IDA

# Acquisition Challenges of Autonomous Systems

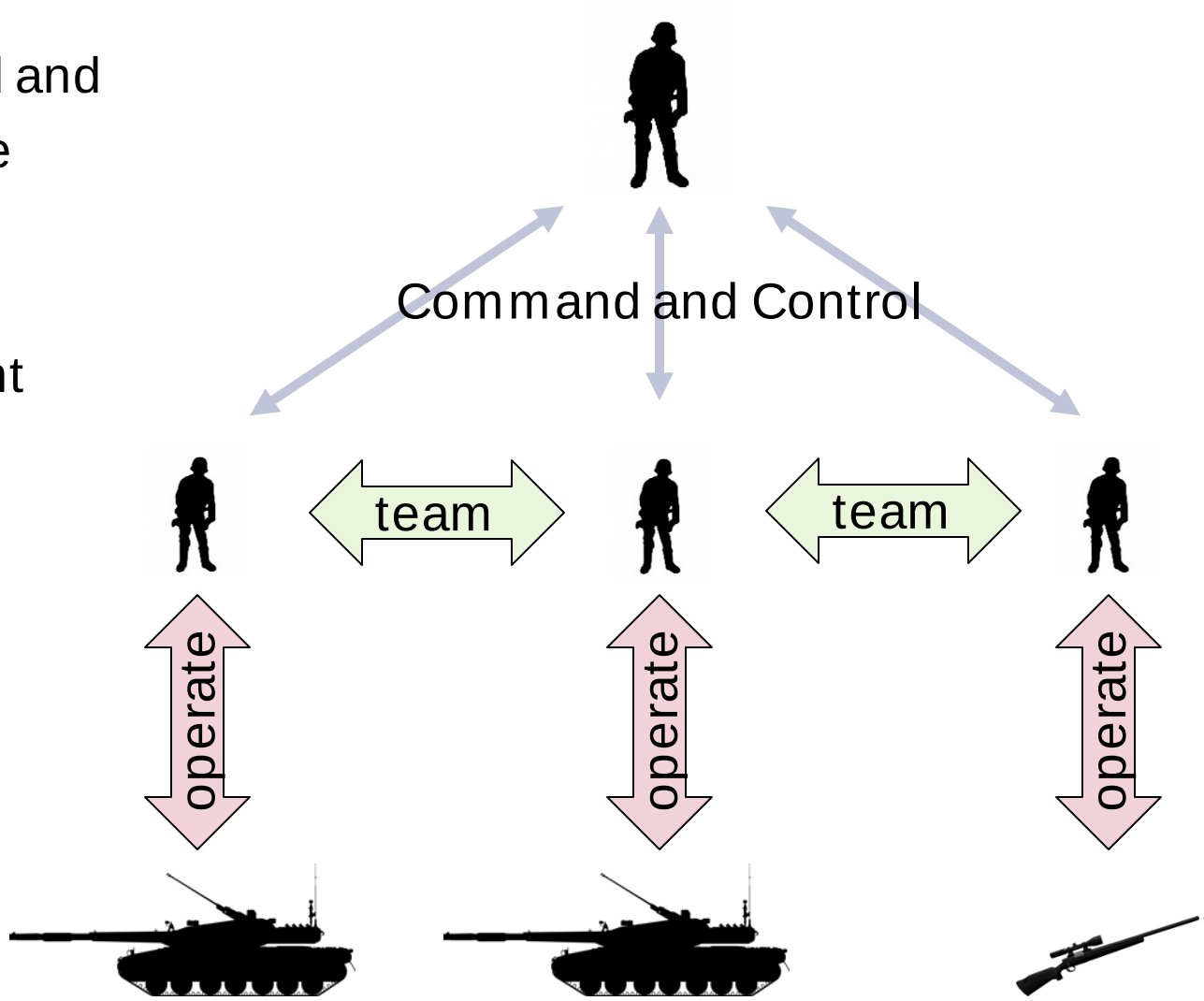
Or,

How Autonomy Breaks Acquisition

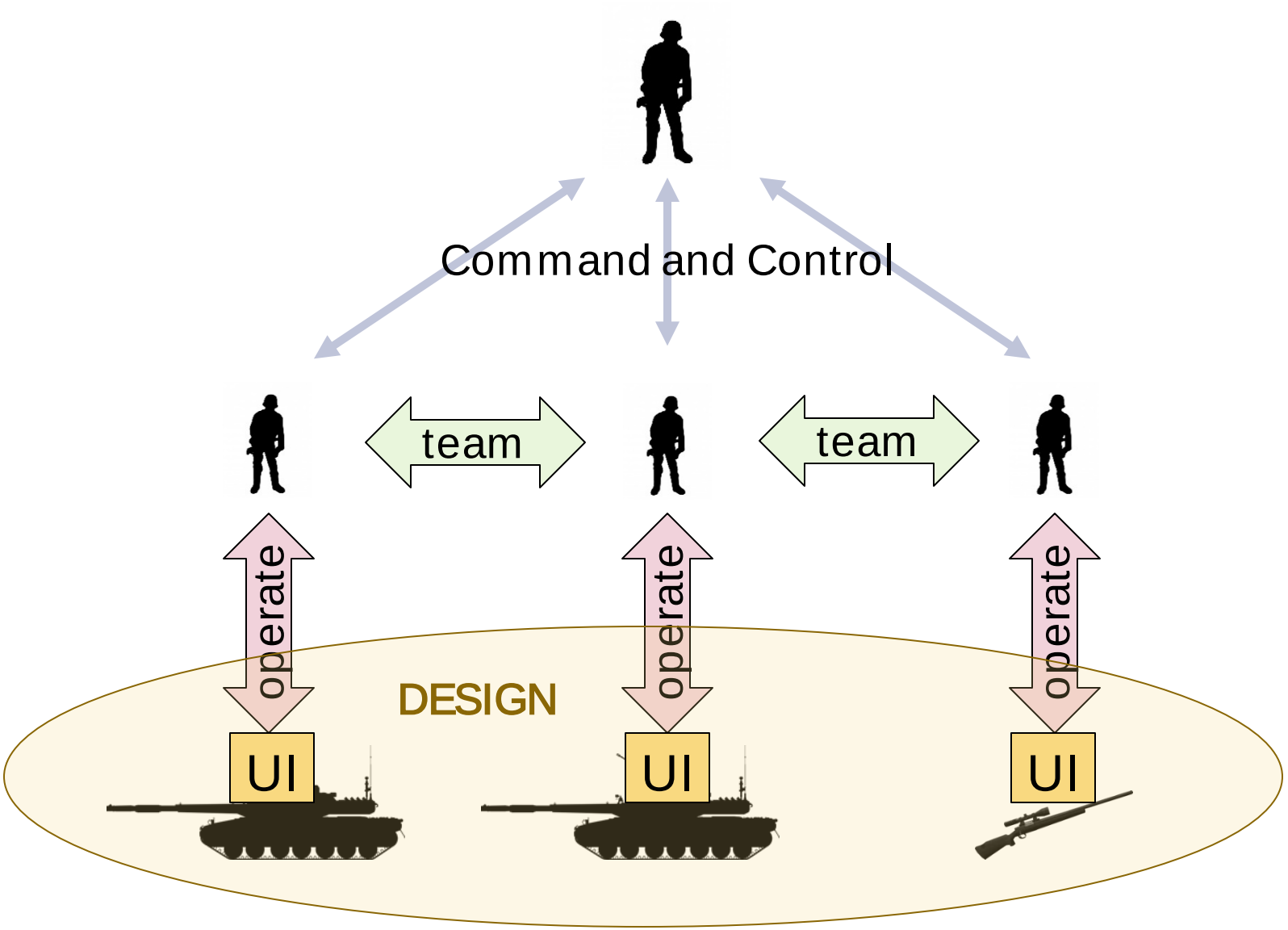
10 May 2018

