Artificial Social Intelligence for Successful Teams (ASIST)

Joshua Elliott
Program Manager, DARPA I2O

Proposers Day

March 14th, 2019

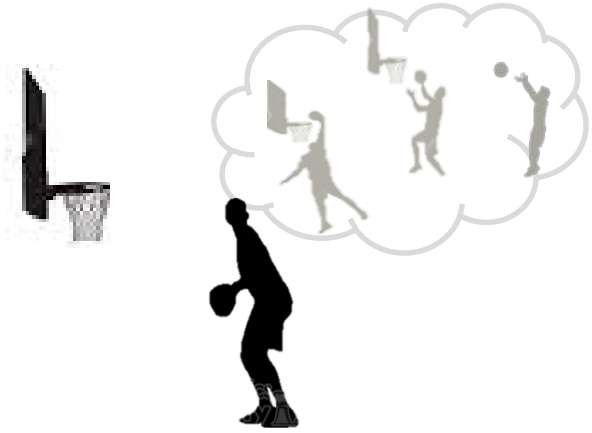
Objective and Impact

ASIST will develop foundational AI theory and systems that demonstrate the basic machine social skills needed to infer the goals and beliefs of human partners, predict what they will need, and offer context aware interventions in order to act as adaptable and resilient AI teammates.

What makes humans good teammates?

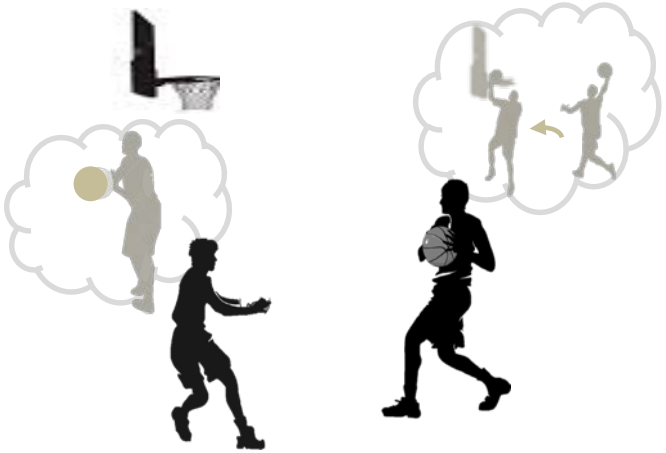
Mental Models of Environment

Humans can build robust mental models of their environment



Mental Models of Others

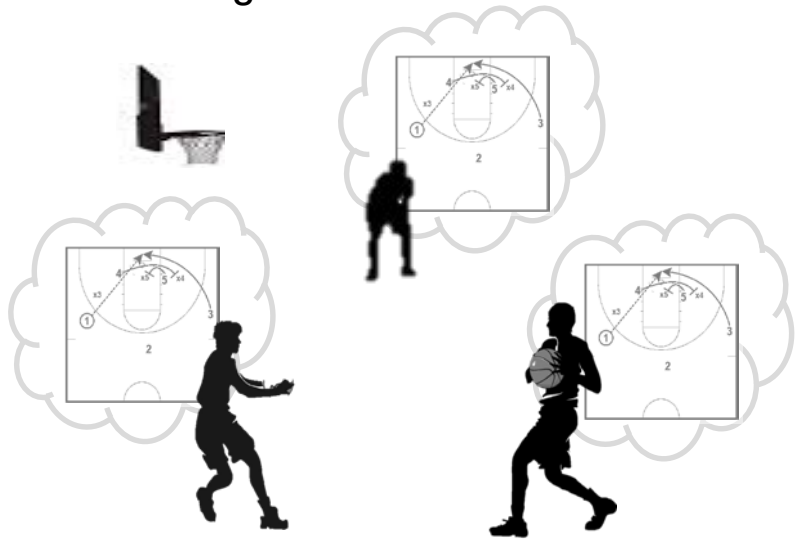
Humans can infer, from observed actions and context, the mental states of other humans



Theory of Mind

Shared Mental Models

To perform in teams, humans use experience and training to align their Mental Models



