# Integrating Context into Artificial Intelligence: Research from the Robotics Collaborative Technology Alliance

Kristin E. Schaefer[1], Jean Oh[2], Derya Aksaray[3], Daniel Barber[4]
*[1]US Army Research Laboratory, [2]Carnegie Mellon University, [3]University of Minnesota, & [4]University of Central Florida*

**Abstract**. Applying context to a situation, task, or system state provides meaning and advances understanding that can affect future decisions or actions. While people are naturally good at perceiving contextual understanding and inferring missing pieces of information using various alternative sources, this process is difficult for artificial intelligence (AI) systems or robots, especially in high uncertainty and unstructured operations. Integration of context-driven AI is important for future robotic capabilities to support the development of situation awareness, calibrate appropriate trust, and improve team performance in collaborative human-robot teams. This article highlights advances in context-driven AI for human-robot teaming by the US Army Research Laboratory's Robotics Collaborative Technology Alliance. Avenues of research discussed include how context enables robots to fill in the "gaps" to make effective decisions more quickly, provide more robust behaviors, and augment robot communications to suit the needs of the team under a variety of environments, team organizations, and across missions.

## Challenges with Developing Context-Driven AI

Integrating context to support artificial intelligence (AI) development provides a number of potential benefits for efficient teaming and collaborative task accomplishment for human-robot teams. For military teams in particular, integration of context into AI architectures is essential to facilitate collaboration and successful operation in complex and dynamic environments. Take for example the event when a Soldier reports that a hostile threat is in a target area. Given this information, a robot is expected to change how it navigates to the target environment, make its primary objective enemy detection, and provide guidance for the movements and future actions of both friendly and adversarial human counterparts so that team members can remain undetected. However, a human teammate's interpretation of the robot's behaviors is directly influenced by the robot's capability to adequately communicate reasoning for its own prior and current actions. Otherwise, its behavior may appear ambiguous or incorrect from a human perspective. Therefore, the robot needs to both understand how context will or could impact its own decisions, as well as how it could impact team members' decisions. By integrating contextual understanding, it will be possible to enable shared situation awareness and shared mental model development; improve joint decision-making and categorization of data; provide better processing times; and enhance learning both online and offline for the team.

While our early work has identified potential benefits of incorporating context in inference and planning models to enhance shared situation awareness and improve intent-based communication of human-robot teams, we have also identified a number of open research challenges in making these types of advances (Schaefer et al., 2017; Schaefer et al., 2019). These identified gaps in the scientific research community include a large number of unknowns about what constitutes context; differing opinions on how to reason about context; differing recommendations for how to acquire new contextual knowledge; and even the specific means to transfer or communicate the teammate's knowledge of context to other members of the team. Here, we explain these open challenges in detail in the following subsections and then describe how the current work by the US Army Research Laboratory's Robotics Collaborative Technology Alliance (RCTA) has been addressing and advancing research in this area.

### Representing context

First and foremost, context means different things to different people. One well-cited, yet broad definition suggests that context is "any information that can be used to characterize the situation of an entity (whereby) an entity is a user, a place, or a physical or computational object that is considered relevant to the interaction between a user and an application, including the user and application themselves" (Dey, 2001, pp. 106). Long, all-encompassing definitions like this one, with a lack of consensus for an agreed upon operational definition have resulted in large variability in the representations of context across the scientific community. Depending on the problem, context might refer to a relevant part of the state-space (Xiong & Huber, 2010), a probability distribution over the concepts in an environment (Singhal, Luo, & Zhu, 2003), a set of relationships between objects (Rabinovich et al., 2007), logical statements that represent cause and effect (Zettlemoyer & Collins, 2009), or a function to select relevant features for object recognition

(Heitz & Koller, 2008). While these representations can capture some aspects of contextual knowledge, none of them are applicable to general settings due to their strong dependence on domain-specific knowledge. Overall, the key gaps are linked to reconciling different views of context and developing a coherent representation that can be used in various ways.

**Inferring context**

In the current literature, inferring context from the world occurs primarily through use of visual and language-based sensor data. However, limitations in current perception systems can inhibit inference capabilities. For example, the presence of a highly cluttered environment, dynamic objects, and even changes in the image resolution can cause failures in accurate scene detection and inference processing. Similarly, inference models for natural language understanding might struggle in the presence of arbitrary sentences with complex grammatical structure and untrained words. Therefore, a future goal is to extend the application areas of existing inference methods and models for generalization. A second goal is to use sensor data other than vision or language, such as temperature, humidity, smell, and audio, to provide additional, and possibly redundant sources of information to support these inference models. As a member of mixed-initiative team, these factors (e.g. heat exhaustion, distraction, inability to hear communications) may influence human team member performance, requiring adaptations to robot behavior for optimal collaboration. Moreover, temperature and humidity support inferences about potential changes in weather conditions (e.g., fog, wet roadways); smells indicate the presence of chemicals affecting safety; and auditory sensing can be crucial in inferring changes in the scene, such as approaching vehicles, people, or danger. Overall, the challenge is in the identification of which combination of sensory information gives an efficient way for better understanding of context to improve decision-making. A major limitation to this area of research is the limited knowledge on how to derive the required types of sensor data in terms of mission specifications that may might make for an efficient context inference.

**Learning new classification schemes**

Given a finite set of features of a context model (e.g., spatial areas, indoor versus outdoor), it is possible to learn classification schemes to determine the current context. However, pre-specified contextual features are unlikely to be sufficient for long-duration robots operating in populated environments. Since people naturally grow their sense of context over time, it is desired for robots to have the same capability for resilient team-based operation. It is not realistic to assume that all possible contextual features are known *a priori* to a mission. In this sense, mining new contextual knowledge online and incorporating it into context models are two crucial issues for improving the performance of robots that are likely to collaborate with humans. There are some research efforts where new algorithms are under development to acquire new knowledge from the perceived world and to reconstruct the inference model as new information is added (Tucker et al., 2017). However, new methods are required to address the problem of updating the current context model in a scalable way for incrementally discovered information. The main challenge here is that there is not a common solution to this problem since the context models are diverse, application dependent, and incapable of accommodating all types of contextual knowledge.