# Traditionally, warfighters operate weapon systems

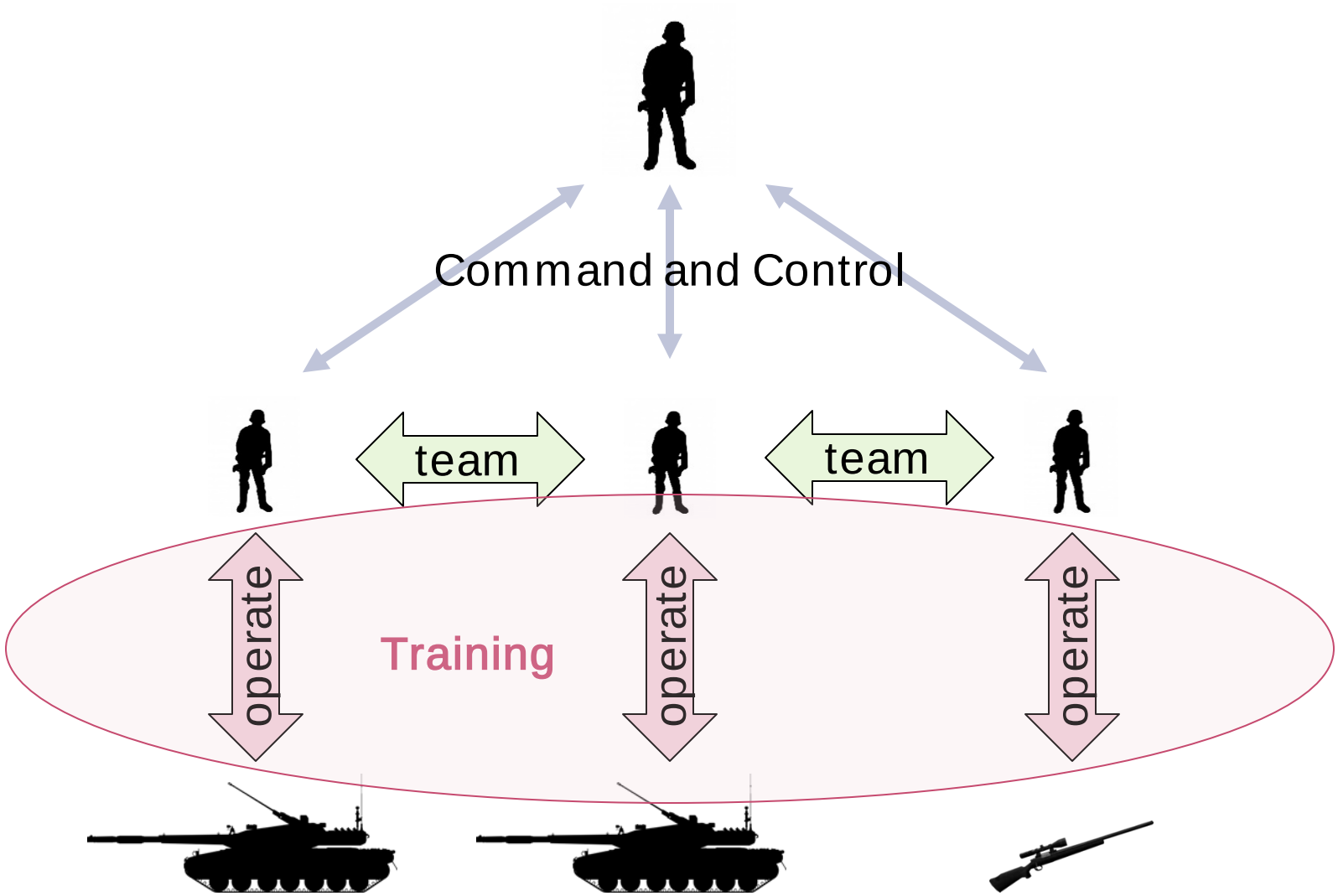
- Receive command and control from above
- Team with other warfighters
- Operate equipment