Social Intelligence

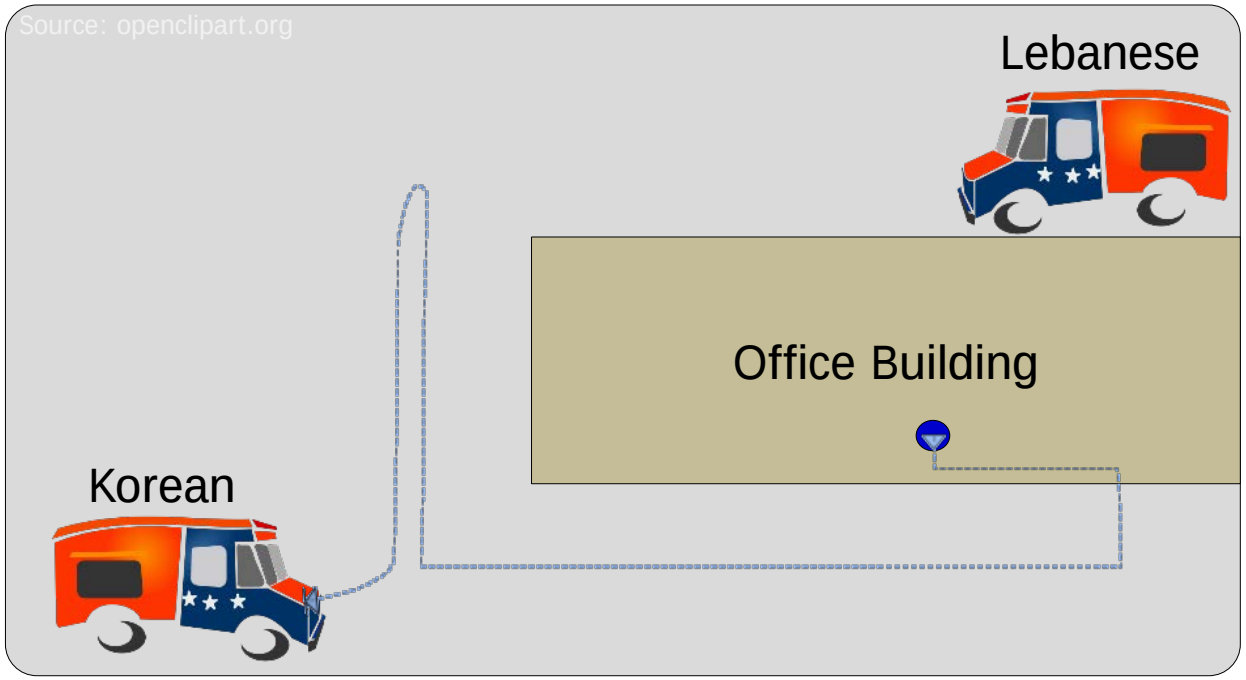
Machine Theory of Mind

Three possible food trucks

- Lebanese
- Korean
- Mexican

Only two parking spaces

What is the agent's favorite food truck?