**Communicating contextual information**

In team collaboration, any contextual variable that drives the inference process must be transparent to team members to build trust and to explain future decisions based on context derived from sensor data. People reason about the world in a way that incorporates diverse contextual information about logic concepts, prior sensor data, or time histories of state estimates. It is possible use this information in their inference and reasoning processes to make decisions. When sharing that level of contextual understanding with robotic team members, it is important to enable direct (e.g., language, text) or indirect (e.g., gestures, emotions, posture) communication methods. Similarly, not all information that has been formed through AI is transparent to humans. For instance, graphical models and neural networks are powerful representations in a robot's inference mechanisms that they can use to instantiate contextual knowledge from a set of sensor data. However, these mathematical representations without annotated explanations may not necessarily be meaningful from the perspective of human. Thus, in a similar fashion of human-to-robot communication, it is crucial to develop a communication architecture that can also support information transfer from robot-to-human in a transparent fashion. In other words, how do we transform what is in the "black box" of machine learning or other algorithms into something a human can easily understand? Furthermore, how do we represent this information through visual or other modes such that a comprehension of the data is quick without overloading human cognition?

**Current work**

This article reviews work on developing context-driven AI for human-robot teaming conducted by the US Army Research Laboratory's Robotics Collaborative Technology Alliance (RCTA) that shows advances to the above listed research gaps. Within the RTCA efforts, context in support of human-robot teaming is defined as *any available information that can fill in gaps, addressing uncertainty, to enable shared understanding and team collaboration*. When integrated appropriately, context supports teaming initiatives for collaborative interactions including planning and prediction, communication, advanced mission goals, and independent as well as collaborative robot decision-making capabilities.

The following sections of this article describe advances in theoretical and applied contributions, as well as research and development efforts, to advance context-driven AI in support of advanced collaborative human-robot teaming. It begins with a general overview of the research goals of the RCTA and how advancing context-driven AI is an integral component of that research. The importance of environmental, mission-specific, and social context for advanced human-robot teaming is discussed. This is followed by specific RCTA research efforts advancing the research and development associated with the above listed challenges for developing context-driven AI. Our work describes the development of a multimodal interface that supports advances in context-driven AI related to natural language, world modeling, and novel concept acquisition. All these technical advances benefit collaborative human-robot teaming.

## Robotics Collaborative Technology Alliance

The RCTA is a large multi-disciplinary research program that includes a consortia of industry and academic partners working directly with government organizations to advance the state of the art in human-robot teaming, or in other words, to revolutionize robots from "tools to teammates" (Phillips et al., 2011). There are four main research thrusts and required capabilities being addressed by the RCTA: 1) optempo maneuvers[1] in unstructured environments, including mobility in dynamic scenes and across rough terrain; 2) human-robot execution of complex missions requiring situation awareness of unstructured environments and distributed mission execution; 3) mobile manipulation in cluttered spaces; and 4) integrated research which combines and assesses capabilities delivered from the other thrusts on multiple robotic platforms. While there are a number of research interests being addressed to advance teaming, *context* plays an important role in all of these areas. In particular, subcategories of RCTA research that drive the advancements in context-driven AI include advancements in semantic perception, adaptive behavior generation, meta-cognition, machine learning, and a hybrid cognitive and metric world model. *Semantic perception* moves robotic perception beyond simply detecting what is or is not an obstacle towards semantic understanding of an environment in a similar way that human team members would perceive or reason about the environment, e.g., recognizing the types of objects and terrains of interest for a specific task, such as navigation (Oh et al., 2015a; Oh et al., 2015b; Oh et al., 2016; Shiang et al., 2017a). *Adaptive behavior generation* combines previously developed robotic planning algorithms, machine learning techniques, and semantic understanding of an environment within the context of a high-level task. This enables robots to generate effective mission plans in partially known and unstructured environments and compute these plans online whenever necessary by following natural language commands (Boularias et al., 2015; 2016; Paul et al., 2016; 2017; 2018; Tucker et al., 2017) or navigating while adapting to social context (Vemula et al., 2017; 2018). *Meta-cognition* enables the use of intui9tive, human-level commands for Soldier-robot communication to facilitate the creation of shared mental models and the development of shared situation awareness (Ososky et al., 2013). The *world model* spans a range including traditional metric data to an associated semantic understanding to underpin the cognitive levels of reasoning (Dean, 2013). Within the RCTA's goal of advancing robots from "tools to teammates", there are three different but interrelated types of context that are directly relevant to improve teaming capabilities: environmental context related to the physical world, mission-specific context derived from the task criteria and goals, and social context related to the agents and interaction of agents within the environment.

[1] Optempo means operations tempo for pace of an operation, its planning, and resupply (Castro & Adler, 1999)

**Environmental Context**

Collaborative team operations occur within an environment which directly influences the perception of team members' actions and actual decision-making behaviors. The main difficulty at present is that humans are very quick to infer meaning from their surroundings. However, the process that robots use for inferring meaning from the world often does not match the process or capabilities of the human, leading to different planning and reason-based decision-making (e.g., path planning; Perelman et al., 2017; 2018; Schaefer et al., 2018). While different is not necessarily a bad thing, it does affect the development of shared situation awareness and shared understanding amongst the team (Chen et al., 2014; Wright et al., 2017).

Context from the environment is semantic information, perceived dynamically or provided *a priori* (e.g., terrain maps). Traditionally, the information extracted from the environment is limited to physical objects and used for navigation and mobility. For example, go from point A to B as quickly as possible. However, the semantic environment includes much more than physics and includes additional data derived and inferred from the aggregation of "world model" information. It can be attributes associated with semantic objects in a robot's world model, part of short and long-term memory, or temporary variables computed in decision-making algorithms. Hence, context becomes important when trying to understand the scene in terms of functionality and affordances, inferred relationships between objects, and influences of the world on the people in that world. Functionality and affordances provide contextual understanding through interpretation of a scene. For example, a person may be sitting on a box, but a box holds things and what is in the box could change the amount of risk associated with operating in that particular area of the environment. Inferred relationships between objects can also directly impact navigation and mobility. For example, a tree is usually an object that should be avoided during path planning, but trees provide cover; operating within close proximity to a tree line could improve stealthy maneuver. It is also important for a robot to be able to interpret and infer how environmental context can influence people (i.e., psychophysiological states including stress, fatigue, and workload). Overall, this type of data supports "priming" of perception systems to better focus on what 'should be' versus what 'might be' in a scene.

