Assessing Human Performance in a Distributed Virtual Battle Experiment

J. Crebolder, C. Cooper-Chapman,
M. Hazen, C. Corbridge

12th International Command and Control Research & Technology Symposium – June 2007

Defence Research and Development Canada
Recherche et développement pour la défense Canada
Canada

Study Objectives

- Conduct a distributed encrypted simulation with multiple runs
- Validate findings from a force defence operational research study
- Evaluate different human factors assessment tools under conditions of distributed simulation with multiple runs

Virtual Battle Experiments

- System level, human in the Loop (HIL), experimentation using a virtual/synthetic environment to stimulate the system
- A next step in the CD&E spiral from constructive simulation modeling of concepts
- Particularly useful for systems/concepts where the human has a large effect on performance such as combat systems

Operational Research Study

- Defence against SWARM attack in littoral
- Netcentricity one approach to increase SA
- Addition of intelligence through UAV surveillance
- But key to usefulness of SA is its timely arrival

Distributed Maritime C2 Experiment
On the human side.....

- Do situation awareness, communication, shared awareness, workload - differ depending on the availability of surveillance information coming from a UAV?

- Situation awareness - decision making dependent on information processed
- Communication - situation awareness is acquired, in part, through sharing information
- Shared awareness - sharing information - building a shared mental model
- Workload – can affect the ability to acquire and maintain situation awareness

Distributed Scenario Runs

- UK-AS-CA-NZ-US contributed in various ways
- UK-AS-CA - produced naval teams and participated in experimental runs
- 2 nations paired on each scenario run
- Each nation paired equally with other 2 nations
- Scenario runs scheduled to minimize effects of time zone differences - challenging

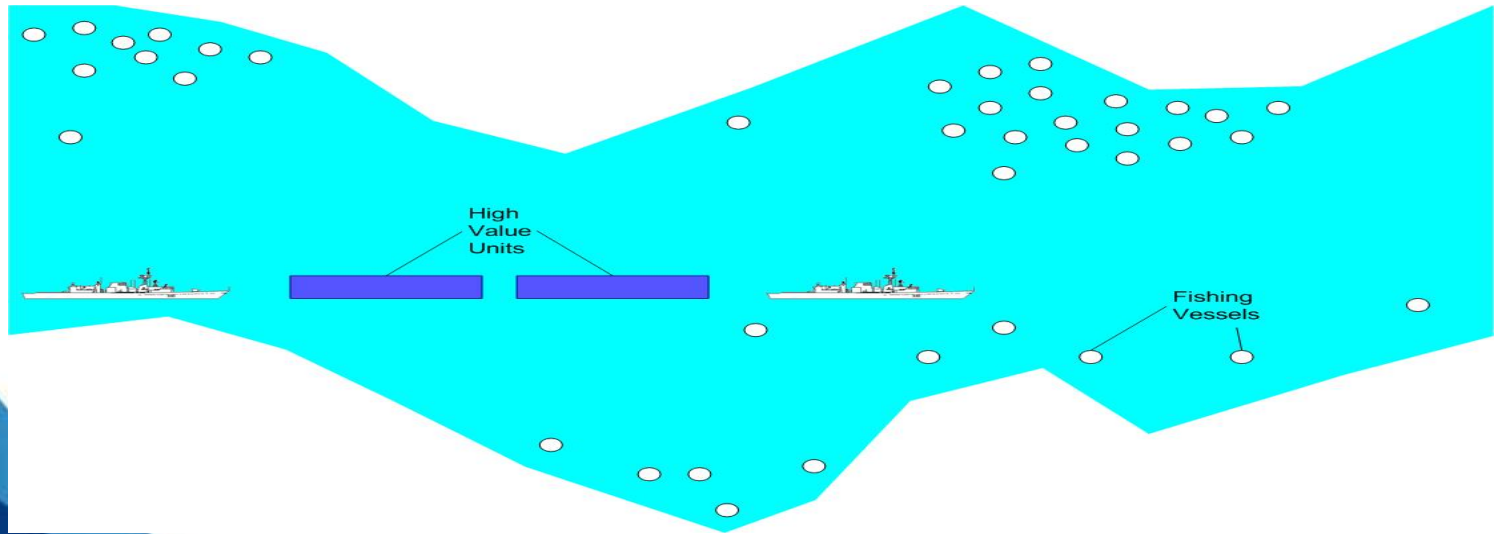
Scenario

Two allied warships tasked to escort two high value units (HVU) through a confined strait of water in a Force Defence operation

Intelligence reports indicate the threat of a terrorist attack

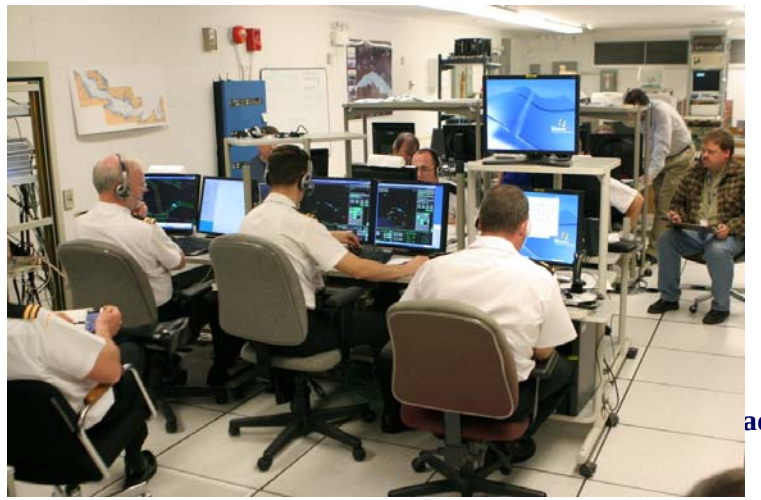
There are numerous fishing vessels and pleasure craft in the area

As the convoy transits the strait a swarm of fast inshore attack craft form up and attack one of the HVUs