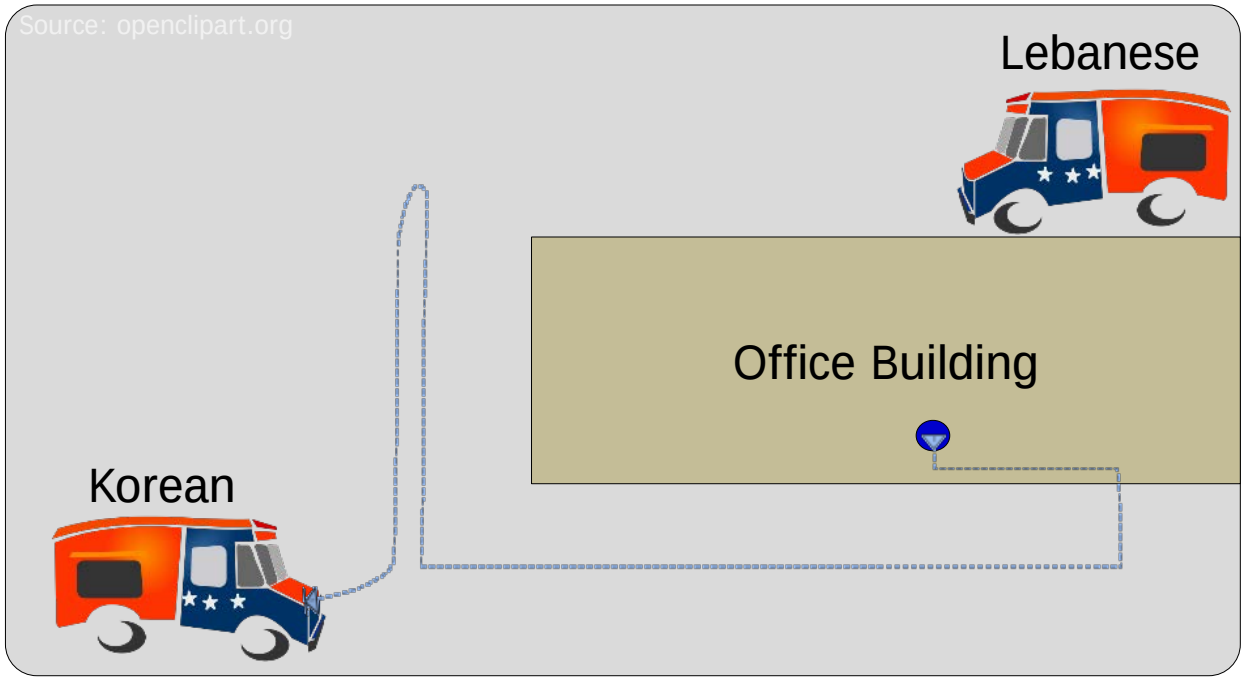
Machine Theory of Mind

Three possible food trucks

- Lebanese
- Korean
- Mexican

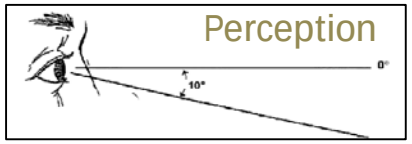
Only two parking spaces

What is the agent's favorite food truck?