Advancement of environmental context-driven AI will produce a better understanding of social situations, as well as an improved detection of specific environmental features that enhance scene understanding and reasoning for observed behaviors. The RCTA research is specifically looking at how the integration of AI that can incorporate elements of environmental context into the decision-making process. This has two major benefits: 1) it allows the robot to reason about world which can impact its subsequent decision-making process, and 2) it supports natural language and bidirectional communication between the robot and other team members to facilitate collaborative decision-making for mission execution.

**Mission-Specific Context**

A task or complex mission goal given to a robot is a rich source of information to be included in AI, and must be able to support decision-making that accounts for rules of engagement, social norms, and the prioritization of objectives. This type of context can therefore reduce the solution space of actions related for a given decision with heuristics or optimize selections within the space with improved weighting of variables. In line with the earlier example of an enemy reported in an area, the type of a mission underway will dictate the robot's goal prioritization and behaviors. For example, should the robot try and scout ahead, stay in formation with the team, or take steps to avoid detection? Thus, context influences the robot's selection of an actions execution (e.g., prioritizing the paths that maximize cover to avoid detection). From a team dynamics perspective, to accomplish a task, each member has their own set of objectives and goals that may be shared or independent each other and the robot. Task or mission context directly tells the robot what role it should be in and what role others will take, such that each individual is doing what is best for overall team performance, especially if the team is distributed throughout an area and not co-located. The RCTA research is specifically looking at how the decomposition of the overall mission goal into tactical behaviors is sensitive to the context of the instruction given by a human teammate related to both the events in the overall environment and to the specific goals of all of the team members.

**Social Context**

Social context is important for teaming because it supports collaborative decision-making. This includes decision-making associated with navigation and how both robots and humans can operate in a shared space (Schaefer et al.,

2016; 2017; Straub & Schaefer, 2018), advanced robot awareness and its understanding of human behavior (e.g., infer a human's intent through pose, location, and social signals; Fiore et al., 2013; MacArthur et al., 2017; Schaefer, Hill, & Jentsch, 2018; Wiltshire et al., 2015), and prediction of future activities based on context related to the scene, objects, and people. Social context guides the robot's interactions with ambient society where people are not bound to mission roles or context and supports detection of team member cues that may result in behaviors outside of mission norms. Teams navigating through crowded environments also requires each team member to have effective models of what the other members of the team are doing at any given time. For robotic team members, the need for real-time awareness of human activity is addressed by studying the approaches to recognize and anticipate the motion of humans, social signals, and the activities in which they are engaged. RCTA efforts are developing approaches to recognize and make inferences about people's actions from the robot's visual data streams to track and predict future actions of people and other dynamic objects. Here, human pose features can be used to provide context to the human's current or potential future activities. Understanding social interactions includes not only being able to predict the actions of other agents but also reasoning about how one's own actions can affect those of other agents. This idea has led the RCTA investigation to socially compliant navigation planning and the computational models for representing such social interactions, for example, using a variation of Gaussian processes (Vemula et al., 2017) and a deep learning model known as Social-Attention (Vemula et al., 2018). These models use context to better support the effectiveness of team communications, minimize the associated cognitive burdens for each teammate, and improve synchronization of human-robot team member tasks.

## Research Advances in Context-Driven AI

Today's Soldiers are required to operate within inherently complex, dynamic environments. Working within teams, regardless of the presence of a robot, includes multiple activities during which these Soldiers must pay attention to their own task execution and their teammates. Inclusion of robots can easily result in teammate overload for a number of reasons. First, robots do not communicate or think like human teammates. Second, while they are equipped with advanced sensors capable of streaming large amounts of data, the data still needs to be parsed for human consumption. Third, there is a degree of required oversight and additional attention for operations based on the capabilities of the autonomy. Research conducted through the RCTA has sought out to address these issues through natural and intuitive bidirectional communication. These efforts have led to the development of a multimodal interface (MMI; Barber et al. 2015; Barber, Howard, & Walter, 2016) that facilitates the exchange of environmental, mission, and social context to AI efforts (see Figure 1). Specifically, this technology and associated research supports the need to improve the development of a common ground for a shared understanding allowing for joint decision-making and collaborative operations.