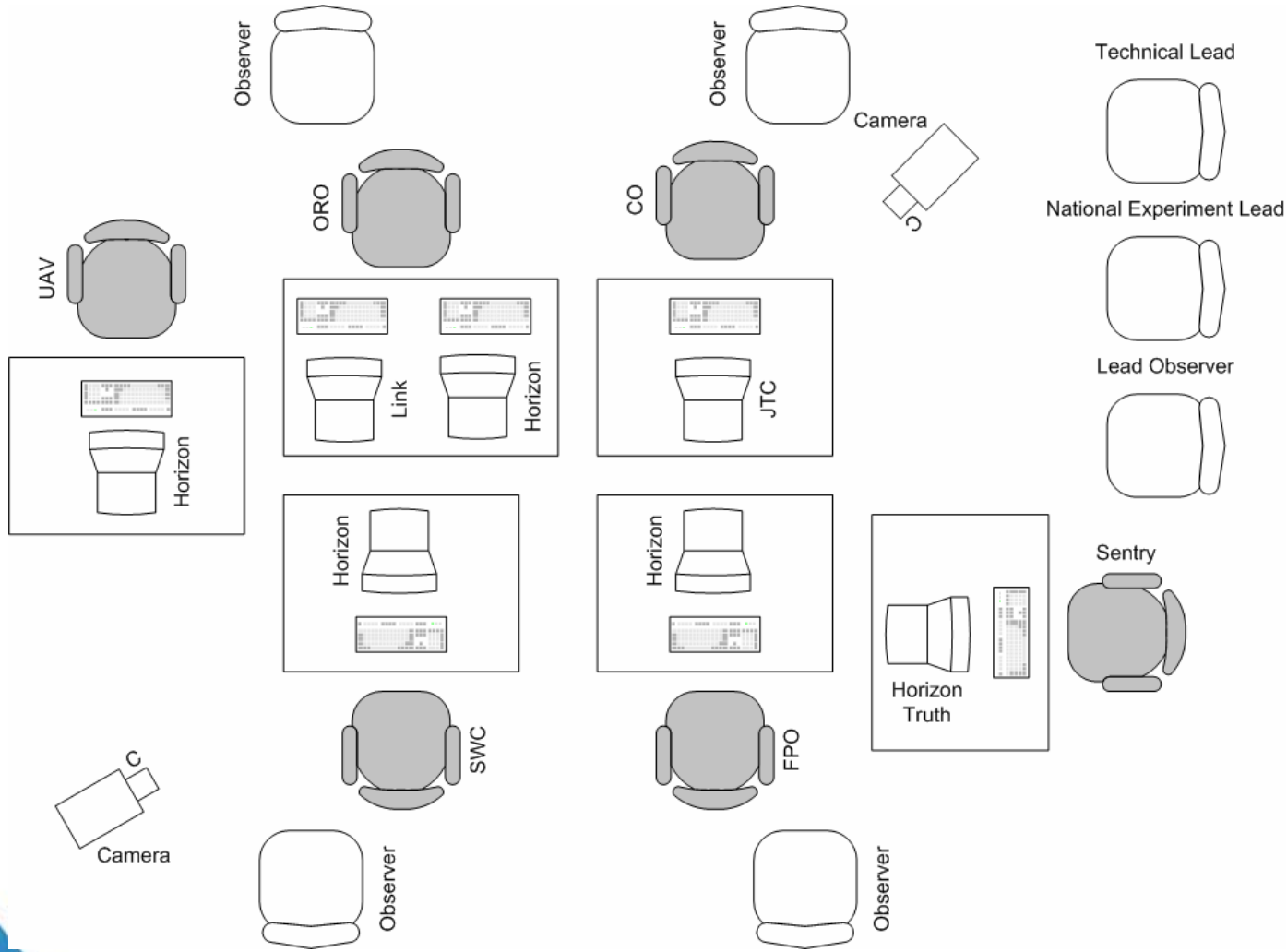


Personnel

- 2 command teams per scenario run – 4 members each:
- Commanding Officer (CO)
- Operations Room Officer (ORO)
- Surface Weapons Controller (SWC)
- Force Protection Officer (FPO)
- One observer assigned to observe each participant + military officer as a lead observer
- Sentry
- UAV operator



Typical Experiment Room Layout



Data Collection

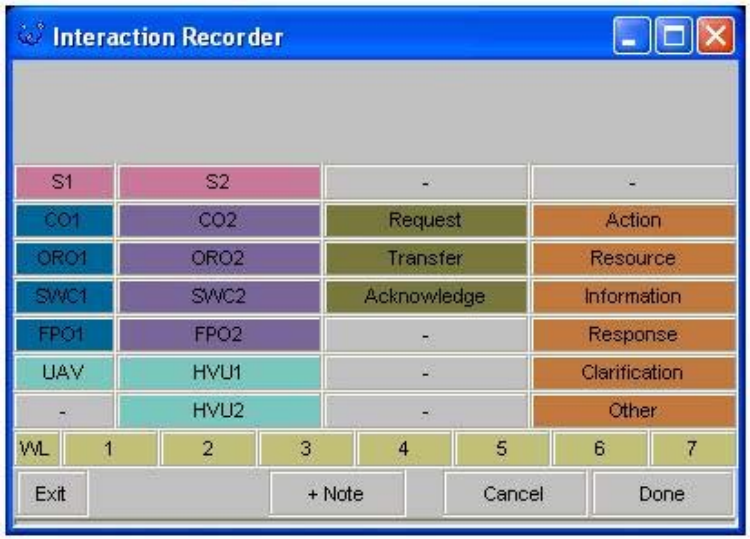
- Headphone/microphone sets for CO and ORO and their observers
- Chat for CO
- Interactive recorder software recorded interactions and workload - Personal Digital Assistant (PDA)
- Pre-scenario questionnaire
- Post-scenario questionnaire for participants and observers
- Audio/video recording of entire room
- System recording
- Observer notes - significant events/beh state