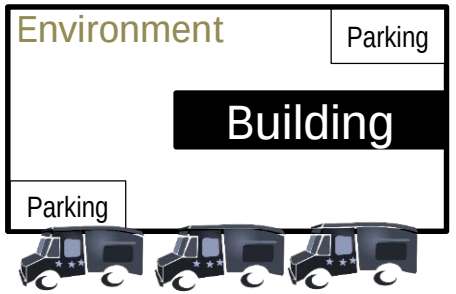


Bayesian Theory of Mind

Ad hoc Representations of the Agent
- Perception
- Reward structure
- Rational choice



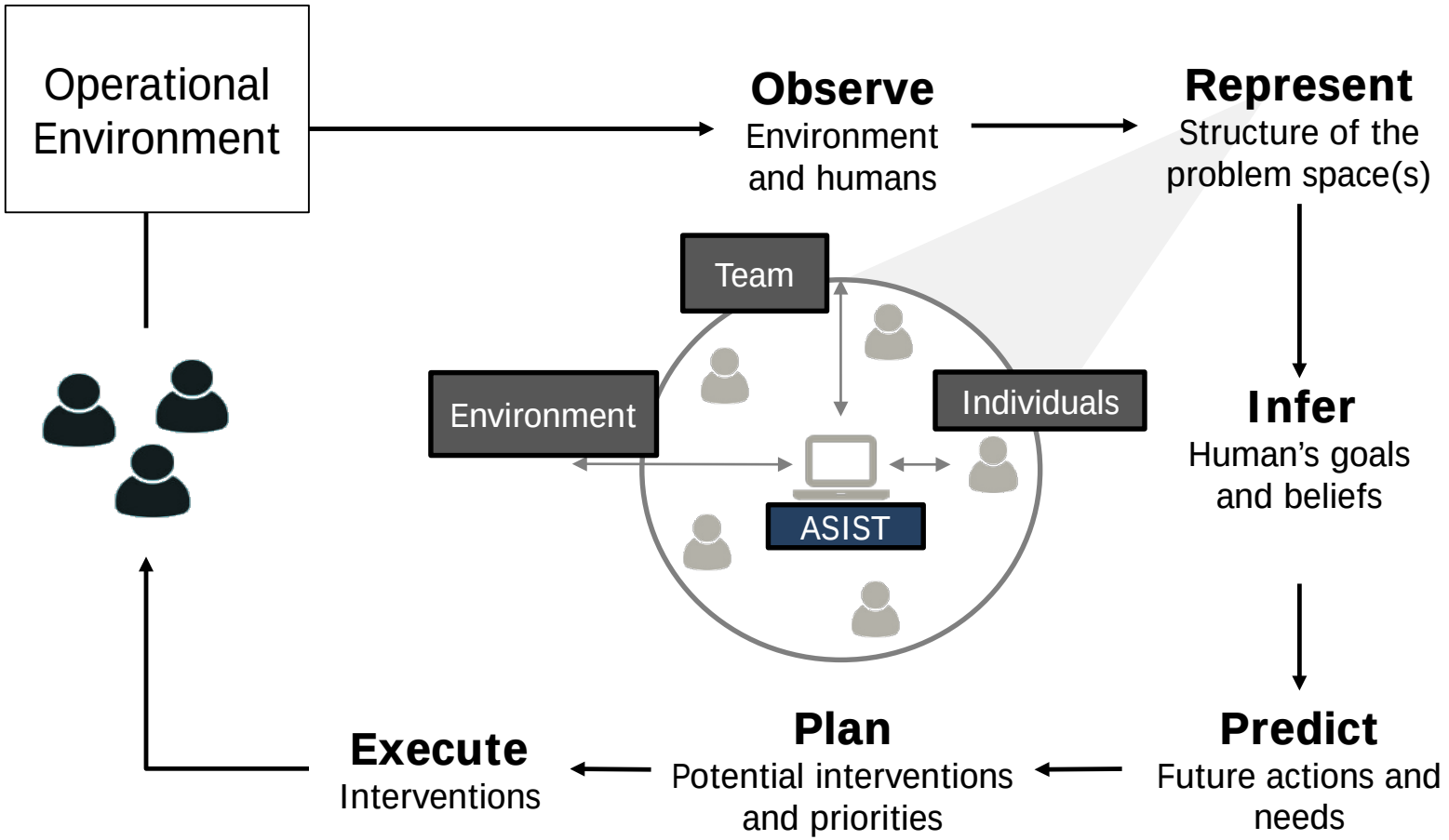
Ad hoc Representations of the Environment
- Objects
- Properties
- States