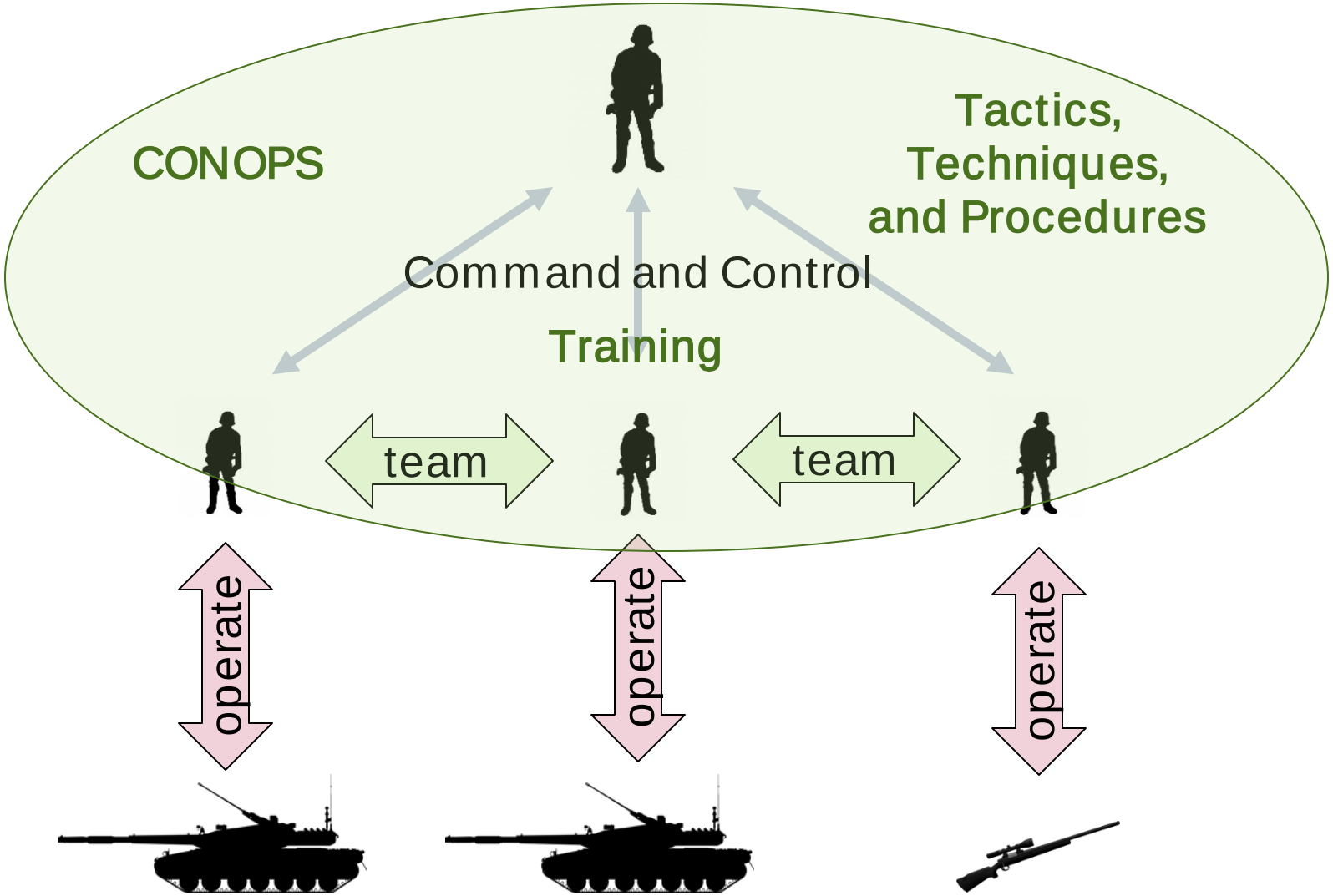
# “System design” only applies to equipment and user interfaces



# “How to operate the equipment” is part of training



# “How to team” is codified in CONOPS and TTPs, but implemented and tuned via training



# System design ends before training is complete (or perhaps even begun)

Designers make a guess at good user interfaces, based on planned CONOPS and existing TTPs

Some interface tweaking during system development

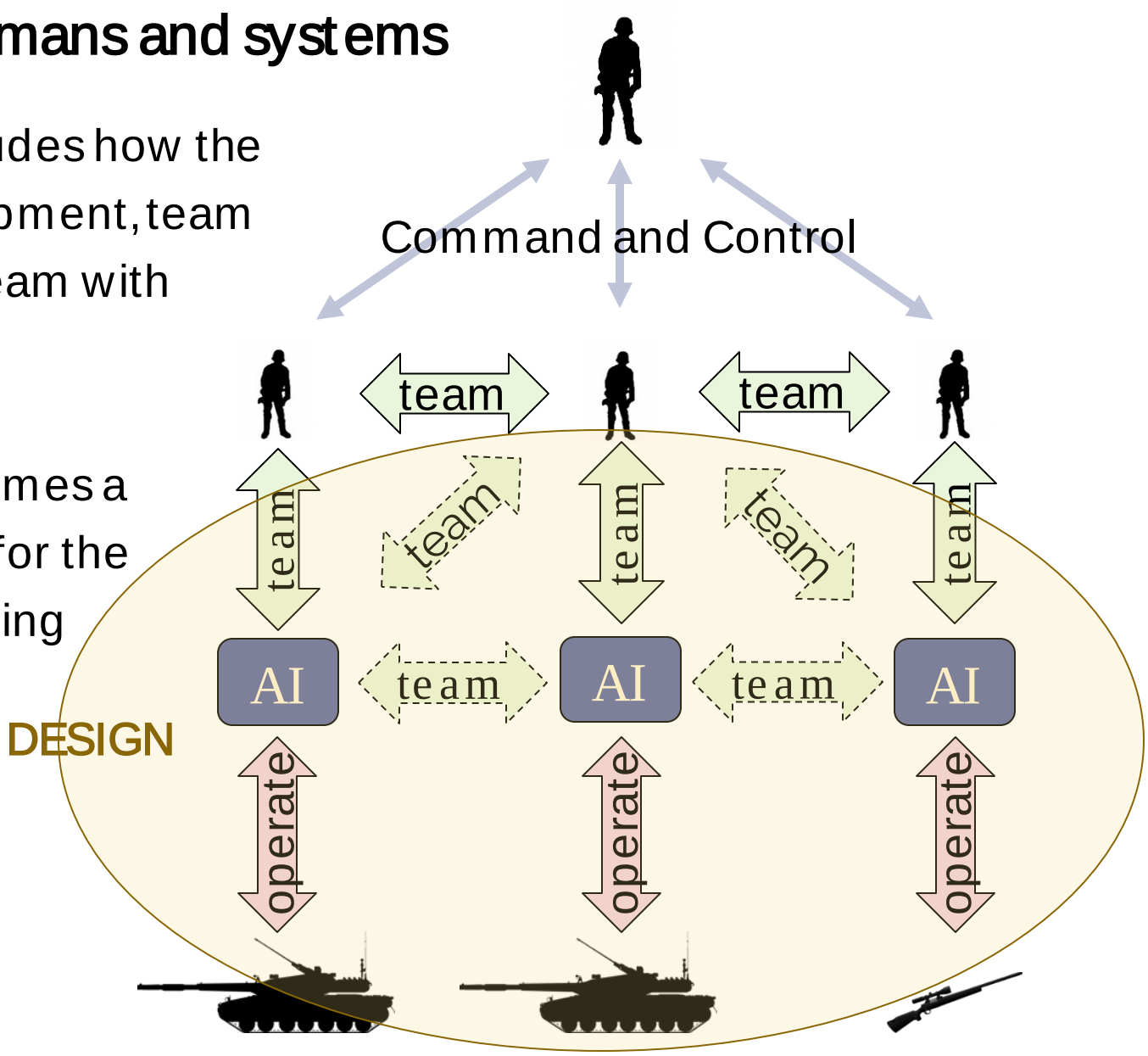
Warfighters figure out how to make best use of the system as it was designed and built