Independent Variables

- One CO was Officer in Tactical Command (OTC) on each run
  - counterbalanced across runs
- UAV was available on half the runs - controlled by OTC
  - counterbalanced across runs
- Performance compared with and without the UAV
  - UAV / Base
- and when OTC and not OTC
  - OTC / not OTC

Performance Measures - *Communications*

- Communications between team members recorded by observers throughout each scenario
- Frequency and Type of verbal communication
- Incoming and outgoing
- Includes voice, and voice over network
- Data collected by observers on PDA or laptop



e.g. CO1> ORO1 > Request > Information

Performance Measures – *Workload*

- PDA also used to collect subjective estimates of workload throughout each scenario
- Verbal response from participants - prompted every 5 minutes
- Scale 1 (can easily complete tasks) – 7 (cannot take on another task)

- Observers also took paper and pencil notes on significant events and behavioural state of participant

Performance Measures – Post-scenario *Workload*

- NASA Task Load Index (TLX) - electronic version
  - Individual - estimate individual workload
  - Team - individual's estimate of team workload
- 6 sub-scales:
  - effort, frustration, own performance, and mental, physical and temporal demand
- Completed after each scenario run

Performance Measures - Post-scenario
SA, Shared SA, Info exchange

- Questionnaire, on-line – subjective responses - completed after each scenario run
- *Situational awareness* - Crew Awareness Rating Scale
- *Shared awareness*
  - what other team members were doing
  - how hard were other team members working
- *Information exchange*
  - completeness and timeliness of information received; clarity of requests
  - adequacy of information available during mission
- *Opinion on scenarios and displays, tools*

Crew Awareness Rating Scale (2000)

Situation Awareness

- 4 levels
  - Perception
  - Comprehension
  - Projection
  - Intention
- Content – what information is available
- Process – ease of processing the information

Data Clean up

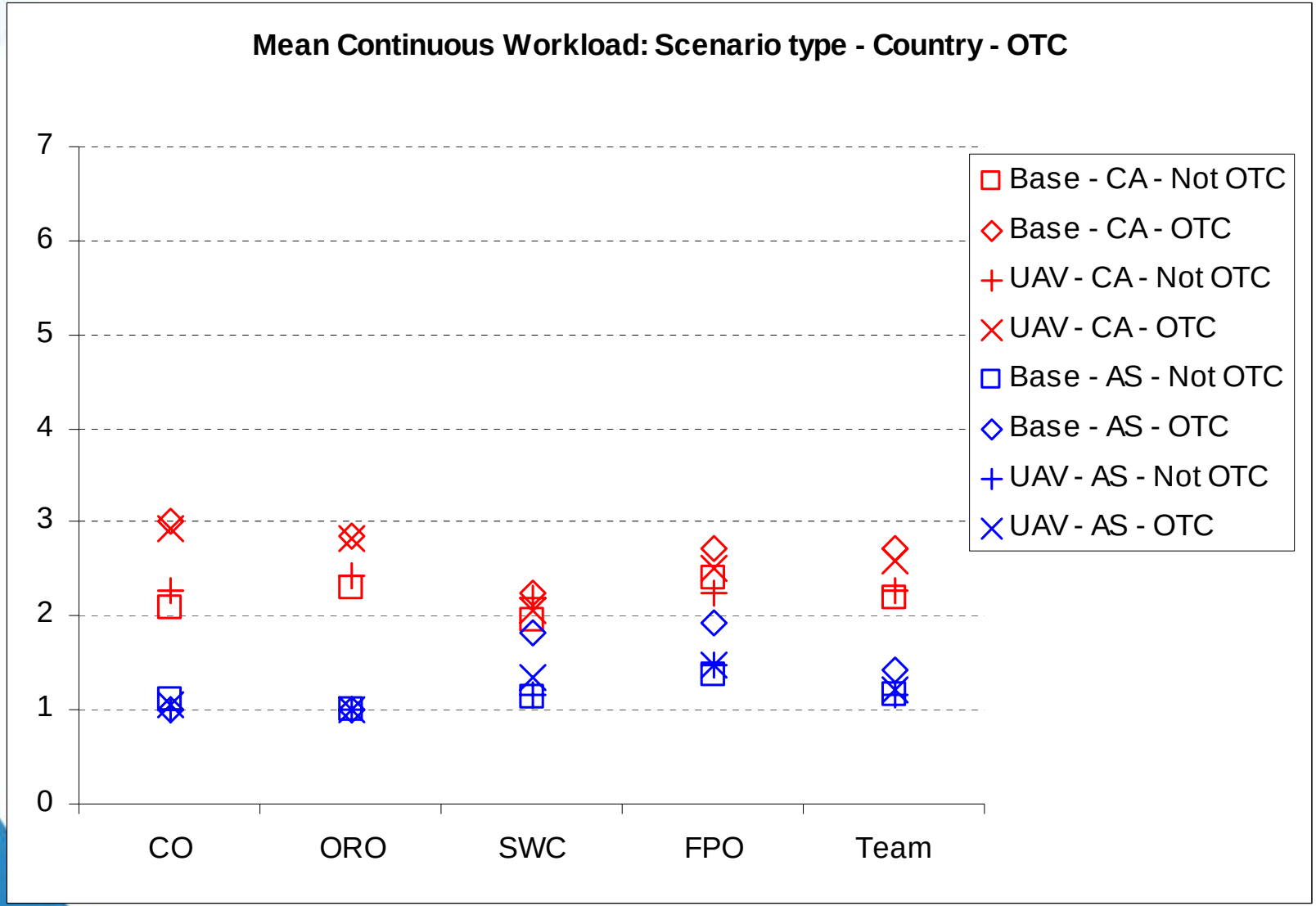
- 15 full runs conducted
  - 5 during week 1 (UK:CA/CA:AS)
  - 10 during week 2 (CA:AS)
- Each scenario was evaluated for data completeness and presence of technical issues
- Non-suitable runs were removed from data set
- 8 matching runs collected in week 2 (AS:CA)

