Army Research Laboratory

ROBOTICS COLLABORATIVE TECHNOLOGY ALLIANCE (RCTA)

FY 2011 Annual Program Plan
Submitted March 2011

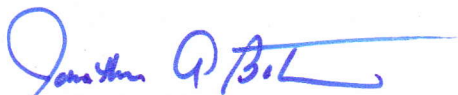| Members |
|---|
| Boston Dynamics |
| Carnegie Mellon University |
| Florida A&M University |
| General Dynamics Robotic Systems (Integration Lead Organization) |
| Jet Propulsion Laboratory |
| QinetiQ North America |
| University of Central Florida |
| University of Pennsylvania |

Robotics CTA FY 2011 Annual Program Plan

For any purpose other than to evaluate the *Robotics CTA FY 2011 Annual Program Plan*, the technical and financial data contained herein shall not be disclosed outside the Government in whole or in part without prior written permission from a duly authorized representative of General Dynamics Robotic Systems. The Government shall have the right to duplicate, use, or disclose the data to the extent provided in the Cooperative Research Agreement.

Robotics CTA FY 2011 Annual Program Plan

The Robotics CTA FY 2011 Annual Program Plan presented in this document has been reviewed and approved by both the Consortium, represented by General Dynamics Robotic Systems, the Integration Lead Organization (ILO); and the Government, represented by Dr. Jon Bornstein, the Collaborative Alliance Manager (CAM).

R R Mitchell

Dr. Robert Mitchell, General Dynamics Robotic Systems
Integration Lead Organization

Jonathan G Bol

Dr. Jon Bornstein, Army Research Laboratory, Robotics Program Office
Collaborative Alliance Manager

Robotics CTA FY 2011 Annual Program Plan

1. INTRODUCTION

Background of Robotics Collaborative Technology Alliance Program

The United States Army Research Laboratory (ARL) issued a Program Announcement (PA)[1] on February 2, 2009 which solicited proposals for a new program entitled the Robotics Collaborative Technology Alliance (RCTA) in order to help fulfill the research and development goals of the U.S. Department of the Army. The stated purpose of this Alliance is to "bring together government, industrial, and academic institutions to address research and development required to enable the deployment of future military unmanned ground vehicle systems ranging in size from man-portables to ground combat vehicles." The PA identified four key technology areas expected to be critical to the development of future autonomous systems, namely:

- Perception
- Intelligence
- Human-Robot Interaction (HRI)
- Dexterous Manipulation and Unique Mobility (DMUM)

It further emphasized the overlap and interplay among these technologies and the need to better understand their interactions through relevant integration and assessment activities.

To accomplish this mission, the PA called for the formation of a "consortium of a small number of industrial and academic institutions acting as equal partners in a research enterprise." The consortium will partner with ARL and other government agencies to advance technology by formulating and executing a number of individual, but coordinated, research tasks. The PA further called for the creation of an Initial Program Plan (IPP) and a series of Annual Program Plans (APPs). The IPP for RCTA, corresponding to work performed during 2010, was "based substantially on the final proposals received from the Consortium," as specified in the PA.

The PA calls for the preparation of a proposed APP for the research in each fiscal year subsequent to that covered by the IPP. The APP provides a detailed plan of research activities, down to the task and even subtask levels. Each year's APP is to be presented to the Research Management Board (RMB) for comments and suggestions. The 2011 APP for RCTA was presented to the RMB on January 21, 2011, and comments and suggestions were duly received.

This document is the revised 2011 APP, which takes into account those recommendations. It consists of six sections. The first section is an introduction which presents the vision of the RCTA, the barriers to achieving that vision, and the primary technical thrusts we are undertaking to overcome the barriers. The introduction also provides a brief overview of each research thrust as well as shows how the thrusts map to the four technology areas identified above, summarizes the process of integrating research outcomes and assessing progress, and shows the relationship of the research to military needs. Section 2 through Section 5 of the APP describes in detail the research by technology area, down to the task and subtask levels. These sections specifically

---

[1] http://www.arl.army.mil/www/DownloadedInternetPages/CurrentPages/CTA/Documents/ROBCTAFINALPA11FEB09.pdf

identify the state of the art related to each research task, describe how the present research moves beyond the state of the art, and identify metrics and goals that quantify the progress toward achieving the research goals. Finally, Section 6 describes the detailed plans for integrating research outcomes in order to achieve needed capabilities that overcome the barriers to our vision for autonomous ground systems.

Vision of the Robotics CTA

Unmanned systems have begun to have a significant impact on warfare; for example, unmanned drones providing sustained surveillance, swift precise attacks on high value targets, and small robots being used for counter-IED missions. While unmanned and highly complex, these systems are still generally remotely piloted systems, reliant upon near-continuous control by a human operator and vulnerable to break-downs of communications links. The future for unmanned systems lies in the development of highly capable systems, which have a set of intelligence-based capabilities sufficient to enable the teaming of autonomous systems with Soldiers. To act as teammates, robotic systems will need to reason about their missions, move through the world in a tactically correct way, observe salient events in the world around them, communicate efficiently with Soldiers and other autonomous systems, and effectively perform a variety of mission tasks. These capabilities certainly do not need to be at a human level, but they do need to be at a level that moves well beyond the current state of tele-operation or closely supervised autonomy.

More specifically, our vision is one where robotic systems have greatly enhanced capabilities in the following five problem domains:

Adaptive Tactical Reasoning. In our vision, robots understand the concept of a mission or task, including stages of progress and measures of success. They work with Soldiers, using the shared concept of METT-TC – mission, enemy, troops, terrain, time, and civilian considerations. They generate tasks to accomplish the mission at hand, reacting appropriately to unforeseen events. They understand their teammates, human or otherwise, and what they need to know during the mission. They make clear distinctions among teammates, adversaries, and non-combatants. They are able to learn from experience, including their own mistakes, generalizing appropriately from specific examples.

Focused Situational Awareness. Future autonomous ground systems maintain situational awareness (SA) that is relevant to the current task and the larger mission. They monitor friendly forces and neutrals and look for threats. They contribute to the general SA of the unit, looking for any salient unexpected events. They continuously predict the future situation so they can better detect anomalies and learn from experience.

Efficient Proactive Interaction with Humans. In our vision, robots interact with each other and especially with Soldiers in an efficient and proactive way relevant to the evolving situation. They receive, understand, and acknowledge orders, asking for clarification if needed. They send relevant messages to their teammates about salient events, using whatever communication means are available. They also receive and understand messages about unfolding events from others and, thus, are able to take any needed actions. They have little or no need for operator control

units (OCUs), instead working with Soldiers as team members utilizing the same modes of command and control that Soldiers use among themselves.

Safe, Secure, and Adaptive Movement. We envision robots that move on orders or their own initiative from one tactical position to the next with little or no reliance on metric inputs such as GPS. They can move, as Soldiers do, to semantically described locations (e.g., "third building on the left after the next intersection"). They also move safely in the presence of people, vehicles, and even animals. They move securely in the context of the current METT-TC, moving with speed or stealth as appropriate. Additionally, they move in a manner that is adaptive to both natural and cultural terrain, including hills, rocks, mud, ice, walls, vehicles, tunnels, and other features.

Interaction with the Physical World. Finally, in our vision, robots are able to observe objects at close quarters to enable 3D interaction with them. They pick-up and move objects, either upon semantic direction or their own initiative. They use tools as necessary for digging, cutting, drilling, etc. They also manipulate doors, windows, hoods of vehicles, etc. as needed to gain access to buildings, vehicles, or confined spaces to execute their missions. They have the dexterity to manipulate a small wire, the strength to pick-up heavy objects, and the range of motion to reach around obstacles. While interacting with the physical world, they can learn, for example, that an object is deformable.

We use a convenient anthropomorphic shorthand – "Think," "Look," "Talk," "Move," "Work" – to encapsulate these five capability building blocks.

Technical Barriers to the Vision

The above vision is an appealing one that promises very great capabilities for future autonomous systems. However, there are significant technical barriers to each envisioned capability. Here, we examine them in turn.

Think. Adaptive tactical reasoning requires both declarative and procedural knowledge with which to reason. Neither exists in current systems, which generally have no data structures for mission level information. Tactical reasoning also requires some kind of model of the other members of the team, both human and robot, so that reasonable predictions of expected behavior can be made. Present systems do not take into account uncertainties in the observed world and the very large decision space in which reasoning occurs. They are forced to reason in a simplified world that does not match reality. Finally, their "thinking" is programmed rather than allowing for adaptation through learning.

Look. The second capability, focused SA, requires a semantic/cognitive description of the robot's environment that current systems do not have. At best, current systems have a map of static and some dynamic obstacles to support navigation. SA also requires a sense of salience, what is important based on a shared understanding among teammates. This sense of salience is missing in the prevailing bottom-up approaches to autonomous perception that are not guided by context. Another critical missing element for effective SA is the ability to learn at a "deep" level, both offline and during operations. Better learning is needed to develop a more human-like

hierarchical understanding of object categories in the first place as well as to refine perception capabilities in the field.

Talk. Existing robotic systems are notoriously opaque and distrusted. For example, they will change course or simply stop during a mission for no apparent reason. They cannot explain what they are doing, primarily because they do not have meta-cognition; in other words, they do not have a model of their own behavior. Current systems also lack the ability to understand human (i.e., semantic) communication of orders or other information. They correspondingly lack the ability to formulate semantic communication to Soldiers to explain what they are doing or ask for guidance.

Move. Safe, secure, and adaptive movement through a complex world is hampered by many technical barriers. First, current systems have insufficient descriptions, or models, of the world in which the robot is moving. They typically have a "green-yellow-red" map of mobility surfaces and possibly a kinematic list of movers. Existing systems struggle to distinguish a stationary person from a barrel or mailbox which represents very different challenges to safe and secure movement. Useful movement is also hampered by the lack of task or mission context so that a robot may persist in trying to reach a particular location that is not needed for the mission. Robots also need to be able to move in crowded and unpredictable environments, where existing algorithmic approaches are probably intractable but new learning approaches may work. They cannot yet adapt to mobility challenges from terrain, weather, etc. by adjusting their gait or form of locomotion.