How is it done today and limitations?

	Bayesian Theory of Mind	Advances in Bayesian Theory of Mind	Learning-to-Learn	Formal Frameworks	ASIST
Operational Complexity	Low	Moderate	Low	High	**High**
Adaptability	Low	Partial	High	Low	**High**
Observations	Perfect/ Single-channel	Noisy/ Single-channel	Noisy/ Single-channel	Perfect/ Multi-channel	**Noisy/ Multi-channel**
Representations					
Environment	Ad hoc	Partly principled	Learned	Ad hoc	**Learned/Transferable**
Individuals	Partly principled	Partly principled	Learned	Ad hoc	**Learned/Transferable**
Group	No	No	No	Synthetic team	**Hybrid Teams**
Infer and Predict	Slow	Moderate	Scalable	Classifying/ Fast	**Both/ Complex**
Humans observed	No	Yes	No	Partial	**Yes**
Interventions	No	No	No	Yes	**Yes**

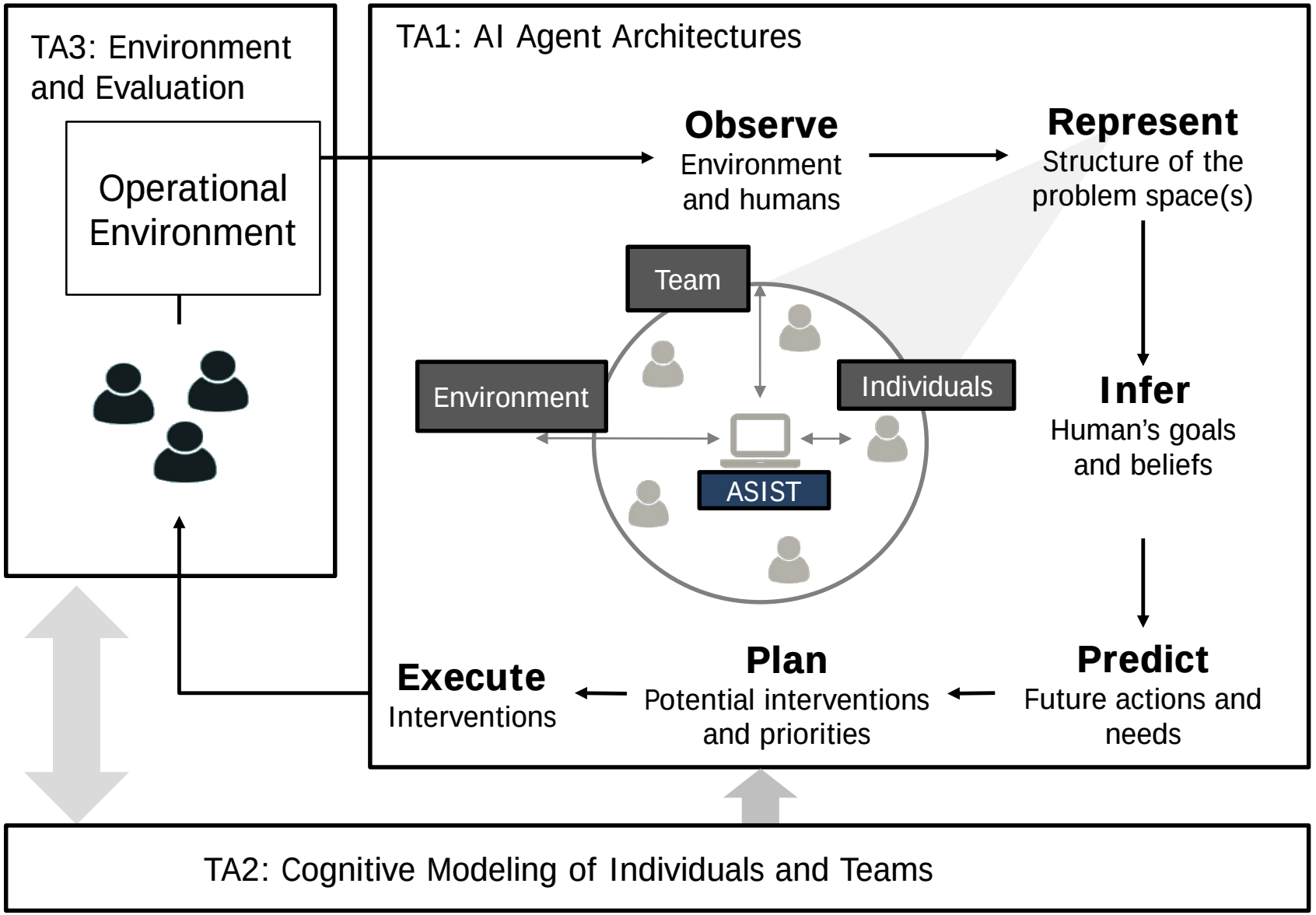
ASIST – Vision



ASIST agents will...

- Operate in increasingly complex and specialized environments
- Adapt to unexpected perturbations
- Contribute to a revolution in cognitive modeling for human-machine teaming

ASIST – Structure



TA1: AI Agent Architectures

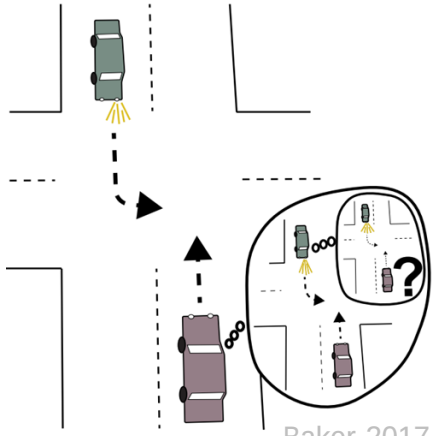
- The goal of TA1 is to develop and implement AI architectures for human-machine teaming

- Complexifiers will be added in stages over the course of the program:
  - By the end of Phase 1 TA1 teams must demonstrate effective AI agent architectures that exhibit key aspects of machine social intelligence in interactions with a single human (e.g. MToM)
  - In Phase 2 this will be extended to teams with multiple humans
    - The goal is an agent that can participate in teamwork, not (necessarily) taskwork
      - E.g. demonstrating the ability to participate in the team's shared mental models
  - Phase 3 will introduce team of humans with specialized roles and skills working collectively toward a complex objective