  = each scenario run with UAV and without UAV (Base), and with CO as OTC and with CO as not-OTC

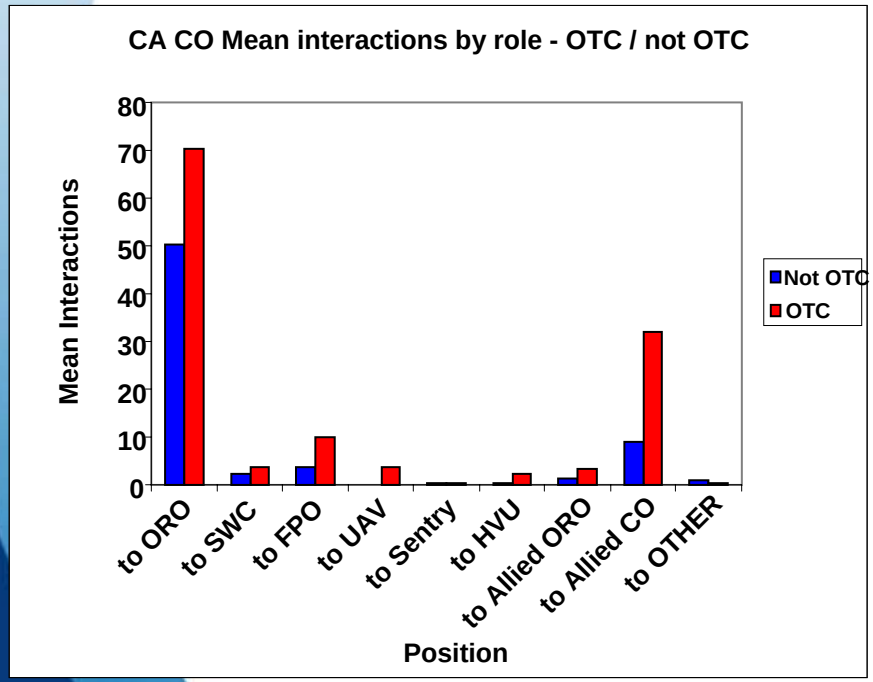
Data Analysis

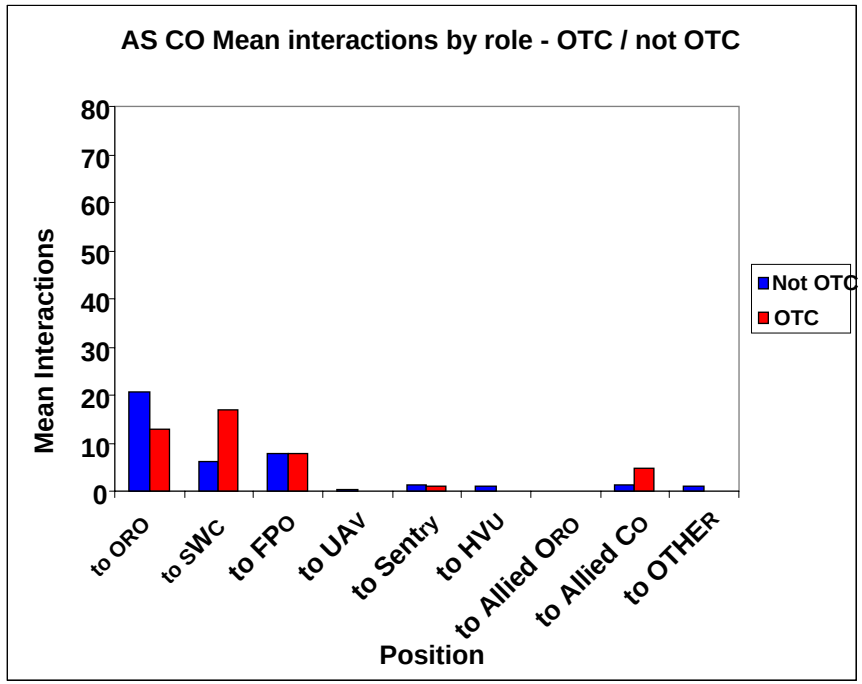
- Too few runs for statistical analysis
- Data was examined for trends in performance
- Focused on:
  - continuous workload
  - communications – frequency of out-going
  - post-scenario assessment of individual workload (NASA TLX)
  - situation awareness (CARS)