Work. The above four capabilities (think-look-move-talk) largely enable the performance of the main goal of the mission – the "work" the robot is to do. The work most often involves direct physical interaction with the world: entering and searching a building or vehicle, loading and delivering supplies, inspecting a suspected IED, etc. This direct interaction with the physical world raises several important barriers. First, there is generally great uncertainty about the objects with which the robot is attempting to interact; for instance, exactly what and where are they? An object may be slippery or deformable. Also, the number of objects and the number of degrees of freedom of a mobile manipulator create a state space that is intractably large. Consequently, current approaches are almost entirely tele-operation based, with some attempts at supervised autonomy.

The barriers described above tend to impact multiple desired capabilities of future ground robotic systems. For example, the lack of effective semantic perception affects not only situation awareness but also the abilities to move safely and securely, to communicate about the world, and to interact with objects in the world. Similarly, the other barriers, while daunting, cut across multiple capabilities; therefore, overcoming each barrier results in multiple benefits.

Based upon the above discussion, we identify five primary cross-cutting technical barriers to achieving our vision:

Simplistic/Shallow World Model. Existing autonomous systems fall into two categories: either they have a World Model that is at only a metric level, thus precluding any cognitive reasoning, or they have a model that exists at only a cognitive level without physical grounding in the

metric world. Neither approach is sufficient for our vision where robots must behave cognitively while interacting in the physical world.

Lack of Semantic Understanding. In existing systems, objects in the world are perceived primarily or only as mobility regions, not as discrete objects of semantic and cognitive importance. Thus, one cannot tell a robot, "Go block the back door of this building" and expect it to do anything useful.

Scripted and Brittle Planning. Robots are almost always tele-operated or, at best, only perform simple scripted behaviors. Scripting all needed behaviors is not tractable and does not allow for learning new or alternative behaviors. Planning algorithms in robots work well only when the planning space is both small and certain enough, but the real world is fraught with uncertainty and high dimensionality. The inability to reason in complex and uncertain environments means that users must intervene frequently in robot operations and are trapped at a close level of "supervised autonomy."

No Shared Understanding of Missions and Roles. Robots now are opaque and distrusted and cannot explain what they are doing. Not only do they not know what they are doing, but also they do not understand what their teammates are doing or what the expectations for roles and communication are. Consequently, current systems must use tedious OCUs to bridge the enormous cognitive gap between humans and robots.

Missing or Shallow Learning Capability. Robots now must be explicitly programmed to do tasks, so producing the needed scope of behavior is intractable. Existing learning capability is shallow and lacks generalization. Thus, we cannot retrain robots without bringing engineers to the field or sending the robots back to the developer.

Figure 1-1 summarizes these barriers as columns and relates them to desired capabilities that are listed in rows. The fact that most of the table entries are filled-in demonstrates how the technical barriers impact many capability gaps.

Barriers to Achieving our Vision -->		Simplistic and Shallow World Model	Mobilty-Focused Perception	Tele-operation or (at best) Scripted Planning	No Shared Understanding of Missions and Roles	Missing or Shallow Learning Capabilities
		World Model is either at only a metric level, precluding reasoning, or at only a cognitive level without physical grounding	Objects in the world are perceived primarily/only as mobility regions, not as discrete objects of semantic and cognitive importance	Bots are almost always tele-operated or at best only perform simple scripted behaviors -- and scripting all needed behaviors is not tractable	Bots are opaque and distrusted, and cannot explain what they are doing -- primarily because they don't know	Bots must be explicitly programed to do tasks, so it is intractable to produce the needed scope of behavior. Any learning capability is shallow and lacks generalization
"Think"	**Adaptive Tactical Reasoning**					
	Understand tasks, missions (METT-TC)	World model needs to represent concepts such as missions, tasks and generally METT-TC		Robots need to generate behaviors pertinent to achieving the mission, adapt to changing situation	Robots need to be able to follow instructions given at a semantic or cognitive level, not just "go to (x,y)"	
	Follow semantic instructions					
	Generate behaviors to achieve mission, adapting to changing situation					
	Understand teammates and what they need to know					
"Look"	**Focused Situational Awareness**					
	Maintain SA relevant to current task/mission	World model needs to represent, maintain, monitor and correct all info needed for SA	Robot needs to contribute to the general SA of the unit, noting salient observations		Robot needs to report on salient observations as needed to other elements of its unit	Robot should learn by comparing its observations and actions to those of its human counterparts
	Contribute to general SA of unit					
	Look for salient unforeseen events					
	Observe and report on salient activity					
"Move"	**Safe, Secure and Adaptive Movement**					
	Move cognitively in relation to salient entities in the world (as people/dogs do) w/o GPS or other metric crutches	World model needs to store and operate upon all entities needed to relate movement to tactical constraints	Robot must perceive all entities in its environment relevant to safe, secure and adaptive movement	Robots must move in a tactically correct manner and react to changes in mission or circumstances		Robot needs to learn from its movement experience, whether from mobility challenges or tactical behavior
	Move in tactically and contextually relevant manner					
	Adjust to mobility challenges such as terrain, weather, barriers					
"Talk"	**Efficient Interactive Communication**					
	Receive and acknowledge semantic instructions	World model needs to have shared mental models as a basis for human-robot interaction	Robot needs to send and information relevant based on shared perception (common ground)		Roobot needs to receive and acknowledge cognitive-level instructions and similarly explain its own behavior	Robot needs to be able to learn through cognitive-level interaction with human teammates
	Explain own behavior					
	Report information relevant to mission					
	Seek guidance as needed					
"Work"	**Interaction With Physical World**					
	Inspect and manipulate objects	World model needs to represent wide variety of objects to be manipulated	Robot needs to perceive well enough to interact effectively with objects in a 3D world	Robot needs to figure out how and when to manipulate or transport objects as needed		Robot needs to learn from interaction with the physical world, e.g., when door is locked
	Transport objects as needed					
	Open doors, windows, hoods, trunks, etc					
	Use tools as needed					

Figure 1-1: Five primary barriers limit the capabilities of autonomous ground systems.

Fundamental Research Thrusts for Overcoming Technical Barriers

Derivation of Research Thrusts

Each of the technical barriers described in the preceding section has spawned one or two technical thrusts to address and overcome it.

1) To replace the existing shallow and simplistic world models, we are developing a Cognitive-to-Metric World Model that is the foundation for much of the work of the RCTA program. This is the first thrust in the Intelligence technical area. The World Model must simultaneously handle both cognitive constructs, such as missions on the one hand and the details of vehicle traction on the other. Thus, our architecture weds a top-down cognitive/deliberative framework to a bottom-up algorithmic/reactive framework and joins them in the middle via statistical reasoning to manage uncertainty. The World Model must ultimately include most or all of the following elements:

- Hierarchical data store combining memory resident and traditional relational database management system (RDBMS) techniques to form a Hybrid Database.
- Bi-directional linking of metric and cognitive data.

BostonDynamics   Carnegie Mellon   California Institute of Technology Jet Propulsion Lab   Florida A&M University   GENERAL DYNAMICS Robotic Systems   QinetiQ North America   UCF   Penn

- Long-term declarative memory which integrates, fuses, and infers from multi-sourced data (perception, semantic relations, a priori, training data, teammate knowledge) across time and space.
- End-to-end support for multi-hypothesis classification and reasoning.
- Prediction: how does the world change due to action X? Inclusion of time in planning cycles.
- Internal data structures to optimize memory and CPU usage to support task-based focus of attention.
- Merging of a priori GIS and sensed metric data.
- Resolution independent storage.
- Support and enable learning.
- Using Shared Mental Models to maintain context and state within a team while minimizing bandwidth.

2) To provide a much more complete description of the world in which a robot moves, we are developing a Perception approach that labels the environment semantically. This semantic labeling can then populate the world model at a cognitive level suitable to support reasoning. We have identified two related thrusts in semantic perception – one focusing on static scene understanding and the other on dynamic understanding and prediction.

3) To move beyond tele-operation and scripted behavior, we have three thrusts in adaptive behavior generation. In the Intelligence technical area, the first thrust uses cognitive reasoning approaches to generate adaptive tactical behaviors like searching for an alternative method to enter a building if the first one fails. In the DMUM area, the second thrust enables adaptation to challenges encountered while interacting with the physical world; for example, changing gait in response to slippery conditions. A third thrust supports both high- and low-level planning: it seeks to overcome the barriers posed by a world that is fraught with uncertainty and complexity. When a high-DOF manipulator needs to grasp an object of uncertain size, shape, and position, the state space quickly becomes unreasonably large. This thrust seeks to bound the state space through better reasoning and sensing. Related work in world modeling tries to bound the problem using new representations of the state space.

4) To make robots more trusted partners, we have two thrusts in the area of transparency and meta-cognition. The intra-team cognition thrust, in the HRI area, develops shared mental models to provide common ground and a basis of higher trust. In the Intelligence area, the transparency thrust gives the robot self-knowledge and the ability for a two-way semantic communication with Soldiers.

5) To overcome the intractable problem of trying to program all needed behaviors, our learning thrust is aimed at designing in the ability to learn rather than attempting ad hoc learning after the fact. Key to our approach is the notion of deep learning – we want robots to learn at a conceptual/cognitive level, as humans do, rather than shallow, imitative learning where the true lesson may be missed.

Figure 1-2 summarizes the barriers we have identified as well as the new approaches we have taken to overcome them. The table also lists the research thrusts we have defined within the

technical areas of Intelligence, Perception, Human-Robot Interaction, and Dexterous Manipulation and Unique Mobility.