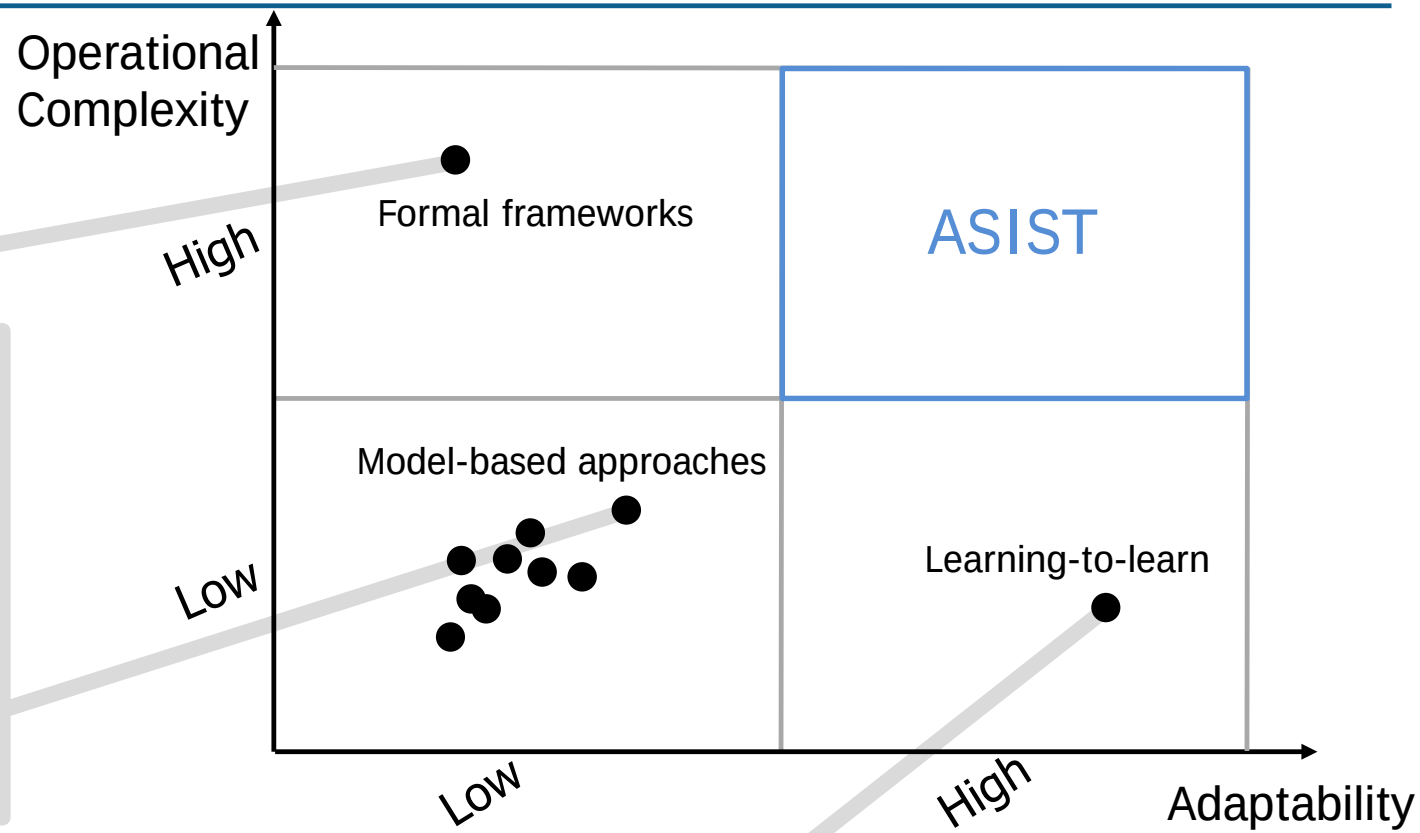
- The AI architectures must include all of the following:
  - Use of generalizable approaches to represent teams and individual partners
  - Demonstrate real-time scalable inference and prediction of human partner actions
  - Ability to scale to increasingly complex operations
  - Ability to effectively handle multiple classes of perturbations

TA1 Architectures: Operational complexity and adaptability

Formal frameworks
- Domain language that includes formal representations of team norms
- Brittle, but applied in complex settings



Real-world prototype of Bayesian ToM



Also: Low-shot imitation learning, demonstration and preferences, latent learning, successor representations, analogical learning, principled models (e.g. physics or psychology), neural networks, real-time approximate inference, constrained models that break the problem up into small pieces, hybrid approaches, ...

Meta-learning approaches
- Multi-step neural net that learns the necessary representations of environment and agent
- Demonstrated in simple 2D-gridworlds, but approach is scalable and adaptable

TA2: Cognitive Modeling of Individuals and Teams

- The primary objective for TA2 teams is to formulate testable hypotheses about human and human-machine teaming and social cognition
- Expected topics of inclusion are:
  - Applicability to hybrid teams of successful theoretical frameworks for human team performance
  - Variability in team performance
  - Norms of behavior and communication
  - Trust
  - Nested mental models
  - ……

Do the successful constructs of human cognition and teaming apply to human-machine teams?