Continuous Workload - every 5 mins

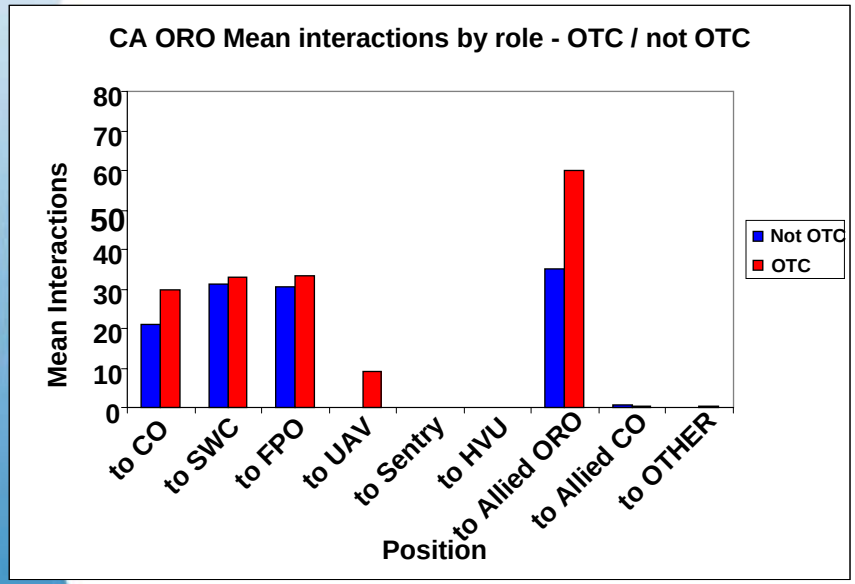


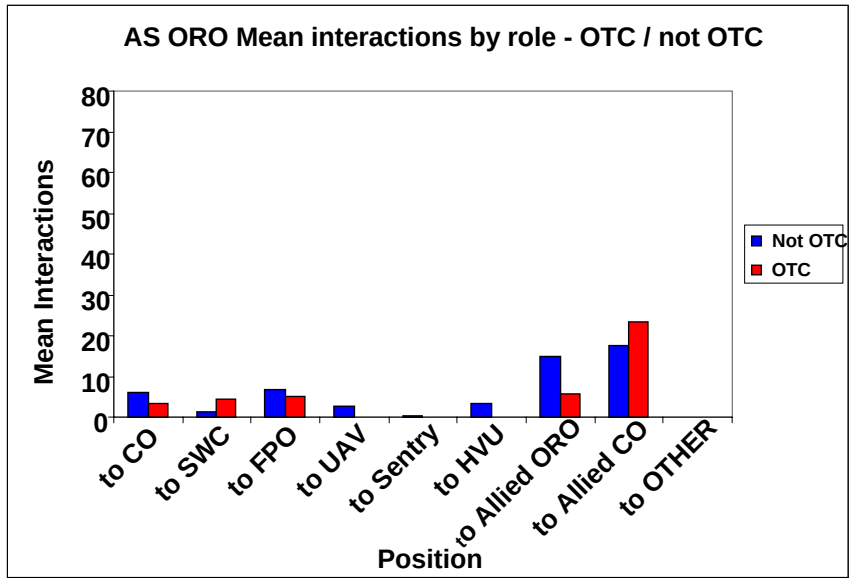
Frequency of Verbal Communications
Out-going for CO