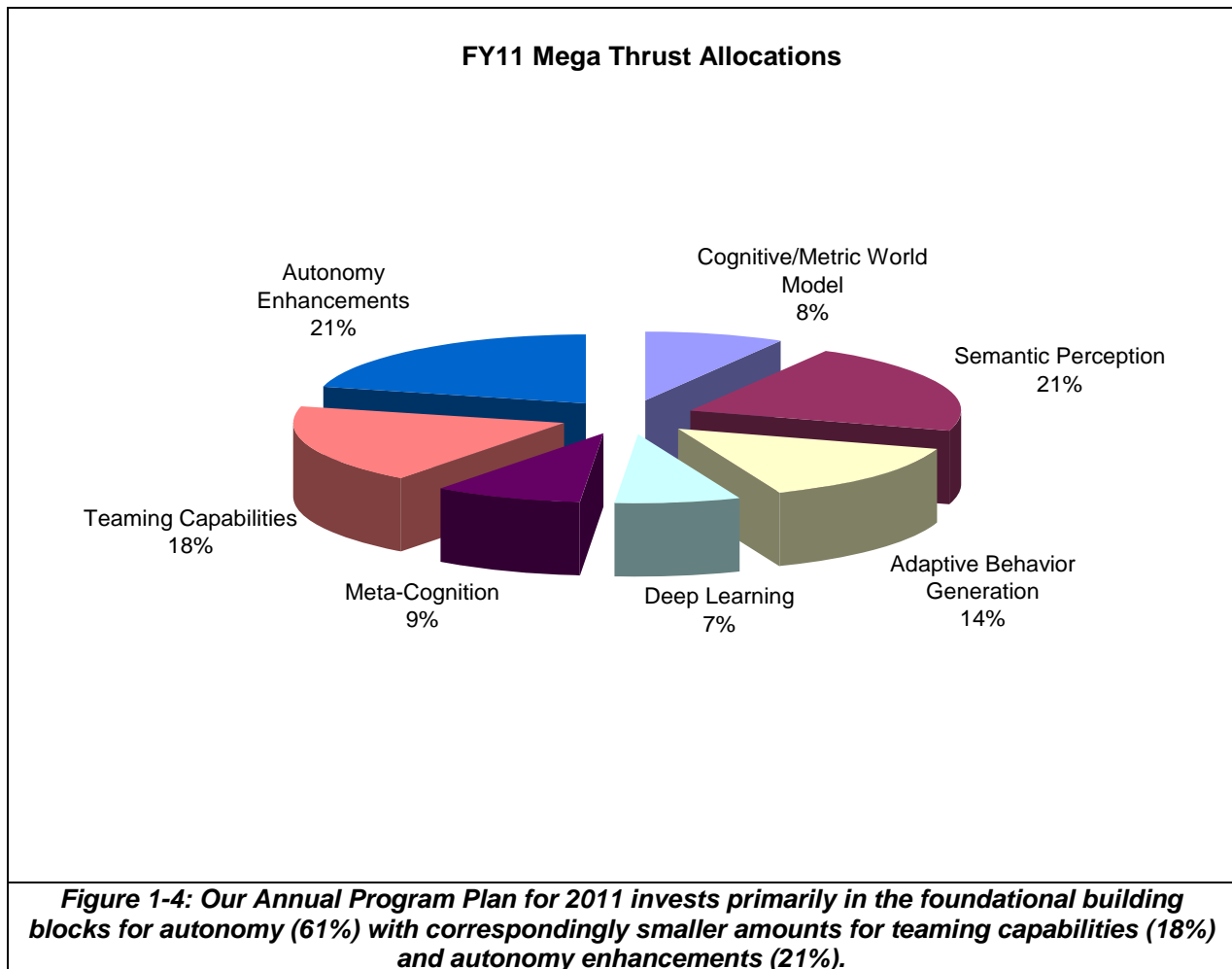| Barrier	Research Thrust	Research Elements
1) Simplistic and Shallow World Model	Cognitive to Metric World Model	New Intelligence Framework (Intelligence)
2) Perception Limited to Generic Obstacle Detection	Semantic Perception using Bottom-up Context and Top-down Guidance	2a. Terrain and Object Classification, Identification and Reasoning (Perception)
		2b. Activity Detection and Understanding (Perception)
3) Scripted, Brittle Planning	Adaptive Behavior Generation From Reasoning with Uncertainty in Very High-Dimensional Cognitive and Metric State Spaces	3a. Cognitive Reasoning and Behavior Generation for Tactical Missions (Intelligence)
		3b. Behavior Generation for Manipulation (DMUM)
		3c. Adaptive and Unique Mobility Behavior Generation (DMUM)
4) Missing or Shallow Learning Capabilities	Deep Learning Coupled to Hierarchical World Model	Learning and Adaptation (Intelligence)
5) No Shared Understanding of Mission and Roles	Shared Mental Models Based on Cognitive World Model	5a. Intra-Team Cognition (HRI)
		5b. Meta-Cognition and Transparency (Intelligence)
		5c. Common Ground for Shared SA (Perception)
Figure 1-2: We have defined new approaches for the five main barriers to autonomy, which lead to nine inter-related research thrusts.		
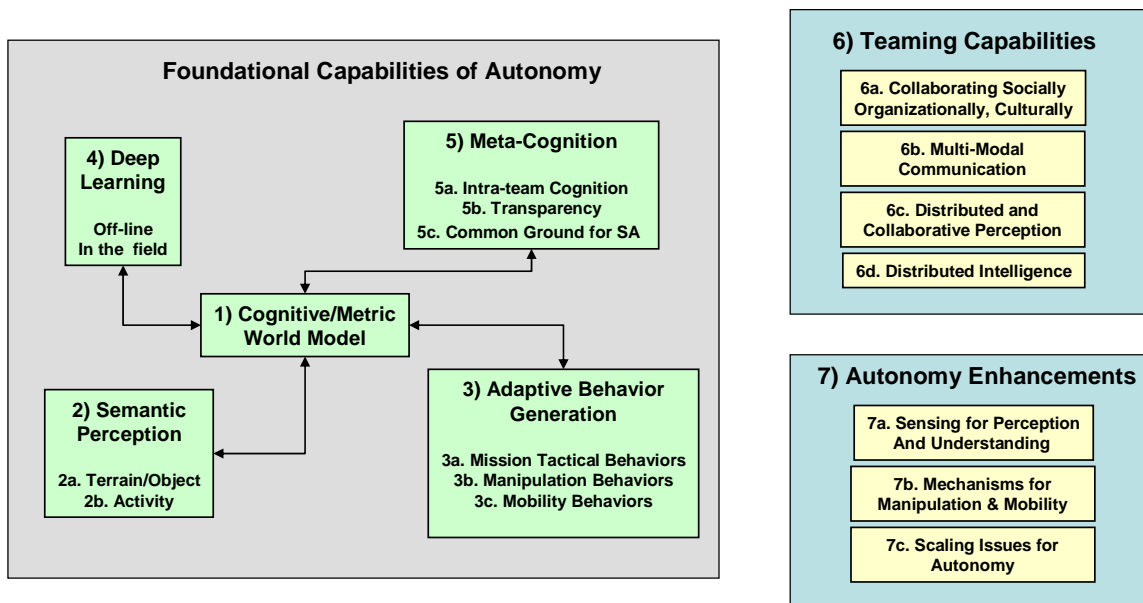
Figure 1-3: Five interrelated thrust areas (green boxes) represent the foundation of our autonomy vision. They correspond to nine technical thrusts in our Annual Program Plan. Seven additional thrusts in the areas of teaming and enhancements for autonomy build on that foundation.

Overcoming the five barriers listed in Figure 1-2 will result in an entirely new level of autonomy consistent with our vision. To overcome these barriers, we have taken the technical thrusts corresponding to the green boxes in Figure 1-3. All of these approaches represent foundational capabilities – cognitive-metric world model, semantic perception, adaptive behavior generation, meta-cognition, and deep learning – which are essential to intelligent autonomy. Without semantic perception, for example, the world model cannot be populated with information to generate adaptive tactical behaviors. Without meta-cognition, the robot cannot understand and communicate its role in a mission. Moreover, without learning, its mistakes can only be corrected through intractable reprogramming.

To implement the five thrusts, we have defined ten research elements as shown in Figure 1-3. For semantic perception, there are two elements: one in terrain and object classification, identification, and reasoning and another in activity detection and understanding. There are three technical components of adaptive behavior generation: one each for mission-level tactical behaviors, for dexterous manipulation control, and for unique mobility planning. Meta-cognition includes both team cognition associated with shared mental models and transparency that arises from introspection.

The above five technical thrusts provide the foundational capabilities for individual autonomy and, thus, merit the highest priority in our program plan. However, we can more fully advance the state of the art in ground autonomy by pursuing two additional research areas – first in the area of teaming and second in an important set of technology enhancements. Through the five thrusts, we have a solid foundation for autonomy. The additional areas build upon that foundation. Figure 1-4 illustrates the relative investment in the five foundational autonomy areas, in teaming research, and in research for autonomy enhancements.

FY11 Mega Thrust Allocations

Autonomy Enhancements 21%

Cognitive/Metric World Model 8%

Semantic Perception 21%

Teaming Capabilities 18%

Meta-Cognition 9%

Deep Learning 7%

Adaptive Behavior Generation 14%

Figure 1-4: Our Annual Program Plan for 2011 invests primarily in the foundational building blocks for autonomy (61%) with correspondingly smaller amounts for teaming capabilities (18%) and autonomy enhancements (21%).

All members of the Consortium are also making cost sharing contributions to RCTA. Through this cost share, they are contributing platforms, simulation software, in-kind research, workshops, seminars, and short courses for the benefit of the entire Alliance.

Through the teaming research area, we will capitalize on the benefits of autonomy as well as test and refine it. To achieve this teaming capability, we have identified four components as indicated by the yellow boxes along the right side of Figure 1-3:

- Collaborating socially, organizationally, and culturally
- Multi-modal communication
- Distributed and collaborative perception
- Distributed intelligence

The second additional area enhances the foundational capabilities previously described in three specific ways. First, we pursue an effort in sensing focused on the specific needs of robotic perception to enhance semantic perception. Second, to enhance the algorithm-focused thrust in adaptive behavior generation for manipulation and mobility, we pursue applied research in the

associated mechanisms as well as basic research in a new generation of actuation materials and approaches. Third, we address the need to develop autonomy across a wide range of platform scale sizes with the associated variability in sensor payloads and computing capacity.

Capability Area	Technical Approach	Research Elements
6) Teaming	Computational Models of Trust	6a. Collaborating Socially, Organizationally and Culturally (HRI)
	Explicit and Implicit Communication Building on Shared Mental Models	6b. Multi-Modal Communication (HRI)
	Shared Semantic Mapping	6c. Distributed and Collaborative Perception (Perception)
	Shared Intelligence Based on Mental Models	6d. Distributed Intelligence (Intelligence)
7) Enhancements to Autonomy	All-weather, multi-spectral and proprieceptive sensing	7a. Sensing for Perception and Understanding (Perception)
	New forms of actuation based on smart materials	7b. Mechanisms for Manipulation and Mobility (DMUM)
	Efficient and distributed algorithms	7c. Scaling of Autonomous Capabilities (Intelligence)

Figure 1-5: We have identified key supporting research in collaboration, sensing, and actuation, which leads to six key supporting research thrusts.

