, his team won the first Cybathlon BMI race in October 2016. Together with his team, he is also designing BMI technology to offer new interaction modalities for able-bodied people.

Sam Gershman, Associate Professor
Harvard University
http://gershmanlab.webfactional.com/index.html
gershman@fas.harvard.edu
Samuel Gershman is Associate Professor in the Department of Psychology and Center for Brain Science at Harvard University. Dr. Gershman received his B.A. in Neuroscience and Behavior from Columbia University in 2007 and his Ph.D. in Psychology and Neuroscience from Princeton University in 2013. From 2013-2015 he was a postdoctoral fellow in the Department of Brain and Cognitive Sciences at MIT. His research aims to understand how richly structured knowledge about the environment is acquired, and how this knowledge aids adaptive behavior. The lab uses a combination of behavioral, neuroimaging and computational techniques to pursue these questions.

Joshua Greene, Professor
Harvard University
http://www.joshua-greene.net/
jgreene@wjh.harvard.edu
Joshua D. Greene is Professor of Psychology and a member of the Center for Brain Science faculty at Harvard University. His research interests cluster around the intersection of psychology, neuroscience, and philosophy. His early work focused on the cognitive neuroscience of moral judgment and the interplay between emotion and reason in moral dilemmas. More recent work focuses on critical features of individual and collective intelligence. His current neuroscientific research examines how the brain combines concepts to form thoughts and how thoughts are manipulated in reasoning and imagination. His current behavioral research examines strategies for improved social decision-making and the alleviation of intergroup conflict. Other interests include effective altruism and the social implications of advancing artificial intelligence. He is the author of Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.

John Laird, Professor
University of Michigan
https://soar.eecs.umich.edu/
laird@umich.edu
John Laird is the John L. Tishman Professor of Engineering at the University of Michigan, where he has been since 1986. He received his Ph.D. in Computer Science from Carnegie Mellon University in 1983 working with Allen Newell. From 1984 to 1986, he was a member of research staff at Xerox Palo Alto Research Center. He is one of the original developers of the Soar architecture and leads its continued evolution. He was a founder of Soar Technology, Inc. and he is a Fellow of AAAI, AAAS, ACM, and the Cognitive Science Society. With Paul Rosenbloom, he is the winner of the 2018 Herbert A. Simon Prize for Advances in Cognitive Systems.

Christian Lebiere, Research Faculty
Carnegie Mellon University
http://www.psy.cmu.edu/new_old_backup/people/lebiere.html
cl@cmu.edu
Christian Lebiere is Research Faculty in the Psychology Department at Carnegie Mellon University. Dr. Lebier directs the FMS Cognitive Modeling Group. He received his B.S. in Computer Science from the University of Liege (Belgium) and his M.S. and Ph.D. from the School of Computer Science at Carnegie Mellon University. During his graduate career, he studied connectionist models and was the co-developer of the Cascade-Correlation neural network learning algorithm that was a precursor of deep learning algorithms. Since 1991, he has worked on the development of the ACT-R cognitive architecture and was co-author with John Anderson of the 1998 book "The Atomic Components of Thought". The ACT-R cognitive architecture has been used by a large international community of researchers in over a thousand publications in the fields of Cognitive Science and Artificial Intelligence.

Dr. Lebiere is a founding member of the Biologically Inspired Cognitive Architectures Society, the International Conference on Cognitive Modeling, and the Journal of Artificial General Intelligence. His research has been supported by NSF, ONR, AFOSR, ARL, NASA, DARPA, IARPA, DMSO, and DTRA. His main research interests are cognitive architectures and their applications to psychology, artificial intelligence, human-computer interaction, decision-making, intelligent agents, network science, and cognitive robotics.

Brad Love, Professor
University College of London
http://bradlove.org/lab
b.love@ucl.ac.uk
Brad Love is Professor of Cognitive and Decision Sciences at University College London (UCL) and a Turing Fellow at the Alan Turing Institute, the UK's national institute for data science and artificial intelligence. Dr. Love is interested in topics that cross psychology, neuroscience, and machine learning. He is interested in understanding consumer behavior using large datasets, such as loyalty card data, topics on how people explore product options and construe product categories, and in relating deep learning networks to brain function.

Joseph B. Lyons, Senior Research Psychologist
US Air Force Research Laboratory
joseph.lyons.6@us.af.mil
Joseph B. Lyons is the Lead for the Collaborative Interfaces and Teaming Core Research Area within the 711 Human Performance Wing at Wright-Patterson AFB, OH. Dr. Lyons received his Ph.D. in Industrial/Organizational Psychology from Wright State University in Dayton, OH, in 2005. Some of Dr. Lyons' research interests include human machine trust, interpersonal trust, leadership, and social influence. Dr. Lyons has worked for the Air Force Research Laboratory as a civilian researcher since 2005, and between 2011-2013 he served as the Program Officer at the Air Force Office of Scientific Research where he created a basic research portfolio to study both interpersonal and human machine trust. Dr. Lyons has published in a variety of peer-reviewed journals, and is an Associate Editor for the journal Military Psychology.