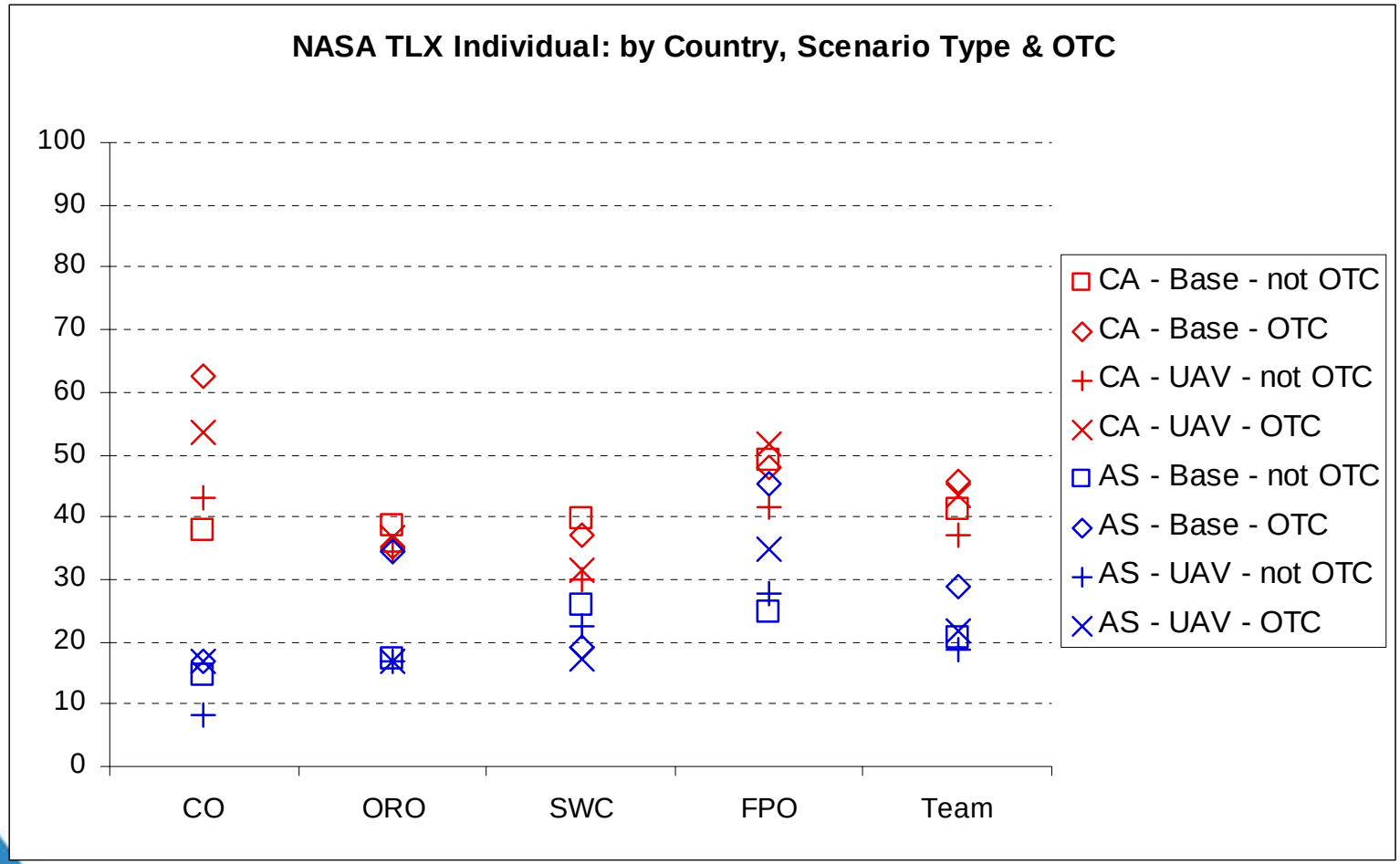


Frequency of Verbal Communications
Out-going for ORO

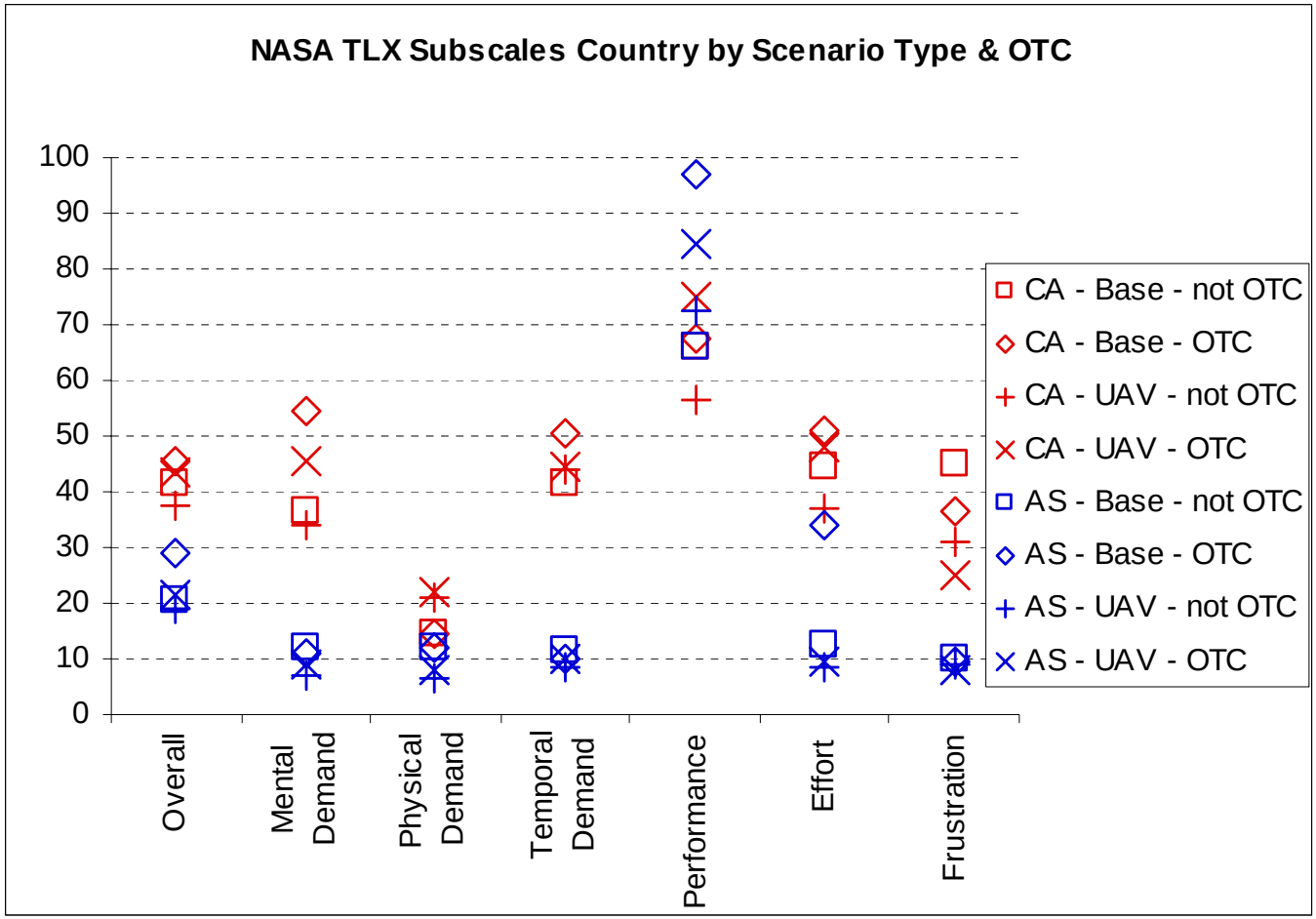




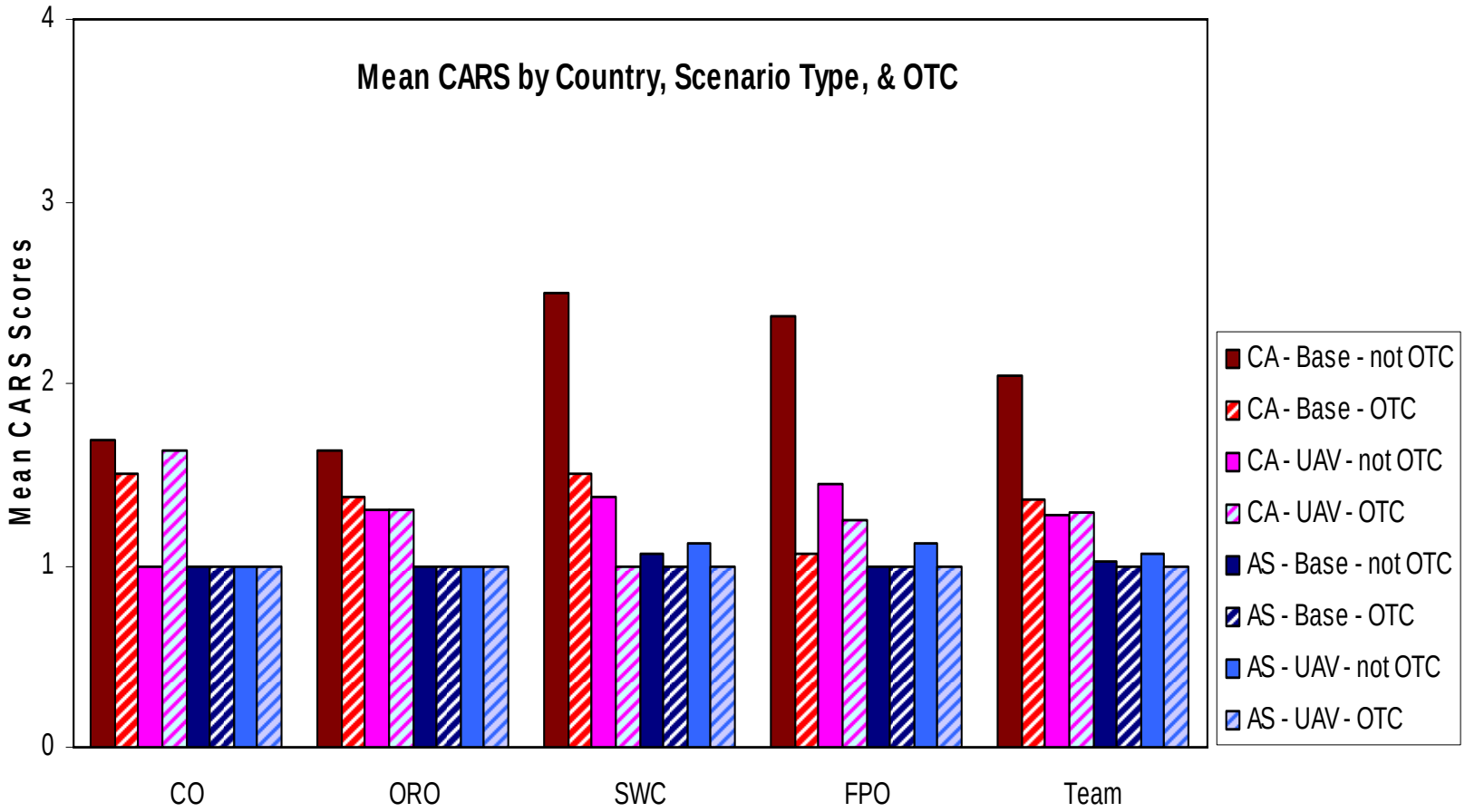
Post-scenario Workload assessment
NASA TLX