Thus, our primary effort is represented by the five foundational capabilities (green boxes) in Figure 1-3 which correspond to the ten research elements of Figure 1-2. There are seven supporting, but very important, research elements as listed in Figure 1-5. All of these thrusts are briefly summarized below. Those detailed discussions of technical research thrusts are organized according to the four technical domains that were previously identified: Intelligence, Perception, Human-Robot Interaction, and Dexterous Manipulation and Unique Mobility.

The third column of Figure 1-2 and Figure 1-5 identifies which of these technical domains corresponds to each of the technical research thrusts discussed below and in the following sections of this document.

Overview of Technical Research Thrusts

This section provides a brief overview of all technical research thrusts in our 2011 Robotics CTA Annual Program Plan. Detailed descriptions of all research, to the task and subtask levels, is given in Section 2 through Section 5 of this document. Figure 1-6 relates all of the tasks described in those sections to the research thrusts previously described.

Mega Thrust	Research Area	Technical Thrust	Task	Page Reference
Cognitive/Metric World Model	Intelligence	Intelligence World Model	I1: Framework for Intelligence (6.1)	39
			I2: Data Mappng for Inference and Focus (6.2)	52
Semantic Perception	Perception	Static Understanding	P3: Static Scene Understanding (6.1)	182
			P4: Perception for Missions in Complex Environments (6.2)	199
		Dynamic Understanding	P5: Dynamic Scene Understanding (6.1)	216
			P6: Perception for Missions in Dynamic Environments (6.2)	228
Adaptive Behavior Generation	Intelligence	Adaptive Tactical Behaviors	I3: Combining Cognitive and Probabilistic Reasoning (6.1)	68
			I4: Generating Adaptive Tactical Behaviors (6.2)	76
	DMUM	Dexterous Manipulation	M1: Theory and Prinicples of Mobile Manipulation (6.1)	364
			M2: Principles of Generalized Grasp Mechanics (6.1)	375
			M4: High Degree-of-Freedom Dynamic Manipulation (6.2)	391
		Unique Mobillity	M5: Theory and Principles of Multi-modal Locomotion Planning and Control (6.1)	398
			M6: Principles of Locomotion Mechanics (6.1)	408
Deep Learning	Intelligence	Learning	I5: Learning Through Experience (6.1)	89
			I6: Life-long Learning (6.2)	97
Meta-Cognition	Intelligence	Transparent Reasoning	I7: Evaluating and Explaining Performance (6.1)	105
			I8: Diagnostics and Confabulation (6.2)	114
	Perception	Collaborative Perception	P7: High-level Abstract Mapping (6.1)	250
	HRI	Intra-Team Cognition	H1: Team Shared Mental Models (6.1)	284
			H2: Situation Awareness (6.1)	292
			H3: Trust in Human-Robot Interaction (6.2)	301
Teaming Capabilities	Intelligence	Collaborative Behaviors	I9: Distributed Intelligence for Human/Robot Teams (6.1)	124
	Perception	Collaborative Perception	P8: Distributed Team and Target Localization (6.2)	262
	HRI	Multi-Modal Communications	H4: Speech and Non-linguistic Communication (6.2)	308
			H5: Gestures, Posture, and Haptics in HR Communication (6.1)	313
			H6: Implicit Communication (6.1)	320
			H7: Integrating Multi-modal HR Communications in Live and Virtual Environments (6.2)	327
		Organizational Collaboration	H8: Dynamics of HR Military Teams and Organizations (6.2)	333
			H9: Social Cues and Behaviors in HR Collaboration (6.1)	342
			H10: Dynamics of Operating within Social Environments (6.1)	351
Autonomy Enhancements	Intelligence	Collaborative Behaviors	I10: Maximizing Performance with Minimal Resources (6.2)	134
	Perception	Sensing	P1: Exploiting Novel Sensor Phenomenology (6.1)	148
			P2: Compact, High Performance Sensors (6.2)	160
	DMUM	Dexterous Manipulation	M3: Sensor-based Dexterous Manipulation (6.2)	380
		Unique Mobillity	M7: Learning Terrain Interactions (6.2)	415
			M8: Dynamic Multi-modal and Reconfigurable Mechanisms (6.2)	420
		Next Generation Actuation	M9: Next Generation Actuators and Materials (6.1)	427

Figure 1-6: The research summarized below is described in detail in Section 2 through Section 5 of this document.

Cognitive to Metric World Model (1)

The new cognitive-to-metric world model is at the heart of the RCTA program. As Figure 1-3 depicts, all other elements of autonomy interact closely with the world model and indeed must operate through it. To construct this new world model, we pursue a "best in class" approach to architecture that weds a top-down cognitive/deliberative framework to a bottom-up algorithmic/reactive framework and joins them in the middle via statistical reasoning to manage uncertainty. In the current state of the art, cognitive architectures have been used to implement sophisticated behaviors in simulation, but these simulators do not capture the difficulties of a real robot interacting with the real world. We investigate the use of a cognitive architecture to recognize and compensate for robot failures using a cognitive model and associated processes for the robot and the task. The idea is to produce a robot system that does not break as soon as the first unexpected event occurs; instead, it is able to recover and plan a workaround.

The addition of legged mobility and whole body manipulation adds multiple degrees of freedom beyond traditional wheeled or tracked based systems, with a focus on smaller platforms. We will investigate the appropriate algebra to simplify the computations required for these systems to interact with the environment. One potential method is to inject traditionally cognitive capabilities into the lower level dynamic planner to allow navigation using a context-based comprehension of the physical environment. This reflects a common theme through intelligence – while we believe in an overall hierarchical architecture to wed cognitive/metric/physical levels, we also feel gains can be made by applying combined techniques at the module level.

For robots to deeply understand environments, terrains, situations, and activities, they require a broad range of data types. We employ a world model strongly tied to the robot's architectural framework. The world model is more than a data repository, rather an adaptive process that answers questions about its aggregated information. It is a substrate for data from widely distributed sources and inferring greater SA.

The use of hierarchal world models within robotics is well proven, yet implementations are constrained to the needs of sub-specialties of robotics or for the support of specific algorithms. Instead, we seek a cross-discipline world model capable of storing and indexing metric, semantic, and cognitive information. The world model should support multiple capabilities for the robot from navigation through manipulation and adaptive behaviors supplied by semantic knowledge, learning, and cognitive control. We will investigate applying state of the art concepts developed in the database and machine learning communities to enable a query-based world model, feeding historically separate algorithms from a common probabilistic data store. Assessment of this concept involves the applicability of this mechanism throughout the intelligence architecture: providing focus of attention to perception, feeding world changes to cognitive agents, and enabling algorithms to move beyond traditional 2D raster map-style representations into volume and depth-based 3D.

Semantic Perception (2)

The world model described above must be populated with information about the physical world in which the robot resides. In current systems, this information is very high in data content but

very low in information content – the robot is often given massive sets of data such as pixels, point clouds, and radar returns, while it actually needs much more compact information about what *objects* are in the world around it and what they are *doing*. It is the role of Semantic Perception, aided by the cognitive level of the world model, to provide this needed information at the appropriate level of abstraction so other elements of the architecture can reason about them. Cognitive guidance via the world model is critical since unaided state of the art perception algorithms appear to be approaching their limit of performance. We divide the semantic perception effort into components, one for the understanding of static entities and the other for understanding activity. Supporting both components, we pursue a third, smaller focused effort to improve robotic sensing.

Terrain and Object Classification, Identification, and Reasoning (2a)
In order to reason beyond safe driving, robots must have a detailed understanding of the world, including a description of objects, material, and other salient features in their environment. In addition to "naming" the entities in the scene, perception should also derive qualifiers (e.g., parked car, occluded wall) and relations between scene parts (e.g., car in front of door) from sensor data. This level of scene understanding remains a challenging problem that has focused the attention of the computer vision community in the past decade. In particular, much progress has been made in object recognition, but these techniques are still brittle.

One major objective of this thrust is to design efficient learning and recognition algorithms through the use of techniques like deep inference that replace intractable global optimization problems commonly used in state of the art vision systems by approximations that are considerably more efficient and which, when integrated, achieve similar or better recognition accuracy.

Another approach is to use the constraints induced by operation in a particular type of environment. For example, we can use the known context of an urban setting to generate accurate representations of the environment from sensor data. Here, we will incorporate facade detection algorithms and combine the urban scene analysis tools with motion and visibility cues. Major advances in scene parsing, scene surface layout analysis, and 3D reconstruction will be combined and leveraged in a uniform framework to advance the state of the art in overall scene understanding.

Activity Detection and Understanding (2b)
A major objective of the second semantic perception thrust is to investigate robust approaches to detecting, tracking, and identifying objects in general configuration, to use the resulting intermediate description to identify behaviors of individuals and groups, and to predict distributions of likely behaviors based on learned models. As part of this thrust, we also develop algorithms for mid- and long-range perception, situational awareness, and gesture recognition.

While vast literature exists in the current state of the art for short-term prediction based on classical temporal filtering techniques, longer-term prediction of motions and behaviors remains open, and recognition of behavioral cues has been confined to well-structured environments.

The other key factor contributing to degraded performance of current state of the art detection and tracking algorithms is that typically environments are densely cluttered, thus causing the perception system to lose track of the objects as they move through extended occlusions. We plan to pursue two approaches to address this difficult problem. The first approach is an application of our general "purposive prediction" model; the second approach is based on forward-simulating the motion of tracked people.

Another area of research of this thrust is in pose estimation for behavior understanding. Recent work in human pose estimation has focused on two different domains. Historically, researchers have worked in laboratory environments, which make a variety of simplifying assumptions to sidestep the difficulty of detecting human body parts in natural images: using multiple cameras or active sensors, a known background, a limited known range of poses, or even motion capture markers. In these simplified settings, researchers have enjoyed success in estimating and tracking 3D locations of joints and classifying basic actions with relatively high accuracy. However, many of the techniques do not translate to our setting for mobile robots in complex and dynamic environments. We propose to handle these computational barriers by using a cascade of tractable models which successfully filter-out more and more unlikely pose configurations, allowing focus of computation resources on the most likely models.