Yuko Munakata, Professor
University of Davis, California
munakata@colorado.edu
Yuko Munakata is Professor in the Department of Psychology and Center for Mind and Brain at the University of California, Davis. Her work investigates child development and environmental influences on children's thinking, using behavioral, neuroimaging, and computational approaches. She is an elected fellow of the Association for Psychological Science and the American Psychological Association. Her work on child development has been funded by the National Institutes of Health since 1998, and has been published in top scientific journals and featured widely in the popular press, including The Atlantic, The Today Show, and Parents Magazine.

Dr. Munakata co-edited two books on brain and cognitive development, and co-authored two editions of a textbook on computational cognitive neuroscience. She served as Associate Editor of Psychological Review, and has received numerous awards for her research, teaching, and mentoring. She received her B.A. in Psychology and B.S. in Symbolic Systems from Stanford University. After earning her Ph.D. in Psychology from Carnegie Mellon University, she conducted postdoctoral research in Brain and Cognitive Sciences at the Massachusetts Institute of Technology. She was a professor at the University of Denver and then at the University of Colorado Boulder before moving to Davis.

Aude Oliva, Principal Research Scientist/Executive Director
MIT
http://cvcl.mit.edu/Aude.htm
oliva@mit.edu
Aude Oliva is the Executive Director of the MIT–IBM Watson AI Lab and the Executive Director of The MIT Quest for Intelligence, an MIT-wide initiative which seeks to discover the foundations of human and machine intelligence and deliver transformative new technology for humankind. She is also a Principal Research Scientist at the Computer Science and Artificial Intelligence Laboratory. She formerly served as an expert to the National Science Foundation, Directorate of Computer and Information Science and Engineering. Her trans-disciplinary work in Computational Perception and Cognition builds on the synergy between human and artificial vision, and how it applies to solving high-level recognition problems like understanding scenes and events, perceiving space, recognizing objects, modeling attention and memory. She was honored with the National Science Foundation CAREER Award, a Guggenheim Fellowship, and the Vannevar Bush Faculty Fellowship. She earned a M.S. and Ph.D. in cognitive science from the Institut National Polytechnique de Grenoble, France.

Charan Ranganath, Professor
University of California, Davis
http://dml.ucdavis.edu/
cranganath@ucdavis.edu
Charan Ranganath is the Director of the Memory and Plasticity Program and a Professor, at the Center for Neuroscience and Department of Psychology at the University of California at Davis. Dr. Ranganath's research focuses on how the brain encodes information about the context (when and where an event took place) of an event, motivational factors that influence the stability of memory, and changes in memories over time. His research laboratory also investigates how motivational and emotional factors influence memory.

Dr. Ranganath was a Section Editor for the journal NeuroImage and an editor for the Journal of Neuroscience. He was recognized as a Visiting Professor and Fellow at the University of Cambridge, UK, and a Sage Center Distinguished Fellow. Dr. Ranganath's work has received several awards, including the Samuel Sutton Award for Distinguished Scientific Contribution to Human ERPs and Cognition, the Young Investigator Award from the Cognitive Neuroscience Society, a Guggenheim fellowship, and the Vannevar Bush Faculty Fellowship.

Randy O'Reilly, Professor
University of California Davis
https://sociology.ucdavis.edu/people/oreilly
oreilly@ucdavis.edu
Randy O'Reilly is Professor of Psychology, Computer Science, and the Center for Neuroscience at the University of California, Davis. He has authored over 70 journal articles and an influential textbook on computational cognitive neuroscience. His work focuses on biologically-based computational models of learning mechanisms in different brain areas, including hippocampus, prefrontal cortex and basal ganglia, and posterior visual cortex. He has received significant funding from ONR, NIH, NSF, IARPA, and DARPA. He is a primary author of the Emergent neural network simulation environment. Dr. O'Reilly completed a postdoctoral position at the Massachusetts Institute of Technology, earned his M.S. and Ph.D. degrees in Psychology from Carnegie Mellon University and was awarded an A.B. degree with highest honors in Psychology from Harvard University.

Brian Scassellati, Professor
Yale University
https://scazlab.yale.edu/people/brian-scassellati
scaz@cs.yale.edu
Brian Scassellati is the A. Bartlett Giamatti Professor of Computer Science, Cognitive Science, and Mechanical Engineering at Yale University. He works at the intersection of artificial intelligence, robotics, and cognitive modeling, and is best known for his work in human-robot interaction. His research focuses on building embodied computational models of human social behavior, especially the developmental progression of early social skills. Using computational modeling and socially interactive robots, we evaluate models of how infants acquire social skills and assist in the diagnosis and quantification of disorders of social development, such as autism.