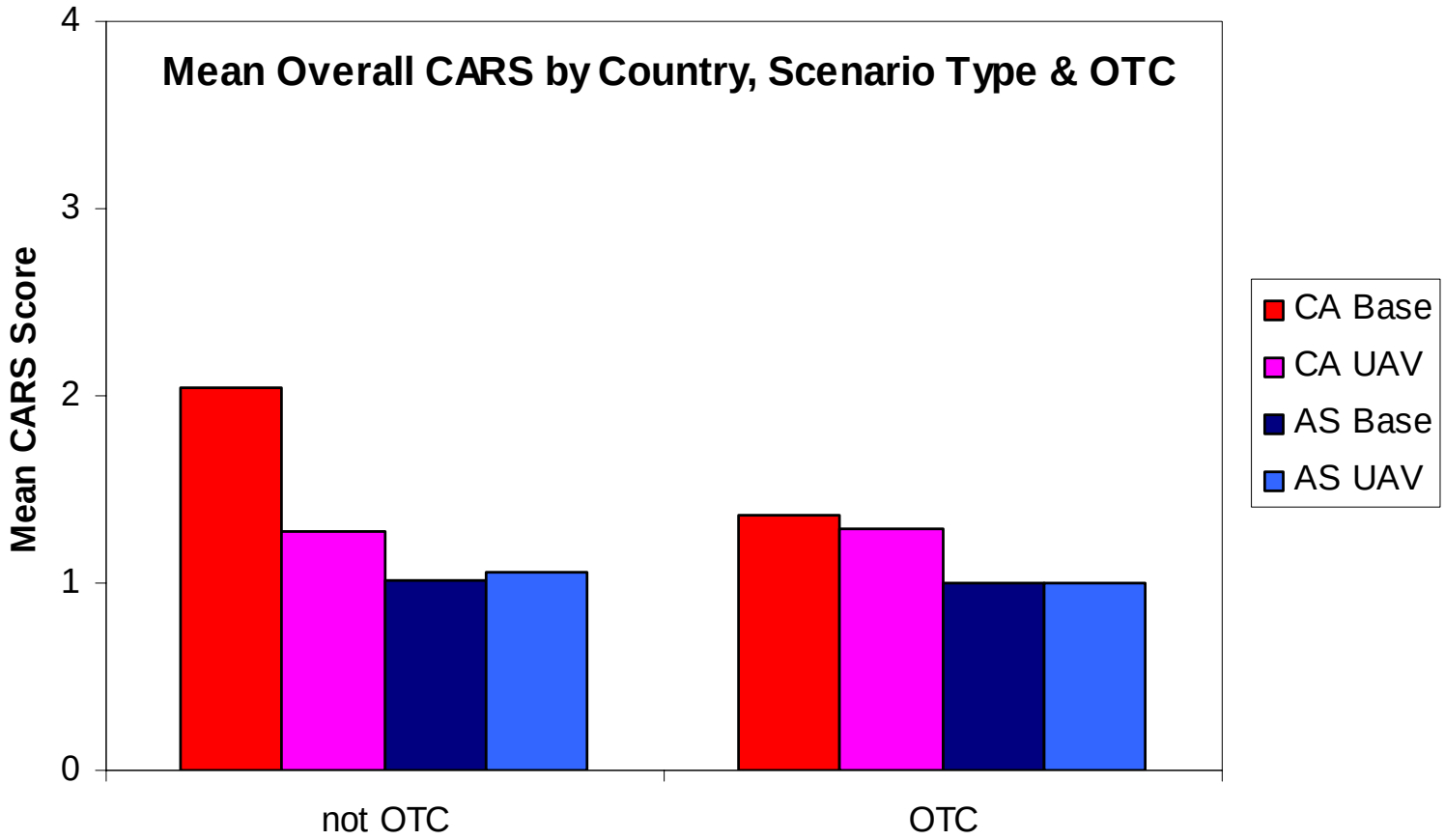
Post-scenario Workload assessment
NASA TLX sub-scales – by team