Also, as part of this thrust, a subtask will be recognizing a small set of actions from "cooperative" subjects. The output of this task will be used in two ways: 1) action recognition for the purpose of communicating with the robot, and 2) to provide data for generalizing our prediction models to sequences of actions.

Adaptive Behavior Generation (3)

Based on the first two foundational approaches, cognitive-metric world model and semantic perception, we posit that we can produce a robot that has the ability to perceive and understand its environment and a world model in which to save and analyze that information over time. However, in order to be useful, the robot must be able to *do* something. Thus, we need a foundational capability for adaptive behavior generation. There are three fundamental types of behavior we wish the robot to possess:

- *Tactical mission behaviors* – this is the macro platform level of planning or behavior generation where the robot decides "what to do next" based on the current circumstances, i.e., METT-TC. Such decisions include changing sensor modes and platform location to provide better overwatch, adjusting speed and course to avoid a pedestrian, and sending a message with important surveillance results.
- *Dexterous manipulation behaviors* – this is the level of behavior where the robot interacts directly with the physical world in order to perform tasks such as picking-up and carrying objects, digging a hole, using tools, or opening a door.
- *Unique mobility behaviors* – this is the level of behavior relating to how the platform interacts directly with the physical world in order to achieve the higher-level mission goals. The robot may, for example, need to change its locomotion in order to climb a wall, navigate rough terrain, or simply get through the mud.

These three behaviors share planning paradigms at an abstract level. They all involve continuously deciding "what to do next" but in differing problem domains, timescales, and scale sizes. Therefore, we group them together conceptually; we anticipate much synergy among the efforts but also recognize the differences. Our fundamental approach is to focus upon the algorithmic aspects of behavior generation while using available approaches to realizing those behaviors through conventional actuation mechanisms.

Each behavior generation area is described in more detail below:

Tactical Mission Behaviors (3a)
To build robust tactical behaviors, we combine knowledge-intensive approaches from cognitive architectures with algorithmic approaches from traditional robotics. To bring the two disciplines together, we quantify, manage, and reduce inherent model uncertainty that arises in the interface between them. Typically, in modern robotic systems, planning is always done under the most likely hypothesis as uncertainty in perception estimates is massive, while decision-theoretic planning under uncertainty is computationally hard. We will develop principal approaches to planning under multiple hypotheses given by perception to support planning with uncertainty at the algorithmic and cognitive levels. Other methods to reduce uncertainty come from defining how the robot interacts with its environment. We devise *policies* for a robot to interact with an unpredictable human behavior to result in a desired outcome. Progress will be assessed via theoretical analysis of the performance guarantees we can provide as well as experimentation both in simulation and on a physical platform. We will test how our approaches to handling, reducing, and reasoning over uncertainty allow for a more robust behavior that focuses perception efforts and produces actions in such a way as to provide a higher level of robustness.

Dexterous Manipulation Behaviors (3b)
We seek a formal understanding and framework that will enable mobile systems to perform highly dexterous manipulation operations, including planning, perception, control, and user interaction. A goal of this research is to provide a foundation for whole body manipulation, where manipulation moves beyond traditional arms to include the abilities of the platform itself to assist with the task. The state of the art provides examples such as the manipulator twisting a door knob, and the platform reverses to pull open the door. To begin moving toward this goal, we investigate abstractions and algorithms for reduced dimensional planning for coordinated mobile manipulation and workspace exploration.

Unique Mobility Behaviors (3c)
The next generation of mobile robots must be able to maneuver through complex three-dimensional environments, including urban, mountain, jungle, and riverine terrain. This requires platforms to go beyond reliance on a single, fixed mobility configuration by exploiting multiple, innovative methods of locomotion. While some platforms of this nature exist in the state of the art, they fall short of their true potential due to a lack of intelligent motion planning and control. We apply learning techniques to current gait models to improve and discover new gaits during online execution of "fixed topology" robots. We also investigate the opportunities and challenges of reconfigurable robots, such as snakes, whose topology may be changed at runtime to meet task goals. The current state of the art focuses upon solitary gaits, allowing us to move beyond

by creating principles for adapting gaits to variable non-steady terrain as well as to transition between horizontal and vertical gaits.

Deep Learning (4)

The technical approaches described thus far provide the basis for a highly capable robot to perceive the world, reason about it, and take useful actions. However, the environment in which the robot operates is full of surprises, and the robot must be able to learn from them rather than repeating its mistakes. To plan explicitly for all possibilities is intractable, so we plan to program what is well understood and learn that which is not. Thus, we have defined a research thrust in learning. The thrust emphasizes deep learning in order to deal with the difficulties of the assignment problem often encountered in conventional learning.

We hypothesize that we can efficiently grow intelligence through intensive training with domain experts (e.g., Soldiers rather than researchers or engineers) and learning from other "smarter" robots, creating an intelligence that continues to expand and grow through new and diverse experiences. To enable this, we develop methods to map gestures and vocal instructions into actions that can be understood by the robot. The techniques that lead to good performance in deep systems of learned modules remain poorly understood, and we will investigate methods such as "no-regret learning" and the use of boosting to leverage existing supervised learning algorithms to handle more complex problems. We investigate generalizations of data sharing between different robots and sensor systems to enable robots to automatically tune themselves.

We perform extensive comparisons on a range of complex tasks that compare performance using fully specified, human programmed implementation of tasks within our system; pervasive learning using only local signals; end-to-end training of the system including using imitation learning to improve performance; and full end-to-end training leveraging both supervised training and self-supervision. We test the speed-up of planning on tasks from complex motion control to cognitive decision making to establish the benefits of enabling adaptiveness in planning.

Meta-Cognition (5)

With learning added to perceiving, reasoning, and acting, the set of foundational capabilities for autonomy is nearly complete. However, we must still lay the groundwork for teaming through technical thrusts in team cognition, transparency, and common ground. We can build additional teaming capabilities with this foundation.

Understanding HR Intra-team Cognition (5a)
In this thrust, we leverage expertise in team research to expand the science of teamwork to the realm of Soldier-Robot Teams. The tasks within this thrust provide data and procedures to give robots the basic cognitive functionality required for effective dynamic collaboration with humans. Our approach involves examining the creation, elicitation, and combination of mental models for both humans and robotic agents to understand the complexity of knowledge required to combine teammates into one cohesive unit. This research will be conducted in close collaboration with Intelligence in determining the best representation for shared mental models

of team structure, populating the SMMs, and dynamically maintaining these models during mission execution. To facilitate this collaboration, key members of research staff from CMU and UCF are members of both the Intelligence and HRI research teams. In the first six months of the project, we have developed a close cooperation between these, specifically with respect to our task H1: Team Shared Mental Models.

Transparency (5b)
In this effort, we address the limited ability to collaborate closely with human teammates and other robots as trusted agents, the limited ability to support distributed intelligence, and the limited ability to communicate effectively with humans and other team members. Our approach is to apply deep modeling to our cognitive architecture, bridge between human and robot cognition, develop techniques that provide introspective ability on the internal workings of robots, as well as enable them to expresses their own operations in terms easily understandable by teammates. We seek meta-cognitive techniques that build upon these introspective abilities to provide added ability for a robot to diagnose its own problems, continue to improve upon its performance, and generalize its behavior to new situations. We will assess progress by evaluating the end-to-end performance of our method, i.e., the degree to which the robot's actions, given some natural language command, match or satisfy human expectations.

Common Ground for Shared Situation Awareness (5c)
The primary objective of this thrust is to create representations of space that will enable communication between humans and robots. Metric maps, such as occupancy grids or coordinates of low-level features, are commonly used. However, more intuitive and more compact representations are required for communication between human and robot team members. This thrust includes collaboration work with ARL. To enable communication, we need to determine a network of traversable space (similar to road networks) plus semantic names for the landmarks to be labeled. Such landmarks are necessary both for a change of mobility (like stairs) and manipulation (elevator door) and for verification of route instructions ("turn right when you pass the restroom").

Teaming Capabilities (6)

The current generation of robots essentially considers the external world, including humans, vehicles, and other robots, as navigational issues rather than as team members, opponents, or part of the ambient culture. In contrast, the Army vision for future robotic systems calls for highly effective Soldier-Robot Teams where each part of the team understands the roles, responsibilities, and required actions of the others and has the capability to provide the communication necessary to make the team successful. Accomplishing this within a mission context, accepted military doctrine, and social norms of the society in which the Soldier-Robot Teams operate will be a major technical challenge but will provide a quantum leap in effectiveness and capability.

Our HRI research also seeks to enable the future robotic team members to communicate with their human teammates using multiple simultaneous communication modalities. This approach allows for (a) integrated situation awareness (SA), world model development, and command interpretation; (b) more natural interaction and communications redundancy; and (c) awareness

of human team member affect. More revolutionary, though, is our approach to integrating robots into the Soldier team structure, into social structures, and into societies. We propose innovations to achieve this futuristic vision by drawing on successful research in human team behavior, human factors, live/virtual/constructive (LVC) simulation, computer science, and neuroscience.

Collaborating Socially, Organizationally, and Culturally (6a)

In this thrust, we are examining multi-level collaboration issues with research aimed at social, organizational, and cultural factors that are required so that robots can be collaborating partners within Soldier-Robot Teams that operate in the real world. This research is fundamental to making it possible for robots to function effectively within human social situations. This thrust addresses three levels of social interaction in separate tasks: within a team, within a social environment, and within a culture. The best form for representing this information is very much a research issue and, as in thrust 1, will require close collaboration with related tasks within Intelligence to be effective.