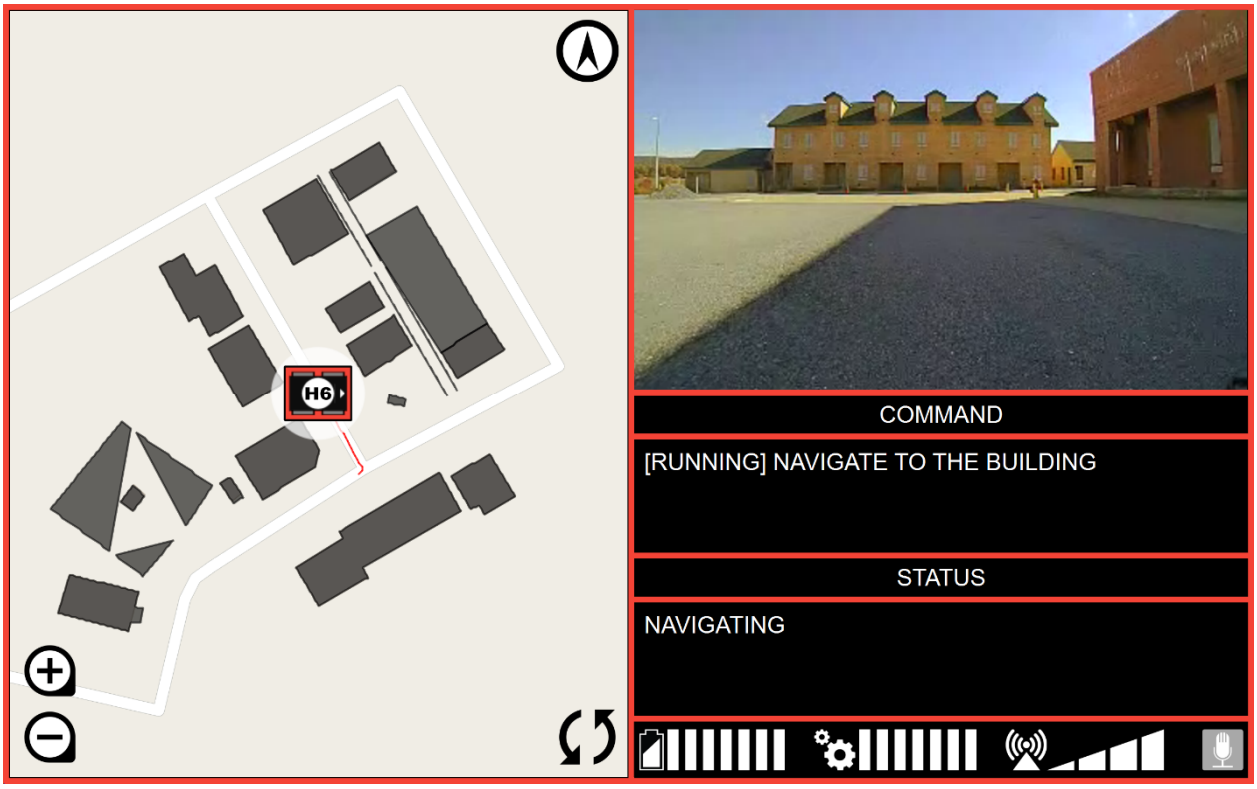


**Figure 1.** RCTA Multimodal Interface (MMI) visual interface. The display includes a semantic map (left), video from the robots perspective or other imagery data (top right), and the robot's action and health status (bottom right).

The RCTA's MMI has leveraged work in speech recognition, natural language understanding, gesture recognition, synthetic speech, tactile displays, and visual displays that model human-robot communication after human-human communications, providing natural and redundant communication channels (e.g., Duvallet et al., 2016; Elliott et al., 2016; Mortimer & Elliott, 2017). These modalities (audio, visual, and tactile) provide additional means to tailor the MMI to adapt the communication of information between teammates to ensure message delivery, robust (fault-tolerant) communication, and shared situation awareness. For example, if a robot determines a Soldier is fatigued or under high workload, the MMI could enable haptic feedback and increase the frequency of radio communications to ensure a message receipt. Moreover, understanding where a Soldier is looking or pointing can add context to the robot's understanding to resolve an ambiguous command such as 'go behind the building' by denoting a specific building or what structures in the environment the team considers 'buildings'.

The MMI acts as a gateway device to facilitate a shared context between robot and human teammates. It provides flexible methods for a Soldier to provide high-level commands to a robot such as 'screen the back of the building' down to specific goal driven semantic navigation, such as 'go to the town square'. It also enables bidirectional communication of a robot's state on a continual basis or through speech dialogues with to request scene descriptions or explanations of its behavior (e.g., 'Where are you going?' or 'What do you see?'). Beyond the provision of dialogue with the robot and a command input, the MMI further attempts to classify the human teammates' states to contribute to the others environmental context. Information about what is in the environment informs the MMI what factors may be impacting changes in the Soldier's decision-making behaviors. Hence, the MMI is not just a portal into the robot for human teammates, but also a sensor about the humans for the robot. This sensor facilitates the acquisition of information for all three categories of context, capturing what each Soldier is currently doing, where they are, physiological and cognitive capacity, and what information they are communicating to all supporting actors. Specific technical advances to context-driven AI support natural language communication, world model development, and novel concept acquisition.

**Context-driven AI and natural language**

One important component of the MMI was development that could support natural language communication. Natural language is a critical capability to facilitate direct human-to-robot mission-specific communication. Speech is the most commonly used method of interaction among human teammates. When a team of agents (human and robot) is performing a shared task, the clarity of the communication and how the context is understood is crucial for the team's success. How language is understood directly impacts development of the shared context, i.e., whether they interpret the task and the environment in the same way such that they can perform the task as one cohesive team. However, our work in developing AI for natural language processing brought to light a number of challenges with integration of robot perception and associated cognition for interpreting human-to-robot communication and performing associated actions that support the team leader's intent. The key finding for this set of research was that the addition of visual descriptors alone does not provide enough contextual understanding to initiate appropriate robot response (Figure 2).

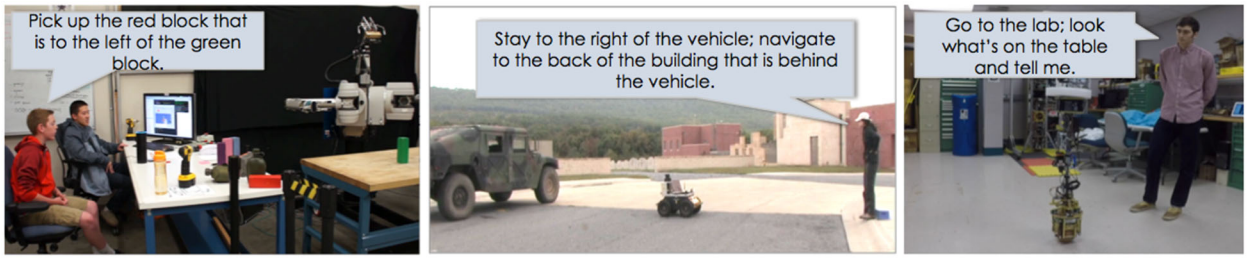


**Figure 2.** This figure represents natural language understanding integrated in various robot platforms: Barrett™ robot arms (left); Clearpath™ Husky robot (center); and CMU Ballbot (right; Lauwers, Kantor, & Hollis, 2005). The difficulty in integrating context-driven AI is in relation to perspective. In the image on the left, perspective is in the word 'left'. What does 'left of the green block' mean -- my left or the robot's left? In the center image, the key word is 'back'. What is the 'back of a building'? In the image on the right, there are questions as to which lab or which table and what should be told to the human.

Natural language processing can also support reduction of ambiguity and improved shared situation awareness by leveraging teammates' inferences of environmental or social context. This added capability can support current technical limitations in robot perception. While recent advances in computer vision show a promising future, to date, robotic perception in real-life environments still remains a difficult challenge for context due to variants in lighting conditions, illuminations, weather, seasons, and more. For instance, recognizing objects in an outdoor environment purely based on computer vision can include both false positive and false negative errors that can result in serious misinterpretation of higher-level tasks. Leveraging language-understanding skills, robots may harvest additional context information to improve their understanding, such as through listening to human teammates in the same environment. For example, consider the navigation task in Figure 2 where a person commands a robot to navigate to the building that is behind a vehicle. Even if the two team members are in the same environment, the way they see the environment may be different. For instance, the robot may misperceive the camouflaged vehicle for part of the background environment, such as a tree, due to its imperfect perception. From the robot's perspective, the command to go to the back of the vehicle, which does not exist in its world model, does not make sense because it is unlikely that the human teammate would use a landmark not close to the robot. Within the context of this task, the robot may update its world model to be more consistent with the command so that the object the robot first recognizes as a tree may be the vehicle of interest.