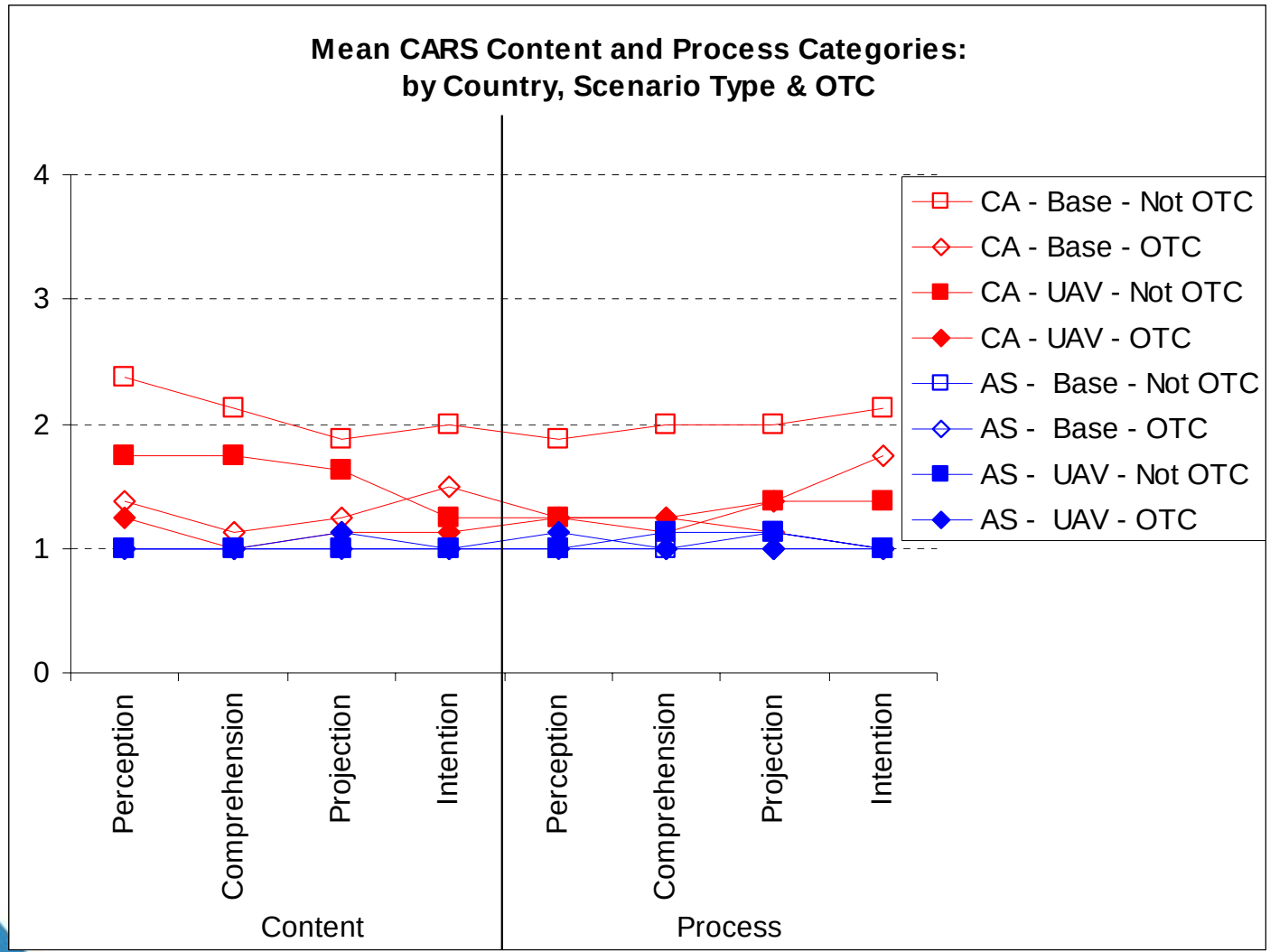
Post-scenario SA assessment
CARS – by position



Post-scenario SA assessment
CARS – by OTC



Post-scenario SA assessment
CARS – Content & Process



Other Data

- NASA TLX (2) - estimate of team workload
  - According to responses on questionnaire team members were unable to reflect on how busy their peers were
  - Useful in research where the focus is shared awareness amongst team members, questions would be specific to a specific aspect in the scenario
- Scenario realism
  - Positive comments - strait transits, number of contacts, difficulty in identifying hostile intent, swarm tactics, ROE
  - Negative aspects – lack of intelligence and ESM information, as well as technical problems, slow responsiveness of system
- Participant displays and observer tools good
- Shared awareness – responses on questionnaires not yet examined

Summary

- Small number of participants and runs led to evaluation of trends rather than statistical analysis
- Performance generally improved with the UAV for the OTC ship
- Positive comments on UAV
- OTC meant more work
- AS lower workload, better SA than CA, and less variability

Lessons Learned

- Most of the lessons learned are fairly obvious, being components of good experimental design
  - however, a multi-national study of this complexity and size brought many challenges, not the least of which were time and resource issues
  - one of the objectives of the study was to discover what sorts of human factors measures of performance and data collection tools would be most useful in this kind of work – this we did
  - and in addition - identified important aspects of complex distributed experimentation and possible ways to ameliorate the challenges in data collection

Lessons Learned cont'd

- Questionnaires should be as short as is feasible of course – we now have a good idea of what type of questions work and what areas to focus on with distributed teams and repeated runs
- Scaled responses are anchored - easiest and speediest for participants to report, easiest to analyze – important when large quantities of data are being collected
- One of the most informative and straightforward measures was the continuous workload
- The process of collecting data post-scenario has to be well structured, streamlined, and supervised – overseers need to be somewhat assertive
- Military officers acting as lead observers were very useful for insight – should also include in pilot study

Lessons Learned cont'd

- Our participants felt they had sufficient training but observation of video tapes suggests otherwise
- Observers need to be trained to criterion and tested for inter-rater reliability – difficult across nations
- Technicians need sufficient training - particularly for using the UAV
  - SA was a problem for the UAV operator in terms of understanding which visual contact they were looking at on their display
- After training and pilot study a dry run that includes data analysis should be conducted
- Sentry and EXCON should be remote

Lessons Learned cont'd

- As with all experimental work that includes human in the loop, future work must ensure that human factors personnel are engaged right from the start
- In multi-national studies representatives from each nation must work together to establish objectives and measures of performance
  - for the benefit of the experiment in general
  - for the benefit if each participating nation


DEFENCE
R&D
DÉFENSE