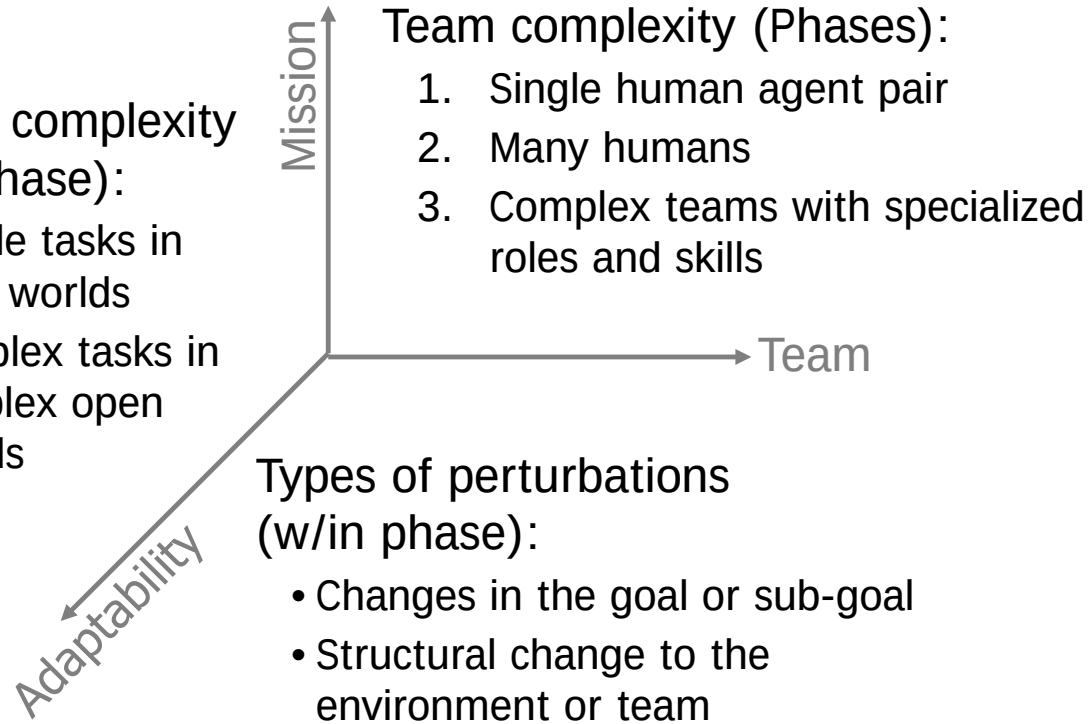
- TA2 performers will pre-register hypotheses in a repository managed by TA3
  - TA2 teams will collaborate with TA3 to develop new measures and experiments to test these hypotheses, and will analyze and publish the resulting data and conclusions
    - Strongly emphasized is the requirement for rigorous data analysis procedures
- TA2 performers are strongly encouraged to work closely with TA1 and TA3 teams throughout to understand what will be experimentally available and to inform the design of agent architectures and experiments

TA3: Environment and Evaluations

The TA3 performer will contribute as follows:

- Develop and implement a scalable testbed and challenge environment(s)
  - Include both passive and active experimental designs
- Coordinate with performers to assure protocols and APIs are stable and functioning prior to evaluations and dry-runs
- Organize and execute tech evaluations and large scale experiments on human and hybrid teams
- Organize and execute milestone events including evaluations, technical meetings, and a technology demonstration

Dimensions of Complexity



Mission complexity (w/in phase):

- Simple tasks in open worlds
- Complex tasks in complex open worlds

Team complexity (Phases):