Multi-modal Communication (6b)

In the second thrust, we will apply our understanding of how communication unfolds (explicitly and implicitly) in dynamic team contexts. Soldier-Robot Team communication in the dynamic team context poses unique challenges; our research seeks to overcome these challenges by taking advantage of all available modalities to both scaffold and augment communication. Our approach includes focused efforts in each prominent modality and a dedicated 6.2 effort into methods for test and integration of multi-modal communication in live and virtual environments. The ultimate goal is to facilitate collaboration between humans and robots at multiple levels. We are addressing the distinctive and complex issues created in socially, organizationally, and culturally charged situations across a series of inter-related tasks. The tasks within this thrust examine both explicit and implicit communication modalities expected to provide effective team communication in ways that are intuitive or intrinsic to humans. The focus on multi-modal and redundant modalities is expected to help overcome interaction problems intrinsic to acoustically noisy and visually challenging situations and to ensure the dynamic bi-directional communication necessary to realize shared team awareness. We include in this research an emphasis on extending modalities that traditionally have been considered line-of-sight (LOS) to be useful even when out of visual contact.

Distributed and Collaborative Perception (6c)

One key element of this thrust is determining how to make use of the volumes of raw data that low cost range sensors can produce. For example, in an indoor environment, it would be useful for the robot to be able to quickly detect relevant objects such as walls, doors, windows, tables, and chairs. As another example, in a hallway, closed doorways may lie flush with the walls but in the image, they differ in appearance. As part of this thrust, we will deliver novel algorithms for rapidly extracting and delineating planar surfaces from fused range and image data. The algorithms will take a 3D point cloud and an associated color image as input. The boundaries of the regions in the image will be used to interpolate and extrapolate the surfaces in order to provide a clearer picture of the layout of the scene.

3D environments impose challenges to localization, and the current literature focuses on approaches without significant efforts to use prior maps. However, operations take place most

times in areas with some prior map information or raw satellite photos. To use such intelligence consistently across humans and robots, we will build robust localization algorithms that can localize based on rough prior blueprints or airborne photos. Prior maps can be used to maintain localization for outdoor operation in GPS-denied urban canyons if one can match features from ground-level views to the map. In contrast to other related work, such as in the MAST CTA, we do not restrict ourselves to building metric or topological maps; rather, we aim to build higher level symbolic or "sketch" maps for the purpose of sharing with humans and other robots. This is in keeping with the general RCTA thrust to create, update, and share information at more abstract and cognitively useful levels.

Distributed Intelligence (6d)
Researchers have shown simple collaboration among heterogeneous teams of humans and robots using clean task decomposition and have applied the results to slow-paced tasks in relatively benign environments. To make collaboration truly effective, we need to address more sophisticated missions and missions that occur under conditions that are hard to predict while building trust in automation. We develop shared mental models to enable heterogeneous human-robot teams to work together with a minimum of communication. We investigate the communication constraints inherent to team collaboration. Leveraging such information, constraint-based planning algorithms show promise for allowing a robot to switch between in response to complexity to achieve improved results by using the solution which best fits the current scenario. We evaluate the approaches using metrics for percentage reduction in computing time, performance error between the high- and low-capability robots, and the utilization of the distributed resources compared to centralized fully informed solutions. Again, in contrast to other related work, such as in the MAST CTA, we do not restrict our representations to metric data; instead, we leverage the cognitive-to-metric world model construct to enable much more compact and efficient sharing of information. We specifically use shared mental models to modulate the sharing of information.

Autonomy Enhancements (7)

In addition to the Foundational Capabilities of Autonomy (1 through 5 in Figure 1-3) and the Teaming Capabilities (6), we have identified a set of needed enhancements to autonomy. While these are not foundational, they extend unmanned system capabilities in three areas:

- Sensing for perception and understanding
- Mechanisms for manipulation and mobility
- Scaling issues for autonomy

Each of these enhancements to autonomy is discussed below.

Sensing for Perception and Understanding (7a)
Closely coupled to the above two tasks of understanding the world through the processing of sensor data is the generation of that data through the sensing process. Research into new sensing technologies has consumed major resources for many decades, and therefore, this program primarily uses the outcomes of previous and ongoing work in sensor research. However, we believe a small amount of focused research in sensing specifically aimed at robotic autonomy is

warranted. Thus, we have an effort that addresses the need for advanced sensing by exploring basic phenomenology, specific challenging environmental conditions (such as dust, smoke, fog), and enabling sensors for legged motion and manipulation.

In active sensing, we investigate LWIR LADAR to provide enhanced 3D perception capabilities through dust, smoke, and obscurants in all-weather environments, critical to UGV applications. A study specifically focused on the particulars of UGV applications is warranted to quantify the advantages LWIR LADAR may have through obscurants. Also, as part of this thrust, we will extend the development of Spectral LADAR, which allows object recognition using spectral signatures in addition to 3D spatial information. In this effort, we build upon a large body of prior work in multi-spectral sensing which is able to identify material types represented in single pixels and, thus, provide shape-independent object classification.

Another area of this thrust is in the use of smart materials systems to incorporate microarrays of pressure sensors into the "feet" of legged robots and in robot grippers as well as micro-electromechanical systems to design embedded temperature and moisture sensors for terrain classification and to integrate these with vision sensors to classify terrain for a variety of legged and wheeled vehicles.

Also, as part of this thrust, we are collaborating with ARL researchers in the areas of MEMS LADAR, acoustics, and very small radars. The MEMS LADAR is capable of real-time 3D images with high fidelity at frame rates and ranges suitable for SUGV application. The acoustic work includes audio classification.

Mechanisms for Dexterous Manipulation and Unique Mobility (7b)
We have previously described foundational thrusts in low-level control and behavior generation for both dexterous manipulation and unique mobility. In order to better realize those behaviors, we have defined a mechanisms thrust that encompasses three efforts: in mechanisms for manipulation, in mechanisms for mobility, and in next generation actuation approaches for both mobility and manipulation.

We explore manipulation systems that take advantage of increased range of motion afforded by a highly articulated mobility base in pursuit of whole body manipulation. Using an existing and readily available experimental platform, the DARPA BigDog, we apply the research in basic manipulation behavior in order to seek a control system that can perform useful manipulation tasks with a high degree of freedom system. These tasks, such as pushing, pulling, lifting, and throwing, are beyond the state of the art. We follow two paths to accomplish this. First, we pursue "legipulation," a bio-inspired approach where a legged system uses one of its legs to manipulate the environment. Second, we will integrate a preliminary manipulator arm onto a BigDog to investigate behaviors with full-body articulation while maintaining balance. We assess these capabilities using success rate and speed of task completion.

To investigate mechanisms for unique mobility, we take advantage of the Canid platform under development though a collaboration between UPenn and ARL. Building upon these topics, we investigate a hybrid framework for selecting, mixing, and transitioning between gaits at runtime.

These topics are then assessed using metrics of locomotion power efficiency, velocity, efficiency per unit weight, and grasp versus release force.

Finally, we investigate potentially revolutionary approaches for next generation actuation of manipulation and mobility mechanism. Actuators based on smart materials have great potential to transform robotic systems by improving the strength-to-weight ratio, speed, range of motion, compactness, efficiency, controllability, and reliability of manipulators, legs, sensors, and other robotic components. The state of the art technologies with respect to robotics include ferroelectric materials, magnetostrictive compounds, shape memory alloys, and dielectric elastomers. These materials have constraints that require special attention during the design process; therefore, synergistic research on material characterization, model development, and development for robotic platforms is critical. To overcome these challenges, we develop active materials for legged robotic platforms. With electro-active elastomeric materials, we expect to enable robotic limbs that can change their shape, stiffness, and potentially viscoelasticity with an applied electric field. We will investigate how electrically and thermally activated materials can be utilized to create passive mechanical joints with variable stiffness and damping. Smart structures utilize shape-changing actuation, simultaneous sensing, and real-time material property control for dynamic adaptation and superior maneuverability. For each material investigated, we will quantify the relevant materials properties and their response to stimulation and compare their capabilities to standard actuation techniques. We will also assess difficulties in and progress toward integration into robotic structures

Scaling Issues for Autonomy (7c)
Robotic systems are likely to span a wide range of scale sizes. The largest may be at the scale of main battle tanks and wheeled fighting vehicles. Smaller platforms of TALON-size may continue to perform EOD and similar roles. Even smaller platforms could climb walls and enter small openings inaccessible to humans. Yet the utility of all of these platforms will be greatly enhanced if they possess a significant degree of the foundational autonomous capabilities we have outlined. Thus, autonomy needs to exist across a range of scale sizes as wide as possible. Our final enhancement thrust is, therefore, focused on the ability to realize autonomous capabilities across a wide range of scales. We are undertaking three approaches to scaling. The first is to develop more computationally efficient algorithms. The second is to use robots as teachers, where a larger and more capable robot takes on the role of teaching a smaller and less capable one. The third approach involves pooling resource across a team of robots, thus compensating for deficiencies in the less capable ones.

Assessment of Integrated Research

Our plan for measuring progress toward autonomy over time is built around our integration and assessment (I&A) plan. As a collaborative fundamental research effort, the Robotics CTA's assessment process differs from that of traditional system development efforts. Instead of building a system to meet a particular performance specification, we deliberately undertake high-risk basic and applied research that may ultimately result in breakthrough technologies. Instead of managing development to meet pre-defined goals, we assess our research against performance benchmarks to evaluate how well that research stands to enhance or even revolutionize robotics and related disciplines.

Our integration and assessment is performed by a team consisting of research integrators and support staff, an assessment team, and I&A management. The I&A management team consists of the four Technical Area Leads, the Integration Lead, and the ARL I&A Lead. The I&A management team plans integration events, defines assessment methodologies, and conducts quantitative assessments. The assessment team prides quantitative and objective results based on accepted experimental practices.

Our assessment protocol is based on a two-stage approach. The first stage consists of task- and/or subtask-level assessments of stand-alone research outcomes. These assessments will typically be conducted by the researchers, but they will be reported to and monitored by the I&A team. Task and subtask assessments are provided in quarterly reports and at other times as requested. An example of an individual assessment is the measurement of precision/recall performance for object or activity detection on a given dataset. Results of such individual task assessments will help determine which research outcomes are ready for integration. Task-level research that is producing demonstrated results beyond the state of the art creates a "push" to be included in the integrated research described below. As described in Section 6, we also define integrated capability goals each year that reflect expected outcomes. Thus, we also create a "pull" to set an expectation for research outcomes. The I&A management team considers both push and pull to decide which research outcomes are suitable for integration and assessment.