Matthias Scheutz, Professor
Tufts University
https://engineering.tufts.edu/people/faculty/matthias-scheutz
matthias.scheutz@tufts.edu
Matthias Scheutz is Professor in Cognitive and Computer Science in the Department of Computer Science and Bernard M. Gordon Senior Faculty Fellow in the School of Engineering at Tufts University. He received degrees in philosophy (M.A., Ph.D.) and formal logic (M.S.) from the University of Vienna and in computer engineering (M.S.) from the Vienna University of Technology in Austria. He also received a joint Ph.D. in cognitive science and computer science from Indiana University. He has over 300 peer-reviewed publications in artificial intelligence, natural language processing, cognitive modeling, robotics, and human-robot interaction. His current research focuses on complex autonomous robots that can be tasked in natural language.

Patrick Shafto, Professor
Rutgers University - Newark
http://shaftolab.com/
patrick.shafto@rutgers.edu
Dr. Patrick Shafto is the Henry Rutgers Term Chair in Data Science and Professor of Mathematics and Computer Science at Rutgers University - Newark. Research in his lab focuses on theoretical and empirical foundations of cooperation and learning in humans and machines. He has received numerous honors and awards including an NSF CAREER award and his research has formed the basis for successful data science start-up companies eventually acquired by Salesforce and Tableau. His research is supported by multiple NSF directorates (EHR, CISE, SBE), DARPA, DoD, the intelligence community, and the NIH.

Greg Trafton, Cognitive Scientist
NRL
https://www.nrl.navy.mil/itd/aic/IntelligentSystems
greg.trafton@nrl.navy.mil
Greg Trafton is a cognitive scientist with interests in cognitive robotics, human robot interaction, and predictive models of humans. He is head of the The Intelligent Systems Section at the Navy Center For Applied Research in Artificial Intelligence (NCARAI) performs state-of-the-art research in cognitive science, cognitive robotics and human-robot interaction, predicting and preventing procedural errors, the cognition of complex visualizations, interruptions and resumptions, and spatial cognition. Dr. Trafton received his B.S. in computer science with a second major in psychology from Trinity University, San Antonio, TX in 1989. He received an M.A. (1991) and Ph.D. (1994) in cognitive science from Princeton University.

Appendix III – Workshop Agenda and Prospectus

Day 1 – Tuesday, July 16, 2019

Time	Title
8:00 – 8:15	**Check-in and Continental Breakfast**
8:15 - 8:20	**Welcome and Introductions and Expectations** John Laird, U Michigan
8:20 -8:45	**Workshop framing talk** John Laird, U Michigan
8:45 – 9:00	**Breakout Instructions and Morning Break**
9:00 – 10:45	**Working Group I: Challenges and Opportunities in Human Machine Teaming** *What is known about human capabilities (from neurocognitive/social research) in this area? What is unknown and would be useful to know? What a re current interaction capabilities in AI systems? Where are there obvious strengths and weaknesses? How do their weaknesses impact interaction with humans today? What capabilities in AI systems can be taken advantage of in human machine interactions? How are humans adapting to current AI systems? How much do AI systems need to be like humans to make them easy for us to interact with?* *Group A – Physical Aspects of Interaction* *Group B – Cognitive Aspects of Interaction* *Group C – Social Aspects of Interaction*
10:45 – 11:00	**BREAK** Transition to main conference room and leads draft outbriefing summary
11:00 –12:00	**Working Group 1: Outbriefing**
12:00 – 1:00	**LUNCH (provided for participants)**
1:00 – 3:45	**Working Group II: Technical Capabilities and Challenges** *What are the promising directions for improving AI systems for human interaction? What are the potential capabilities of AI systems beyond what we have in humans (different sensing modalities, access to knowledge basis not available to most humans, …)?* *Group A – Physical Aspects of Interaction* *Group B – Cognitive Aspects of Interaction* *Group C – Social Aspects of Interaction*
3:45 – 4:00	**BREAK** Transition to main room and leads draft outbriefing summary
4:00 – 4:45	**Report Out from Breakout II**
4:45 – 5:00	**Summary of Day** Charan Ranganath, U California - Davis
5:00	**MEETING ADJOURNED FOR THE DAY**

DAY 2—Wednesday, July 17TH, 2019

Time	Title
8:00 – 8:15	**Check-in and Continental Breakfast**
8:15 – 8:30	**Welcome and Day 1 Recap** Sam Gershman, Harvard U
8:30 -9:30	**'White Space' Discussion I** Discussion of topics which did not fit into the framework of day 1, but need to be discussed.
9:30 – 10:30	**'White Space' Discussion II** Discussion of particularly far-out (or long-term), high-risk, high-impact ideas.
10:30 – 10:45	**BREAK**
10:45 – 11:45	**Discussion of Key Ideas/Components for Report**
11:45 – 12:00	**Closing Remarks**
12:00	**DEPARTURE**