To account for these issues, the RCTA has looked into how language understanding improves vision-based object recognition on an open data set (NYU Depth Dataset V2; Silberman et al., 2012). Context-based reasoning is modeled as multimodal understanding where the robot continuously updates its task context by fusing new information from vision and language understanding (Figure 3). Graphical model approaches such as Random Walk (Shiang et al. 2017a) and Conditional Random Fields (Shiang et al. 2017b) use context for object recognition where the nodes and the edges of a graph represent objects and their relationships, respectively. These models represent the robot's world model where the contextual relationships among object types can be updated using both vision and language. This

general idea has been applied to various human-robot teaming problems to enable RCTA robots to perform complex tasks in natural language (Boularias et al., 2015; Oh et al., 2015b; Oh et al., 2016). Our intelligence system for multimodal understanding has been integrated to on several physical robot platforms with varying capabilities (refer to Figure 2). These examples demonstrate the use of domain-specific contextual information for various tasks including manipulating table-top objects, navigating in outdoor environments, and on indoor perception.

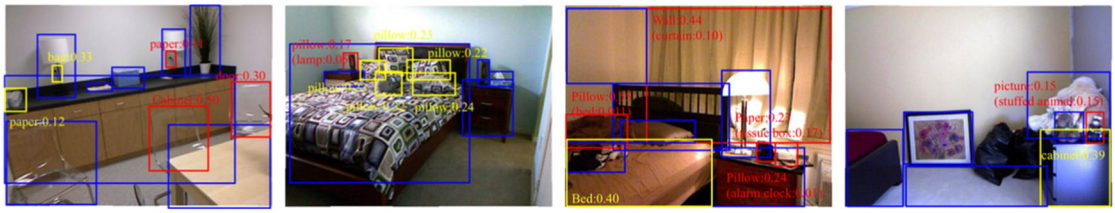


**Figure 3**. Examples illustrating how fusing vision and language can improve object recognition results. The blue boxes represent the correct classification by vision; the red boxes represent misclassification by vision; and the yellow boxes represent those that have been corrected by using language with one to three spatial descriptions (e.g., the chair is near the table).

**Context-driven AI and world modeling**

Context can be useful when robots operate in an unknown or partially known environment. Consider a navigation command, "Go to the barrel behind the building," when an agent is situated in the location pictured in Figure 4 (left image). When this task is given to humans, the majority of people come up with a plan where they hypothesize the rear side of the building not shown in the picture, the space behind the building, and a path towards an imaginary goal.


**Figure 4.** An example of how people plan for the command "Stay to the left of the building; then go to the barrel behind the building" given the leftmost image showing an outdoor environment. The drawings on the right show two examples where human subjects drew their plans for the given command, depicting the start position of the person, the assumption of where the barrel will be located 'behind' the building, and the person's planned path for attempting to reach the barrel.

The RCTA robot planning approach follows this human-like planning concept, known as *assumptive planning*, where the robot must hypothesize the unseen part of an environment to fill the gap between the given command and its world model. The idea of assumptive planning enables the robot to plan and execute commands that require environmental context reasoning in an unknown or partially known environment. Figure 5 illustrates how the robot solves the same navigation command using camera and 3D LiDAR sensors. The semantic analysis is solely based on what has been sensed. The robot then gradually propagates the information to include the hypothesized space. At the same time, the robot computes a metric cost map representing the commander's preference, 'stay to the left of the building', with the knowledge it has gained by having been trained offline via a machine learning technique called "imitation learning" where experts demonstrate desired behaviors (Boularias et al. 2015; 2016). The robot continuously updates its world model of the environment, known as the world model, and its subsequent plan as it perceives more information. Using context, the robot was able to accomplish tasks that it had previously been unable to complete due to missing

information. In two sets of extensive outdoor experiments during the RCTA assessments, the task-level performance for navigation improved from 50% to 75-93% under conditions with varying difficulties (Oh et al., 2016).

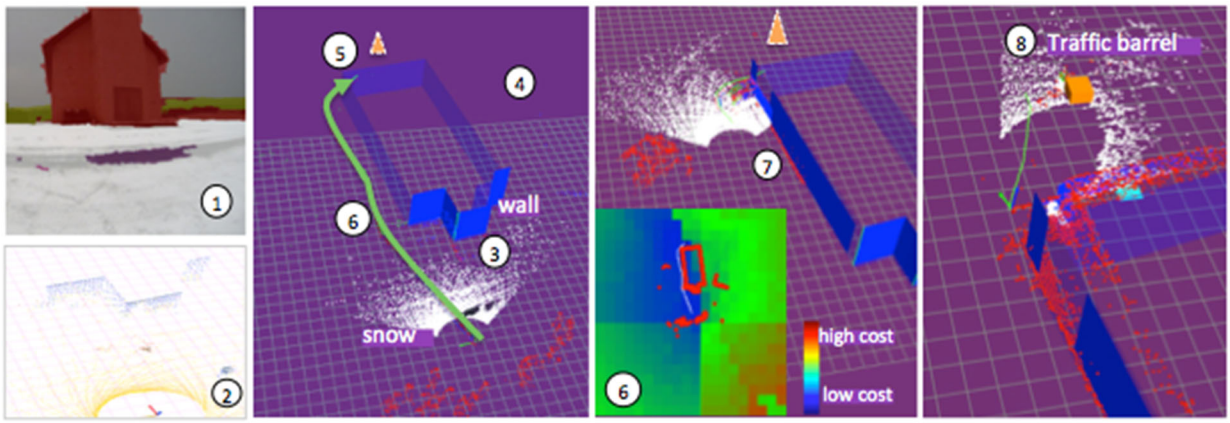


**Figure 5.** This is an example illustration of the assumptive planning approach for a robot given the command: 'Stay to the left of the building; then go to the barrel behind the building' (Oh et al., 2015b). Steps 1 and 2 show the camera and LiDAR sensor data. Steps 3-6 represent the hypothesized space for robot reasoning about the spatial constraint 'behind the building' needed to locate the hypothesized barrel as a target goal and generate a plan. Steps 7-8 demonstrate the robot's capability to continuously update its world model of the environment and its subsequent plan as it perceives more information about the actual environment.