The second stage of assessments focuses on integrated capabilities that result from bringing together results from multiple research tasks. At this second stage of assessment, we conduct a series of experiments which we call integrated research assessments (IRAs). Each IRA combines two or three outcomes from research thrusts to achieve a capability from the think-look-talk-move-work spectrum. Similar assessments will be repeated and extended over time to provide regression testing and integrate improving technologies. The assessments will involve formal experimental design in collaboration with the government with reported results. They will include elements of both modeling and simulation and laboratory experiments appropriate for the assessment of basic and applied research outcomes. Multiple IRAs are planned for 2011 and 2012 and are described in more detail in Section 6. The exact timing of the IRAs will depend upon the status of research outcomes. The pace of IRAs will likely increase as more results become available during the course of the program.
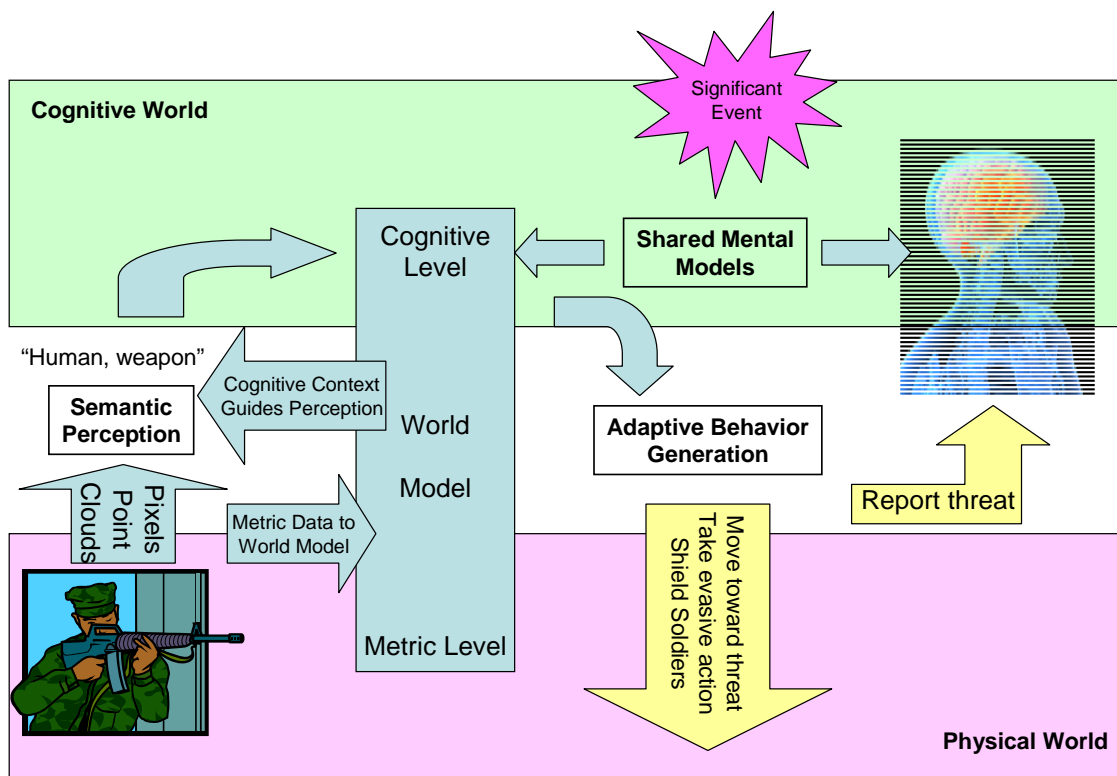
Figure 1-7: For our first assessment, we integrate elements of semantic perception, shared mental models, and adaptive behavior generation.

As an illustrative example of an early integrated research assessment, we plan to combine elements from three research thrusts into a basic "look-think-talk" capability. In this assessment, an autonomous system observes a series of actions. Its semantic perception capability populates the cognitive level of the world model with information about observed activity. This perception uses state of the art "bottom-up" techniques but is guided by contextual information from the world model's long- and short-term memories. In this case, the system observes a Soldier with a weapon. Based on a shared mental model of what the current mission is, the system identifies one or more activities as salient events. In this case, the Soldier may be a member of the unit or may represent a threat, depending on context and prior knowledge as well as the immediate visual evidence. If an event is important enough, it triggers a number of possible actions, such as making a report, moving to gain a better vantage point, taking evasive action, etc. This integrated assessment combines semantic perception, shared mental models, and adaptive behavior generation, all of which are mediated by the new world model. Of course, the first integration of these capabilities is likely to reveal significant shortcomings. Thus, we will subsequently add improved capabilities as well as assess more complex situations.

In our assessments, we apply established principles of scientific experimentation. The assessment plan includes the platforms, sensors, human participants, and simulation tools used, along with a specification of datasets to be collected. The datasets will include data to be sequestered as well as data for possible posting to the broader research community.

For the above "look-think-talk" example, the overall experimental hypothesis is that an autonomous sensing system can observe a wide range of human activity, recognize a subset of that activity as salient to the mission and conditions at hand, and then report those salient observations in a human-understandable form such as, "A man in a green coat just left the safe house." If that activity actually occurs and the system reports it correctly, that constitutes a true positive. If the system does not report salient activity, we have a false negative. In this example, the human activity which is sensed constitutes the independent variable, while the reported message is the dependent variable.

In addition to the end-to-end performance, we will also examine the individual components of the experimental system. In this case, the components are:

- Activity detection
- Salience assessment
- Activity reporting

For the integrated end-to-end capability to work correctly, the activity of a man departing a building must be detected, that activity must be correctly assessed as salient to the mission, and a suitable report must be constructed. If the overall system fails in a given case, we need to understand what the cause of failure was. Thus, we will separately assess each of these capabilities, which can be tested both in an integrated form and individually. For example, we can assess the activity reporting module by providing many examples of output that could come from the prior stages and evaluating how well it constructs messages that are both accurate and readily understood by humans. Similarly, the salience assessment module can be fed a variety of activity detections across a variety of mission contexts. A given activity may be salient in one context but not another.

This type of integrated assessment is quite challenging because we are conducting fundamental research: we seek to create capabilities that do not exist, rather than simply making incremental improvements to existing capabilities.

Relationship to Military Needs

In order to focus on the development of capabilities for ground autonomous systems over the next decade, it is helpful to consider specific ground robotics capability needs that have been identified by Army and Joint Service Future Operating Capability documents, Army S&T Master Plans, COCOM Priorities, Defense Planning Guidance, Prioritized Capability Lists, Warfighter Capability Gaps, Unmanned Systems Integrated Roadmaps, and feedback from current conflicts. These resources use a variety of methods to categorize requirements. For example, COCOM requirements are grouped into JCAs including Battlespace Awareness, Command and Control, Force Application, Logistics, and Protection, while Army Documents such as FM 3-0 Operations consider Movement and Maneuver, Intelligence, Fires, Sustainment, Command and Control, and Protection. TRADOC Pamphlet 525-66 provides perhaps the most comprehensive set of requirements, detailing the mapping between Joint Functional Concepts and Force Operating Capabilities. One way to frame our discussion is in the context of four Joint Functional Concepts most relevant to unmanned systems: Battlespace Awareness, Force Application, Protection, and

Focused Logistics. Other examples of military needs come from the operational experiences of commanders in the field. For example, LTG Rick Lynch, former commander of III Corps at Ft. Hood, has proposed four Joint Operational Needs Statements (JONS) for robotic systems: Route Clearance, Persistent Stare, Robotic Convoy, and Robotic Wingman. These specific robotics-related needs map well to the more general Joint Functional Concepts as follows:

- Route Clearance → Protection
- Persistent Stare → Battlespace Awareness
- Robotic Convoy → Focused Logistics
- Robotic Wingman → Force Application

If we consider the missions of EOD and Route Clearance, ISR, logistics, and combat support, the following fundamental capabilities are important:

- For all mission types – Receive, understand, and acknowledge instructions/orders. This underlying capability is required in order to initiate any mission.
- For route clearance/EOD – Move to a verbally (semantically) described location, find the described object(s) of interest, inspect and/or manipulate the OOI, optionally transport the OOI. This capability is a leap-ahead from the current state of the art where, for example, EOD robots are tele-operated or, at best, perform simple waypoint navigation to go downrange to the vicinity of a suspected IED. In our vision, the robot recognizes and proceeds to the suspected object, inspects it, and even manipulates with permission and possibly some supervision.
- For persistent stare – Move to an appropriate location, survey an area of interest (AOI), recognize salient activity, construct a human-understandable report about the activity, and answer follow-up questions about the activity.
- For convoy/logistics – Identify and locate needed supplies, load the supplies, move to where the supplies are needed, and unload the supplies as needed and/or ordered.
- For robotic wingman – Move in tactically correct manner with unit, provide overwatch as needed, and provide useful response to hostile actions.

The above capabilities overlap a great deal in the technical challenges they impose, and they also differ in important ways. To better understand these differences and similarities, it is useful to decompose each mission capability into a sequence of mission elements. Figure 1-8 summarizes this decomposition for our four mission types.

Matrix of Autonomous Capabilities by Mission				
TRADOC Pamphlet 525-66	**Protection**	**Focused Logistics**	**Battlespace Awareness**	**Force Application**
JONS (Lynch)	**EOD/Route Clearance**	**Convoy**	**Persistent Stare**	**Robotic Wingman**
Receive, understand and acknowledge orders ("Think, Talk")	Receive orders, e.g., verbal or text, translate into internal WM representation, request clarification if needed, acknowledge/restate orders, answer questions if posed			
Report to unit, configure for mission, form up ("Move, Talk")	Load/request needed material for mission, e.g., tools, marsupial robots	Provision vehicle for mission, configure sensor payload, obtain relevant map data	Provision vehicle for mission, obtain relevant map data	Provision vehicle for mission, configure weapon payload, obtain relevant map data
Proceed to next specified location -- phase line,… ("Move, Think")	Move downrange to suspected IED site(s). Area has been cleared for EOD mission	Move in traffic to combat outpost location, either as part of manned unit or wholly unmanned	Move to suitable location for surveillance, taking account of both vantage point and concealment	Move to next objective in tactically correct manner, providing overwatch as needed
Maintain SA and comms -- react to unforseen events ("Look, Talk")	Look for additional IEDs, react to IED events and possible related small arms attacks	Adaptively replan to disruptions to planned route, respond to IED or other events	Recognize time-critical ISR events and report/react accordingly	Perform 360 SA, recognize threats to self, leader and unit, report salient information
Perform specific mission tasks ("Work, Think, Talk")	Inspect possible IED objects. Manipulate and move objects as necessary	Select, pick up and load/unload needed materials	Observe activity and report salient events	Provide 360 SA, provide and accept targeting information, engage with permission
Conclude mission and return to unit ("Move, Think")	Return to EOD or Route Clearance unit, unload tools as needed, receive/await new orders	Report to commander at logistics mission destination, receive/await new orders	Return to ISR unit commander/HQ, provide more data as requested, receive/await new orders	Return from mission with manned leader, receive/await new orders

Figure 1-8: Matrix of autonomous capability challenges by mission type.