1. Single human agent pair
2. Many humans
3. Complex teams with specialized roles and skills

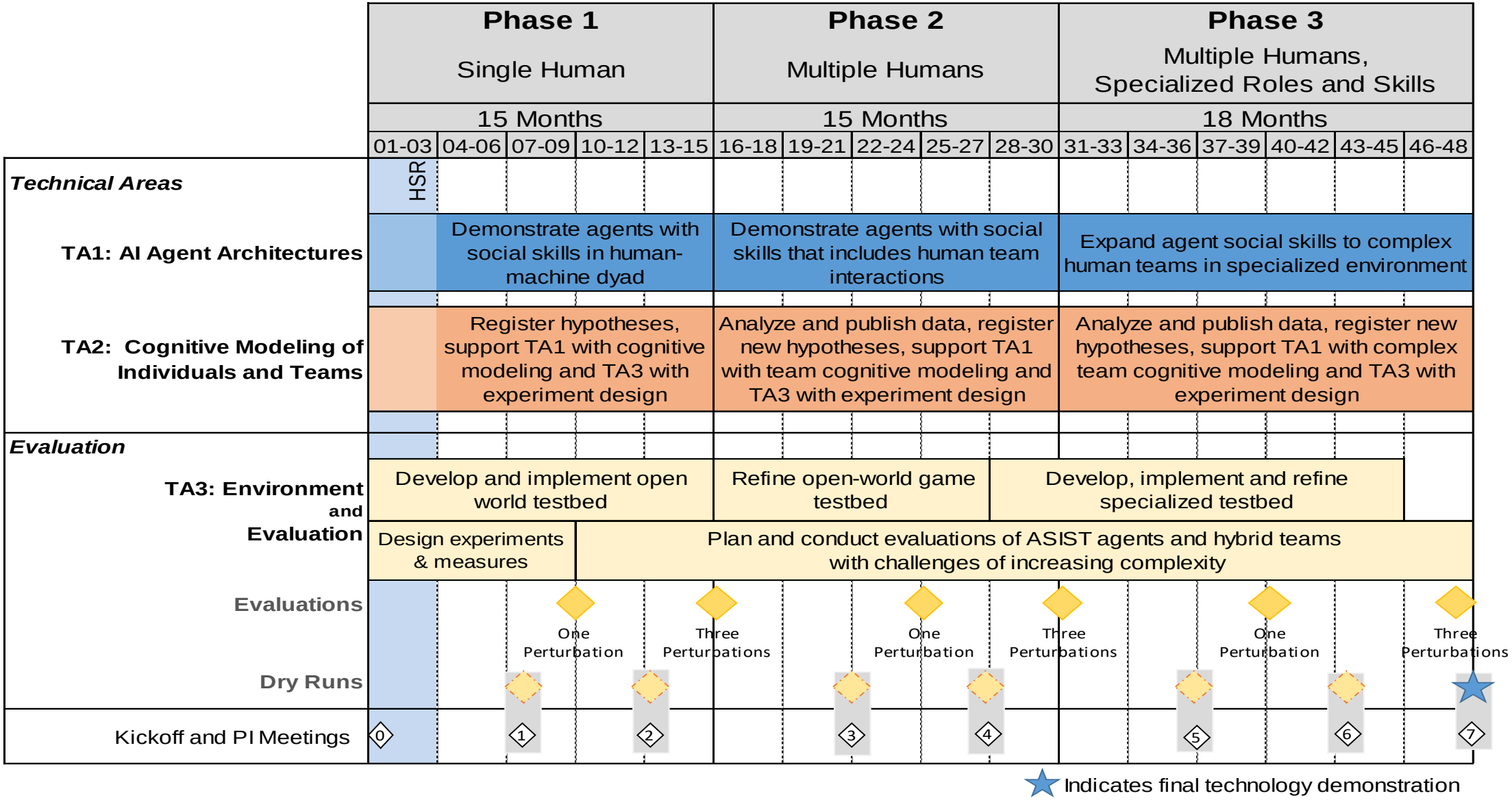
Types of perturbations (w/in phase):

- Changes in the goal or sub-goal
- Structural change to the environment or team
- Sudden change in strategy

TA3: Program meetings and experiments

- The program will start with a 3-day kickoff meeting hosted by DARPA
- TA3 will then organize, design, and execute the following:
  - Six two-day **evaluation events** (two per phase)
    - Hosted at the site of the testbed
    - Reps from TA1/2 teams included
  - Host **technical meetings** approximately three months before each evaluation
    - Technical meetings will include:
      - Three-day dry-run experiment
      - Two-day all-hands PI meeting
    - Location for technical meetings will be local to the TA3 performer
  - Organize and host a final three-day technology demonstration in conjunction to a 2 day final PI meeting at month 48 in the Washington D.C. area

ASIST Schedule



Some proposal details

- Each proposal may only address a single technical area, but proposers may submit multiple proposals
  - See Section III.D of the BAA for additional restrictions

- Performers selected for TA3 will not be selected to perform in either TA1 or TA2
  - This is intended to avoid organizational conflicts of interest (OCI) situations
  - Proposers may submit proposals for all three technical areas

- Human Subjects Research (HSR) is expected for TA3 proposals

- HSR may be included in TA1 and TA2 proposals

Deliverables

- Performers are required to provide, at a minimum, the deliverables described below:
  - Copies of any and all technical papers derived from work funded by ASIST
  - Annotated slide presentations within one week after program kickoff and each program milestone
  - **Quarterly technical reports** within 15 days after the end of each quarter; using a PowerPoint template that details:
    - Key technical accomplishments
    - List code/results delivered and the location
    - Plans for the next quarter
    - Issues (technical, programmatic, financial)
  - **Monthly financial status reports** within 15 days after the end of each calendar month
  - TA1 and TA3 performers must deliver software source code prior to each dry-run and evaluation
  - Final report summarizing the project at the end of the overall period of performance

www.darpa.mil