**Context-driven AI and novel concept acquisition**

It is essential for a robot to be able to communicate known and unknown information. By quantifying what is unknown, it becomes possible to dictate needs for contextual understanding to fill in the gaps in a robot's knowledge and reasoning process. To reach this aim, a model was developed under RCTA research efforts to enable learning the meaning of a large variety of phrases in complex environments to facilitate learning new words and objects online. Overall, such a model contributed to language-guided models that enabled online concept acquisition in complex partially-known workspaces (Tucker et al., 2017). For example, the leftmost picture in Figure 6 illustrates a robot and four objects located in the environment. Initially, the robot perceives an unknown object and a jar and does not know the 'box'. When the robot is given a command as 'go to the box', it incorporates environmental context into its decision-making (i.e., reasoning about the existence of a box-object around itself) and infers that the perceived unknown object is supposed to be a box, and then approaches. Moreover, as it approaches, it collects various visual images of the object and use this information in its model to recognize boxes in future instances. Overall, the outcome of this work assists in reasoning about the world containing known and unknown objects and gaining new knowledge through human-robot communication.

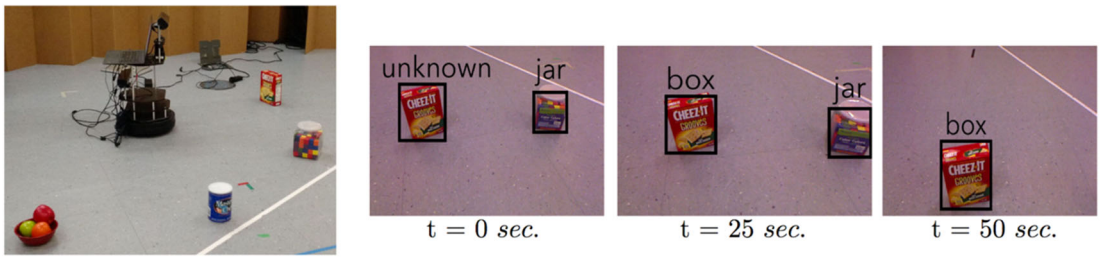


**Figure 6.** This figure demonstrates a grounding scenario (left image) with the TurtleBot mobile platform in an environment populated with objects (i.e., box, jar, can and fruits in clockwise order) from the YCB data set (Calli et al., 2015). The robot is equipped with a Kinect sensor with a limited field of view (62°×48.6°). The goal is to acquire new grounding symbols (right images; Tucker et al., 2017). For example, the robot has a model for grounding a jar object but was not trained to recognize or ground a box object. The robot receives a command "move to the box". Due to the presence of an unknown object in its perceived world, the model grounded the unknown phrase "the box" to the

unknown object and drove to the box. Online re-training was performed with the acquired set of visual observations and the lexical token "the box". Inference occurred in 2.34 seconds.

Humans are good at understanding abstract notions such as a group of cars, a row of blocks in front, or the nearest door from the row of three doors. Recent work has extended existing grounding models to accommodate abstract notions (Paul et al. 2016) or acquired factual knowledge (e.g., 'lift the box that I put down', or 'this is my helmet', 'pick *it* up') from natural language instructions (Paul et al. 2017). These works significantly extend the space of commands that robots can understand and enable a better understanding of temporal and spatial context for more effective and efficient human robot interaction.

## Conclusions

The research efforts in the RCTA have demonstrated that environmental, task, and social context can significantly support human-robot teaming through the development of effective bidirectional communication. To date, the primary work has been on improving robot perception and learning to support tasks related to navigation. The development of an MMI has facilitated major improvements in the development of context-driven AI through natural language understanding, updating the world model, and learning through the identification of unknown objects. Additional investigations are underway to understand how to improve the robot's understanding of context through effective bidirectional dialogues. These dialogues include natural language understanding capabilities to better categorize speech into commands, queries, and reports for the joint construction of context using the research we described here.

Future human-robot teams are envisioned to operate in high-risk optempo and cluttered environments. Therefore, it will be essential for the robot to be able to autonomously adapt its communication modalities and types of information it possesses to better support team member needs in times of high workload or stress. One of the ongoing research efforts within the RCTA aims to integrate a better contextual understanding of the human team member through the real-time classifications of human's state with wearable technologies (e.g., heart rate sensors) to infer human workload, stress, and potentially even trust. This source of information will be used to enable the MMI to adapt a relevant modality of communication, frequency of communication, and types of information transmitted among human and robot teammates while also supplying additional input to AI decision-making modules to maximize effectiveness of interactions and decision-making.

***Acknowledgments***. Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Numbers W911NF-10-2-0016. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. We would like to thank Dr. Nicholas Roy, MIT, and Dr. Susan G. Hill, ARL, for their reviews and guidance on this article. We would also like to acknowledge the members of the RCTA Consortium from the US Army Research Laboratory, Carnegie Mellon University, Florida State University, General Dynamics Land Systems, Jet Propulsion Laboratory/California Institute of Technology, Massachusetts Institute of Technology, QinetiQ North America, University of Central Florida, and University of Pennsylvania who have supported these research efforts.

**References**

*+ denotes work completed in conjunction with the RCTA*

+Barber, D.; Abich IV, J.; Phillips, E.; Talone, A.; Jentsch, F.; and Hill, S. 2015. Field assessment of multimodal communication for dismounted human-robot teams. In *Proceedings of the 59th Human Factors and Ergonomics Society Annual Meeting,* 59(1), 921-925. Los Angeles, CA: SAGE Publications.

+Barber, D.; Howard, T.; and Walter, M. 2016. A multimodal interface for real-time Soldier-robot teaming. In *Proceedings of SPIE Defense, Security, and Sensing – Unmanned Systems Technology*, 98370M, 1-12. Baltimore, Maryland USA.

