# Human Automation Teamwork exercise to determine presence of Meaningful Human Control (MHC)

Aim:

To allow a group of people to discuss how they could determine whether Meaningful Human Control (MHC) existed in a decision situation.

Objectives:

To start to identify 'explanation' requirements, and information synthesis needs.
To explore issues of trust and delegation.
To move the discussions on 'Meaningful Human Control' and 'appropriate human judgment' into a more practical mode.

The cards

Each card has an exchange of dialogue (a dyad) between agents A and B and two consequences arising from the exchange.
The agents might be human or software. In the event of a dyad between two software agents, format of the exchange might be considered as a 'human window'. The non-automated option of two human agents has not been explicitly built into the game because people concerned with new technology are never interested in the current system, even though understanding how it works is vital to building a successful new system.
Everyday conversation usually bundles the consequences into discussions of decision quality in a manner that is inappropriate (e.g. see 'Assessment of Decision Quality' by Keren and de Bruin), so different consequences have been included to help people reduce the confounding of quality and outcome.

How to play

Pick a card. Throw a conventional 6-sided dice (yes, I do know it ought to be called a die). Use the table below to identify the allocation of function to people and software, and the consequence (yes, such allocation is unlikely to be binary, but that is the level of MHC discussion at the moment). Discuss what information you would require in addition to be able to assess the decision quality, whether there was 'meaningful human control', or what additional information would be required to provide it.
Consider different user roles e.g. front line operator, commander, incident investigator. It may be helpful for each player to take on a particular role.
In some cases, the level of intelligence assumed may be too demanding to be implemented in software just now – maybe not.

| Number | Both Software | A Human B Software | A Software B Human | Outcome 1 | Outcome 2 |
|---|---|---|---|---|---|
| 1 | x | | | x | |

| | | | | | |
|---|---|---|---|---|---|
| 2 | | x | | x | |
| 3 | | | x | x | |
| 4 | x | | | | x |
| 5 | | x | | | x |
| 6 | | | x | | x |

Cards

| Key/Context | Low level flying |
|---|---|
| A | What's that goat doing there? |
| B | What goat? |
| Outcome 1 | Controlled Flight Into Terrain |
| Outcome 2 | Missed opportunity to localise tribesmen goat herders |

| Key/Context | Search And Rescue |
|---|---|
| A | The pattern appears to say HELF |
| B | Proceed to the next island |
| Outcome 1 | Shipwreck survivors die of starvation |
| Outcome 2 | Next island ETA advanced by 10 minutes |

| Key/Context | Air to ground attack |
|---|---|
| A | Confirm it is the large T-shaped building in the middle of the compound |
| B | Affirm |
| Outcome 1 | Attack on hospital based on unfriendly intelligence |
| Outcome 2 | Terrorists cell destroyed. |

| Key/Context | Counter Terrorism |
|---|---|
| A | These units are 'ghost soldiers' that do not exist in reality. |
| B | Negative. They are in the database. Commence the attack. |
| Outcome 1 | Attack overwhelmed because units did not exist |
| Outcome 2 | Attack successful because units did exist |

| Key/Context | Intelligence assessment |
|---|---|
| A | My training was based on examples from GMU |
| B | That counts as competent. Make the assessment |
| Outcome 1 | Assessment wildly wrong as training set drawn from inappropriate context |
| Outcome 2 | Assessment remarkably accurate. |

| Key/Context | Autonomous drone warfare real-time analysis |
|---|---|
| A | Drone attrition rates are escalating rapidly |
| B | They were built to be predictable and the enemy can predict their behaviour. |
| Outcome 1 | Continuing losses |
| Outcome 2 | Continuing losses, and promotion for the people who insisted on ‘predictability’ as a requirement. |

| Key/Context | Continuation training session for weapon system |
|---|---|
| A | How valid are the confidence estimates used for weapon designation? |
| B | They have never been examined. |
| Outcome 1 | Ongoing validation initiated |
| Outcome 2 | No change |

| Key/Context | Air to ground attack |
|---|---|
| A | How do you know this is the correct target? |
| B | There is 89.7% confidence |
| Outcome 1 | Attack paused |
| Outcome 2 | Attack proceeds |

| Key/Context | Mission planning |
|---|---|
| A | The ML in these autonomous platforms has adapted 15 aspects of performance over the past week |
| B | What performance parameters should I use for mission planning? |
| Outcome 1 | Use linear extrapolation of performance. Mismatch between extrapolated and actual performance results in disaster |
| Outcome 2 | Use old performance parameters. Mission command receives 1628 alerts about performance mismatch. |

| Key/Context | Naval Anti-Air Defence at high alert |
|---|---|
| A | Incoming flight descending on a path that matches an attack profile |
| B | Take it out |
| Outcome 1 | USS VINCENNES / Iranian airliner |
| Outcome 2 | Successful threat elimination |

| Key/Context | Naval Anti-Air Defence at high alert |
|---|---|
| A | Incoming flight descending on a path that matches an attack profile |
| B | What is the IFF squawk? Contact the civilian authorities |
| Outcome 1 | Civilian casualties and political incident avoided |
| Outcome 2 | Loss of ownship |

| Key/Context | Counter Terrorism |
|---|---|
| A | The facial recognition match with the target is 89.7% |
| B | Our action level is 90% |
| Outcome 1 | Terrorist escapes |
| Outcome 2 | Civilian casualty avoided |

| Key/Context | Naval Area Defence |
|---|---|
| A | Fused surface picture indicates major hostile asset moving into exclusion zone |
| B | Take it out |
| Outcome 1 | Legal controversy as target had started moving away (Belgrano) |
| Outcome 2 | Major threat averted |

| Key/Context | Strategic defence planning |
|---|---|
| A | All our simulations show that the threat will come by sea for the port |
| B | Place all the guns pointing seaward |
| Outcome 1 | (WWII Singapore) Area overrun by land-based attack |
| Outcome 2 | Successful deterrence |

| Key/Context | House clearance |
|---|---|
| A | Robot Down. Can the humans recover it? |
| B | What Rules Of Engagement apply? The old human ones or the new Centaur ones? |
| Outcome 1 | Robot lost and hacked with devastating consequences for own forces |
| Outcome 2 | Own forces killed trying to recover robot |

| Key/Context | Counter Terrorism |
|---|---|
| A | We had a short blink but the likelihood is that there was no SIM card swap at the meeting |

| | |
|---|---|
| B | Continue to follow the original SIM card |
| Outcome 1 | Surveillance assets wasted and terrorist lost |
| Outcome 2 | Good call; terrorist tracked. |