Little fear that the system as built will be incompatible with the intended CONOPS or existing TTPs

# Autonomous systems operate themselves (to some extent) and team with humans and systems

System design includes how the AI will operate equipment, team with humans, and team with other AI

System design assumes a CONOPS and TTPs for the human side of teaming relationships



# System design and training design are no longer separable

Designers make a guess at

- AI control algorithms (to operate systems)
- Optimal human-AI teaming protocols
- Optimal AI-AI teaming protocols
- What CONOPS and TTPs the humans will use

This guess might be wildly wrong – there is (as yet) no science or engineering body of knowledge to enable prediction of how the proposed autonomy design will actually perform

# Getting the design right will require experimentation

Arriving at an effective design will require **experimentation** involving **representative warfighters** and **prototype systems**

This experimentation might discover that the intended CONOPS and/or existing TTPs **will not work** for this system

Achieving success might involve rethinking not just the algorithms, but also training and doctrine

# Our acquisition system hates experimentation

All of our acquisition processes are based on an assumption that post-Milestone B programs know exactly what they are building, and how to build it

(This is why GAO hates immature technologies at MS B)

A key feature of experimentation is that you don't know in advance how long it will take, or what the outcome will be

# How does this break acquisition?

The schedule estimate will underestimate how much time will be spent experimenting with alternative designs

The cost estimate will underestimate R&D costs, and will be based on a design that is not the one that will be implemented

The test plan will be conceived in terms of “test events,” rather than as ongoing experimentation and discovery

Training and doctrine are not normally within the scope of what programs are allowed to tinker with

# Developmental Test and Evaluation (D&TE) will be particularly challenging

DT&E supports three distinct goals:

1. **Characterization** of system behavior
2. **Diagnosis** of undesired behavior
3. **Certification** of adequacy/ compliance

# Autonomous capabilities complicate all three of these goals

Design of Experiments isn't enough to characterize

- Response is not smooth between design points
- Too many factors, possibly not all known

Diagnosis is extremely difficult

- Is the problem in the sensors, the world model, the reasoning algorithm, the decision algorithm, the training data, or the teaming protocols?
- How can we tell?

Proving a negative is impossible

- Need to assure that the system will NOT ___.
- All testing becomes like cyber or safety assurance

# Autonomy demands new kinds of instrumentation

To support diagnosis and certification, we will need to instrument the internals of both autonomous decision-making and of human-AI interactions (including trust)

This poses not only technical challenges, but also potential intellectual property and data rights issues

# Summary (1)

Autonomous systems pose unique challenges to the acquisition system

These challenges arise because there is no mature scientific theory or engineering practice to predict how proposed systems will behave

Current cost and schedule estimation approaches are even less suitable for autonomous systems than for historical systems

# Summary (2)

Assuring system dependability will require new test and evaluation approaches that are incompatible with current standard practices:

- Novel instrumentation

- Ongoing experimentation (LVC)

- Ongoing intelligent adversarial testing


IDA