+Boularias, A.; Duvallet, F.; Oh, J.; and Stentz, A. 2015. Grounding spatial relations for outdoor robot navigation. In *Proceedings of the IEEE Conference on Robotics and Automation (ICRA)*, 1976-1982. Seattle, WA: IEEE.

+Boularias, A.; Duvallet, F.; Oh, J.; and Stentz, A. 2016. Learning qualitative spatial relations for robotic navigation. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 4130-4134. New York: AAAI Press.

Calli, B.; Singh, A.; Walsman, A.; Srinivasa, S.; Abbeel, P.; and Dollar, A.M. 2015. The YCB object and model set: Towards common benchmarks for manipulation research. In *Proceedings of the International Conference on Advanced Robotics*, 510-517. Istanbul: IEEE.

Castro, C.A. and Adler, A.B. 1999. Optempo: Effects on Soldier and unit readiness. *Parameters: Journal of the US Army War College*, 29: 86-95.

+Chen, J.Y.C.; Procci, K.; Boyce, M.; Wright, J.; Garcia, A.; and Barnes, M. 2014. Situation awareness based agent transparency, Technical Report ARL-TR-6905. Aberdeen Proving Ground, MD: US Army Research Laboratory.

+Dean, R.M. 2013. Common world model for unmanned systems. In *Proceedings SPIE 8741, Unmanned Systems Technology XV*, 87410O (17 May 2013); doi:10.1117/12.2016606

Dey, A.K.; Abowd, G.D.; and Salber, D. 2001. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction* 16: 97-166.

+Duvallet, F.; Walter, M.R.; Howard, T.; Hemachandra, S.; Oh, J.; Teller, W.; Roy, N.; and Stentz, A. 2016. Inferring maps and behaviors from natural language instructions. In *Experimental Robotics. Springer Tracts in Advanced Robotics*, edited by M. Hsieh, O. Khatib, and V. Kumar, vol 109, 373-388. Springer, Cham.

+Elliott, L.R.; Hill, S.G.; and Barnes, M. 2016. Gesture-based controls for robots: Overview and implications for use by Soldiers, Technical Report, ARL-TR-7715. Aberdeen Proving Ground, MD: US Army Research Laboratory.

+Fiore, S.M.; Wiltshire, T.J.; Lobato, E.J.; Jentsch, F.G.; Huang, W.H.; and Axelrod, B. 2013. Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemics behavior. *Frontiers in Psychology* 4(859): 1-15.

Heitz, G. and Koller, D. 2008. Learning spatial context: Using stuff to find things. In *Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science*, edited by D. Forsyth, P. Torr, and A. Zisserman, vol 5302, 30-43. Berlin, Heidelberg: Springer.

Lauwers, T.; Kantor, G.; and Hollis, R. 2005. One is enough! In *Robotics Research: Results of the 12th International Symposium ISRR*, edited by S. Thrun, R.A. Brooks, and H. Durrant-Whyte, 327-336. Springer.

+MacArthur, K.R.; Stowers, K.; and Hancock, P.A. 2017. Human-robot interaction: Proximity and speed – Slowly back away from the robot. In *Advances in Human Factors in Robots and Unmanned Systems. Advances in Intelligent Systems and Computing*, edited by P. Savage-Knepshield and J.Y.C. Chen, vol. 299, 365-374. Springer, Cham.

+Mortimer, B.J.P. and Elliott, L.R. 2017. Information transfer within human robot teams: Multimodal attention management in human-robot interaction. In *Proceedings of the 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 1-3. Savannah, GA: IEEE.

+Oh, J.; Navarro-Serment, L.; Suppe, A; Stentz, A.; and Herbert, M. 2015a. Inferring door locations from a teammate's trajectory in stealth human-robot team operations. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5315-5320, Hamburg, Germany.

+Oh, J.; Suppe, A.; Duvallet, F.; Boularias, A.; Vinokurov, J.; Navarro-Serment, L.; Romero, O.; Dean, R.; Lebiere, C.; Hebert, M.; and Stentz, A. 2015b. Toward mobile robots reasoning like humans. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1371-1379. Austin, TX: AAAI Press.

+Oh, J.; Zhu, M.; Park, S.; Howard, T.M.; Walter, M.R.; Barber, D.; Romero, O.; Suppe, A.; Navarro-Serment, L.; Duvallet, F.; Boularias, A.; Vinokurov, J.; Keegan, T.; Dean, R.; Lennon, C.; Bodt, B.; Childers, M.; Shi, J.; Daniilidis, K.; Roy, N.; Lebiere, C.; Hebert, M.; and Stentz, A. 2016. Integrated intelligence for human-robot teams. In *2016 International Symposium on Experimental Robotics,* edited by D. Kulić, Y. Nakamura, O. Khatib, and G. Venture, 309-322. Springer International Publishing.

+Ososky, S.; Phillips, E.; Schuster, D.; and Jentsch, F.G. 2013. A picture is worth a thousand mental models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 57(1), 1298-1302.

+Paul, R.; Arkin, J.; Roy, N.; and Howard, T.M. 2016. Efficient grounding of abstract spatial concepts for natural language interaction with robot manipulators. In *Proceedings of the Robotics: Science and Systems (RSS)*, Ann Arbor, MI.

+Paul, R.; Barbu, A.; Felshin, S.; Katz, B.; and Roy, N. 2017. Temporal grounding graphs for language understanding with accrued visual-linguistic context. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI),* 4506-4514. Melbourne, Australia: AAAI Press.

+Paul, R.; Arkin, J.; Aksaray, D; Roy, N.; and Howard, T.M. 2018. Efficient grounding of abstract spatial concepts for natural language interaction with robot platforms. *The International Journals of Robotics Research*, 37(10), 1269-1299.

+Perelman, B.S.; Evans, A.W. III; and Schaefer, K.E. 2018. Attitudes toward risk and effort tradeoffs in human-robot heterogeneous teams. In *Proceedings of the Human Factors and Ergonomics Society,* 62(1), 1098-1102. Philadelphia, PA.

+Perelman, B.S.; Evans, A.W. III; and Schaefer, K.E. 2017. Mental model consensus and shifts during navigation system-assisted route planning. In *Proceedings of the Human Factors and Ergonomics Society*, 61(1), 1183-1187. Austin, TX.