Future Directions Workshop: Human Machine Teaming

Basic Research Office, Office of the Secretary of Defense

16–17 July 2019
Basic Research Innovative Collaboration Center
4100 N. Fairfax Road, Suite 450 Arlington, VA 22203

Co-Chairs: John Laird (Michigan), Samuel Gershman (Harvard), Charan Ranganath (U.C. Davis)

Interactions with technologically sophisticated AI agents are now commonplace, but we nonetheless routinely encounter difficulties because we have only an incomplete understanding of the technology, and the technology has an incomplete understanding of us. Over the last ten years, there has explosion in research in Artificial Intelligence and in our understanding of the human mind and brain. In this Future Directions Workshop, we will explore different sides of how recent research can inform how humans and intelligent machines can work together. How can our knowledge of the human mind inform the development of intelligent machines so that they can interact more effectively with humans? How are human-to-machine interactions similar to human-human interactions and how are they fundamentally different? How do current AI approaches fare in capturing human capabilities and interactions? What knowledge, representations, and methods do humans use in interacting with each other that need to be modeled and possibly duplicated in machines, and which can be ignored? What do people need to know about what is happening inside AI systems to support effective interaction? What aspects of human-human and human-computer interaction are still a mystery where additional research is needed?

This Future Directions in Human Machine Teaming workshop will gather researchers from the AI and Cognitive Science communities to discuss opportunities and challenges for how knowledge about humans and AI systems can inform each other, while also informing how humans and AI systems can productively interact. The workshop is designed primarily around small-group breakout sessions and whole-group discussions rather than a standard conference format, and aims to shed insight on three overarching questions:

- How might the research impact science and technology capabilities of the future?
- What is the possible trajectory of scientific achievement over the next 10–15 years?
- What are the most fundamental challenges to progress?

The discussions and ensuing distributed report provide valuable long-term guidance to the DoD community, as well as the broader federal funding community, federal labs, and other stakeholders. Workshop attendees will emerge with a better ability to identify and seize potential opportunities at the intersection between the two fields of study. This workshop is sponsored by the Basic Research Office within the Office of Secretary of Defense, along with input and interest from the Services and other DoD components.

Agenda

Day One: The majority of the first day will be spent in small-group breakout sessions on fundamental challenges to progress and technical capabilities. The overarching questions to explore include:

- What is known about human capabilities (from neurocognitive/social research) in this area? What is unknown and would be useful to know?
- What are current interaction capabilities in AI systems? Where are there obvious strengths and weaknesses? How do their weaknesses impact interaction with humans today? What capabilities in AI systems can be taken advantage of in human machine interactions?
- How are humans adapting to current AI systems? How much do AI systems need to be like humans to make them easy for us to interact with?
- What are the promising directions for improving AI systems for human interaction? What are the potential capabilities of AI systems beyond what we have in humans (different sensing modalities, access to knowledge basis not available to most humans, …)?

To explore these questions, the participants will be split into three small-groups for breakout sessions to examine these questions from three separate perspectives:

1. **Physical aspects of interaction.**
   How is information transmitted between individuals, including human-human and machine-human? This includes modalities of production and receiving information across different levels of abstraction including sound, speech, language, vision, gesture/activity, emotion, brain-machine interfaces, etc. What is common across human-human and human machine interactions at the physical level, and how are interactions different today or may become different in the future? How do these commonalities and differences affect human machine interaction?

2. **Cognitive aspects of interaction.**
   What cognitive capabilities are critical to supporting interaction, such as decision making, problem solving, planning, language, access to world knowledge (common sense), and learning? Which aspects of these capabilities are missing from current AI systems and how might they inform the development of cognitive capabilities or other aspects (robustness) of AI systems? How important is it for AI systems to have the same reasoning styles and capabilities as humans?

3. **Social aspects of interaction.**
   What social capabilities do humans have that need to be replicated in AI systems, such as common ground and joint attention, theory of mind, social awareness? How to keep up one's side of the interaction, dialog, establishing and maintaining trust, etc.?

Day Two: The second day of the workshop is a half-day consisting of white-space, whole group discussions on topics that did not fall into the Day 1 framework or were especially ambitious and/or high-risk. Participants will also discuss areas that require more growth, as well as the trajectory of this intersectional area over time. At the end of the day, the whole group will discuss the overarching themes of the workshop that should be included in the final workshop report.

We don't want to be myopic: build a machine that solves a technical problem. Even machine
Not everyone needs to be thinking about how machine understands models