Our research portfolio is driven by capabilities over near-, mid-, and far-term time horizons. Thus, we will assess progress in capability over time. As an example, we consider the integrated capability elements of the EOD mission listed in the first column of Figure 1-9. This mission information was collected based on the interaction of RCTA technical managers and researchers with several EOD Soldiers in late 2010 and early 2011.

- Mission Preparation and Initiation
- Transit Downrange
- Mission Situational Awareness
- Perform EOD Task
- Return to Designated Location

For each of these mission elements, we have worked with the Naval Explosive Ordnance Disposal Technology Division (NAVEODTECHDIV) to identify specifically how RCTA can impact EOD unmanned operational capabilities (UOCs) in the 3-, 5-, and 7-year timelines. Figure 1-9 summarizes UOCs we envision by EOD mission phase and lists specific foundational

elements of autonomy that apply to each element. Here, we summarize the relationship of each of the EOD mission elements both to the RCTA vision outlined in the beginning of this section and to the foundational capabilities of autonomy we are developing to achieve our vision.

Mission preparation. Currently, EOD robots are simply devices that must be operated step-by-step to achieve a mission. They must be configured by operators, transported to where they are needed, and then tele-operated to go downrange and perform the mission. Our vision is to endow the robot with the ability to receive orders or instructions, configure itself for the specified mission, and position itself to begin executing the mission. To achieve this vision, RCTA will give the robot a measure of adaptive tactical reasoning using the cognitive-to-metric world model so that it understands what an EOD mission is, can receive and acknowledge orders, and knows how to prepare itself for the mission.

Transit downrange. Instead of the current practice of tele-operating the robot(s) downrange, our vision is for autonomous movement to the appropriate site downrange. For example, the robot may be instructed to go to a map reference or to a semantically described point, e.g., "the rubble pile behind the blue track." It then navigates through the environment, which may include rugged terrain and heavy urban clutter. It can also transition from outdoors to indoors if needed, possibly using new mobility mechanisms to negotiate stairs or other obstacles. This tactically adaptive movement requires the understanding of terrain, objects, and activity in the environment. We achieve this through parallel efforts in semantic perception and adaptive mobility behavior instantiated in our new world model.

Mission situational awareness. During current EOD missions, SA is achieved only through human "eyes on." This may be direct human vision or vision through robot sensors. The latter requires Soldiers to be "heads down," which is not conducive to overall SA. Thus, we strive for our vision in which the robot builds its own understanding of the environment – both in the immediate vicinity of the object of interest (OOI) and in the broader environment where other IEDs or other forms of threat may exist. The goal is shared SA among both human and robotic team members, combining both the "look" and "talk" capabilities in our vision. The RCTA will achieve this "common ground" through integrated research in semantic perception and meta-cognition.

Perform EOD task. This mission element is more specific to the EOD mission than other elements. It is currently performed through tedious tele-operation that requires very experienced operators. There is now some degree of automation through memorizing some manipulation steps. However, we aim to develop the basis for a much higher level of autonomy where the robot can, for example, search through a pile of rocks and other debris with little or no supervision. To do so, it must understand the objects in its environment as well as its own ability to interact with that environment. The capability will come from our research in adaptive behaviors as they relate to manipulation tasks. Furthermore, our research looks at joint manipulation and mobility issues to enable a robot, for example, to brace itself to move a heavier object than it could otherwise move. Also, instead of simple imitation learning of simple manipulation steps, we are exploring deep learning – in combination with semantic activity understanding and a hierarchical world model – to enable the learning of more complex tasks.

Return to unit. The final mission element is largely a combination of the previous elements. It requires the robot to exercise judgment about when to return, whether to seek guidance, what it encounters on the way back, and how it should communicate with its unit upon return.

EOD Mission Element (Foundational Capabilities)	Near-term ~3 years	Mid-term ~5 years	Far-term ~7 years
Mission Preparation and Initiation (1, 3a, 5a, 5b)	Receive and acknowledge simple mission orders, enabling more "launch and forget" operation.	Receive and acknowledge more complex orders; assemble inside host vehicle.	Deploy (assemble and exit host vehicle) and respond adaptively to changing tactical situation, e.g., hostile action.
Transit Downrange (1, 2a, 2b, 3c, 7b)	Navigate to a single or series of grid coordinates in rugged environments (moderate clutter, positive/ negative obstacles, day/night, GPS/GPS-denied, indoor/outdoor).	Navigate to described location in rugged terrain with heavy clutter (positive/ negative obstacles, day/night, GPS/GPS-denied, indoor/outdoor).	Navigate to described location in rugged terrain with mobility challenges and heavy clutter (positive/ negative/ water obstacles, day/night, GPS/GPS-denied, indoor/outdoor).
Mission Situational Awareness (1, 2a, 2b, 5c, 7a)	Generate 3-dimensional virtual environment annotating user selected locations/items of interest, allowing for 3rd person platform operation and scene analysis (linear distance measurement).	Generate high-res 3-dimensional virtual environment, updated in real-time annotating user selected locations/items of interest, allowing for 3rd person platform operation and forensic scene analysis (distance, angles, volumes).	Generate high-res 3-dimensional virtual environment, updated in real-time, annotating user selected and automatically identified locations/items of interest, allowing for 3rd person operation and forensic analysis.
Perform EOD Task (1, 2a, 2b, 3b, 4, 6, 7a)	Provide user with haptic feedback and intuitive control in a high latency environment. Simultaneous operation of multiple robots by a single operator.	Automate dual arm task based manipulation operations. Automate passive cooperative behaviors with multiple robots. Robots do not simultaneously physically interact with objects.	Learn simple manipulation tasks from Soldier training. Automate active cooperative behaviors with multiple robots. Robots do simultaneously physically interact with objects, e.g. pick-up and carry a payload.

EOD Mission Element (Foundational Capabilities)	Near-term ~3 years	Mid-term ~5 years	Far-term ~7 years
Return to Designated Location (1, 2a, 2b, 3a, 3c, 4, 5b, 7a, 7b)	Return efficiently to user-defined location using multi-sensor data.

Upon loss of communications, return efficiently to user-defined rally location or until communications are re-established | Return efficiently to user-defined location using multi-sensor data.

Upon loss of communications, return efficiently to user-defined rally location or until communications are re-established

Demonstrate increased efficiency from near-term years. | Rejoin unit on-the-move using own/mixed initiative; enter host vehicle autonomously. |

Figure 1-9: EOD Mission Integrated Capability progress over time.

In order to enhance EOD mission operational capability as indicated in Figure 1-9, we must address the technical barriers to those capabilities. Therefore, we have identified those barriers, examined them in light of our planned research, and verified that our research is targeted to surmount those barriers. While this discussion uses the EOD mission as one example, similar analysis shows that the research is also targeting the other relevant mission areas.

Conclusion

The RCTA program has established a vision for a new level of autonomy in ground robotics. That vision includes five key capabilities: (1) adaptive tactical reasoning, (2) focused situational awareness, (3) efficient and proactive interaction with humans, (4) safe, secure and adaptive movement, and (5) interaction with the physical world. We anthropomorphically describe these capabilities as "Think," "Look," "Talk," "Move," and "Work."

Currently, there are major technical barriers which lie in the way of this vision. At the heart of the problem is the need for a world model, which can be instantiated on robots and which represents the range of entities, spatiotemporal scales, and abstractions that must be reasoned about. There is also a shortcoming in a robot's ability to sense the environment and understand it at a semantic level; this ability is needed in order to populate the world model with new and relevant information. Robots have only a rudimentary capability to plan behavior, and their planning is focused almost entirely on navigation and is brittle even in that limited domain. Furthermore, cognitive concepts such as mission, enemy, troops, terrain, time, and civilians (METT-TC), which are at the heart of a Soldier's planning process, are entirely missing in current robots. Finally, the robot behaviors that have been achieved thus far have been

programmed for specific applications; any correction or new application requires re-programming. Therefore, without a robust learning capability, our vision of adaptable robots cannot be achieved.

To overcome these barriers, we have identified five major fundamental research thrusts which will lay the foundation for a new level of autonomy. These foundational thrusts are:

- Cognitive-to-metric world model
- Semantic perception
- Adaptive behavior generation
- Deep learning
- Meta-cognition

These thrusts provide the basis for *individual* robotic autonomy. We also build upon that foundation to enable *collaborative* autonomy among a set of robotic and human teammates. We have identified a small set of enhancements that will make autonomy much more effective at both the individual and collaborative levels. Our current plan invests over 60% of our resources in the foundational thrusts listed above; it invests smaller, approximately equal, efforts in collaboration and in selected autonomy enhancements.

We have built a fundamental research portfolio to implement our technical thrusts. The portfolio is structured in four technical areas: Intelligence, Perception, Human-Robot Interaction, and Dexterous Manipulation and Unique Mobility. The following four sections of this Annual Program Plan detail the tasks that comprise our research portfolio in each of those technical areas. The sections include, at the task and subtask levels, a description of the current state of the art, how the currently proposed research advances the state of the art, and the metrics that will be used to assess progress. Each section also contains specific research plans and objectives for 2011.

In order to harness this fundamental research to achieve our vision, we must integrate research outcomes from across the technical areas into cross-cutting technical capabilities. The final section of this document describes our approach to integrating research outcomes and assessing the resulting integrated capabilities. These capabilities, by design, align very closely with the capabilities in our vision for robotic autonomy. Thus, our integrated research assessments will directly measure our progress toward achieving the RCTA vision.