+Phillips, E.; Ososky, S.; Grove, J; and Jentsch, F. 2011. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 55(1), 1491-1495.

Rabinovich, A.; Vedaldi, A.; Galleguillos, C.; Wiewiora, E.; and Belongie, S. 2007. Objects in context. In *Proceedings of the 11th International Conference on Computer Vision*. Rio de Janeiro, Brazil: IEEE. doi:10.1109/ICCV.2007.4408986

+Schaefer, K.E.; Aksaray, D.; Wright, J.L.; and Roy, N. 2019. Challenges with addressing the issue of context within AI and human-robot teaming. In *Computational Context: The value, theory and application of context with AI,* edited by W. Lawless, R. Mittu, and D. Sofge. CRC Press.

+Schaefer, K.E.; Brewer, R.; Putney, J; Mottern, E.; Barghout, J.; and Straub, E.R. 2016. Relinquishing manual control: Collaboration requires the capability to understand robot intent. In *Proceedings of the International Conference on Collaboration Technologies and Systems*, 359-366. Orlando, FL. IEEE.

+Schaefer, K.E.; Hill, S.G.; and Jentsch, F.G. 2018. Trust in human-autonomy teaming: A review of trust research from the US Army Research Laboratory Robotics Collaborative Technology Alliance. In *Advances in Human*

*Factors in Robots and Unmanned Systems. AHFE 2018. Advances in Intelligent Systems and Computing*, edited by J.Y.C. Chen, vol 784, 102-114. Springer, Cham.

+Schaefer, K.E.; Perelman, B.; Brewer, R.W.; Wright, J.L; Roy, N.; and Aksaray, D. 2018. Quantifying human decision-making: Implications for bidirectional communication in human-robot teams. In *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation. VAMR 2018. Lecture Notes in Computer Science,* edited by J.Y.C. Chen and G. Fragomeni, vol 10909, 361-379. Springer, Cham.

+Schaefer, K.E.; Straub, E.R.; Chen, J.Y.C.; Putney, J., and Evans III, A.W. 2017. Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research: Special Issue on Situation Awareness in Human-Machine Interactive Systems,* 46: 26-39.

+Shiang, S.; Gershman, A.; and Oh, J. 2017b. A generalized model for multimodal perception. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4603-4610. San Francisco, CA: AAAI Press.

+Shiang, S.; Rosenthal, S.; Gershman, A.; Carbonell, J.; and Oh, J. 2017a. Vision-language fusion for object recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4603-4610. San Francisco, CA: AAAI Press.

Silberman, N.; Hoiem, D.; Kohli, P.; and Fergus, R. 2012). Indoor segmentation and support inference from RGBD images. In *Computer Vision. ECCV 2012. Lecture Notes in Computer Science*, edited by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, vol 7576, 746-760. Berlin, Heidelberg: Springer.

Singhal, A.; Luo, J.; and Zhu, W. 2003. Probabilistic spatial context models for scene content understanding. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 235-241. Madison, WI, USA: IEEE.

+Straub, E.R. and Schaefer, K.E. 2018. It takes two to tango: Automated vehicles and human beings do the dance of driving – Four social considerations for policy. *Transportation Research Part A: Policy and Practice.* doi:10.1016/j.tra.2018.03.005

+Tucker, M.; Aksaray, D.; Paul, R.; Stein, G.J.; and Roy, N. 2017. Learning unknown groundings for natural language interaction with mobile robots. *In Proceedings of the International Symposium on Robotics Research (ISRR)*. Puerto Varas, Chile.

+Vemula, A.; Muelling, K.; and Oh, J. 2017. Modeling cooperative navigation in dense human crowds. In *Proceedings of the IEEE Conference on Robotics and Automation (ICRA)*, 1685-1692. Singapore: IEEE.

+Vemula, A.; Muelling, K.; & Oh, J. 2018. Social attention: Modeling attention in human crowds. In *Proceedings of the IEEE Conference on Robotics and Automation (ICRA)*. Brisbane, Australia: IEEE.

+Wiltshire, T.J.; Lobato, E.J.; Garcia, D.R.; Fiore, S.M.; Jentsch, F.G.; Huang, W.H; and Axelrod, B. 2015. Effects of robotic social cues on interpersonal attributions and assessments of robot interaction behaviors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 801-805.

+Wright, J.L.; Chen, J.Y.C.; Barnes, M; and Hancock, P.A. 2017. Agent reasoning transparency: The influence of information level on automation-induced complacency, Technical Report, ARL-TR-8044. Aberdeen Proving Ground, MD: US Army Research Laboratory.

Xiong, X. and Huber, D. 2010. Using context to create semantic 3D models of indoor environments. In *Proceedings of the British Machine Vision Conference*. Aberystwyth, Wales, UK.

Zettelmoyer, L.S., and Collins, M. 2009. Learning context-dependent mappings from sentences to logical form. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 2, 976-984. Singapore: ACM.

**Dr. Kristin E. Schaefer** is an Engineer with the US Army Research Laboratory. She received her Ph.D. in Modeling & Simulation from the University of Central Florida in 2013. Her research interests lie primarily in the areas of artificial intelligence and modeling & simulation approaches to enhance the development of bidirectional communication and trust in human-robot teams.

**Dr. Jean Oh** is a Systems Scientist (research faculty) at the Robotics Institute at Carnegie Mellon University. Her current research is focused on the intersection among vision, language, and planning in robotics. She is passionate about creating persistent robots that can co-exist with humans in shared environments, learning to improve themselves over time through continuous training, exploration, and interactions.

**Dr. Derya Aksaray** is currently an Assistant Professor in the Aerospace Engineering and Mechanics Department at the University of Minnesota. She received her Ph.D. degree in Aerospace Engineering from the Georgia Institute of Technology in 2014. After her Ph.D., she held post-doctoral researcher positions at Boston University from 2014-2016 and at the Massachusetts Institute of Technology from 2016-2017. Her research interests lie primarily in the areas of control theory, formal methods, and machine learning with applications to autonomous systems, robotics, and human-robot teaming.

**Dr. Daniel Barber** is a Research Assistant Professor at the University of Central Florida. He has extensive experience in the field of robotics, simulation development, training environments, and human state assessment. His current research focus is on human system interaction and training assessment including multimodal communication, user interaction devices, teaming, physiological assessment, and adaptive systems.