Studying Complex Human-System Behaviour: Human-in-the-loop Simulation Requirements

Mr. David Crone
Air Operations Division, DSTO
david.crone@dsto.defence.gov.au

Prof. Penelope Sanderson
ARC Key Centre for Human Factors, The University of Queensland
psanderson@itee.uq.edu.au

Dr. Neelam Naikar
Air Operations Division, DSTO
neelam.naikar@dsto.defence.gov.au

The Defence Science and Technology Organisation (DSTO) is required to provide advice to customers for the procurement of future military systems using the high fidelity human-in-the-loop simulation (HILS) facility housed in the Air Operations Simulation Centre (AOSC). A program of research is under way that compares two work analysis techniques (traditional task analysis and Cognitive Work Analysis) on the basis of whether the human-system performance measures that they suggest are sensitive to system modifications and so may be used for system evaluation. In this paper we show that representing aircrew's tactical environment as a series of concentric "rings" resulted in the development of HILS requirements that let us evaluate the measures derived from both work analysis approaches. Using the rings to frame the experiment and develop simulation requirements was beneficial for several reasons including participant involvement, validity of the system and operator behaviour observed, and completeness of the study.

INTRODUCTION

The problem of evaluating whether a change to a system will be effective may be approached from two distinct work analysis techniques. On one hand, traditional task analysis (TA: Kirwan and Ainsworth, 1992), as used widely in the Human Factors Engineering (HFE) community, comes from the view that the basic unit of analysis is "the task". On the other hand, Cognitive Work Analysis (CWA: Rasmussen et al, 1994 and Vicente, 1999) emphasises constraints placed on the system. There has been no work that compares the two approaches on their usefulness for determining measures that are sensitive to system change. A program of research is currently under way at DSTO that addresses this issue (Crone et al, 2003). One hypothesis of the research is that CWA-based measures will be more sensitive to a system change than task-based measures when tested in a human-in-the-loop simulation (HILS) environment. Given the philosophical differences between the two approaches, any study that aims to evaluate the approaches empirically in terms of the type of measures they produce is complex. One of the many issues is how to design human-system experiments that take into account both tasks and constraints.

In this paper we report on one aspect of the research – the development of HILS requirements for conducting experiments to determine whether the measures are sensitive to system change. We show that representing aircrew's tactical environment as a series of concentric "rings" resulted in the development of requirements that allowed us to evaluate measures derived from both work analysis approaches. After providing some background for the study, the experiments that were performed are briefly discussed. Next the application of the rings to the work is described and a number of examples of the use of the rings to derive requirements are offered. Finally, the use of the rings to derive the requirements is discussed.

BACKGROUND

The Defence Science and Technology Organisation (DSTO) provides advice to customers for the procurement of future military systems. The high fidelity HILS facility housed in the Air Operations Simulation Centre (AOSC) is increasingly being used to provide the context in which to evaluate these novel systems. One of the benefits of the HILS facility is the opportunity to explore realistic human and system behaviour. However, with the increase in realism there is a commensurate increase in the number of possible system and human performance measures that can be used for evaluation. The question arises: What measure(s) of performance *should* be used to evaluate the system of interest?

A Human Factors Engineering (HFE) evaluation of a system may incorporate several stages. Each stage is a separate point on a continuum that ranges from tightly controlled but less representative experiments on one end, to less controlled studies on the other end that are more applied and are more representative. HILS supports studies toward the applied end of the continuum, and operational testing is even further toward the applied end. For a system to be accepted into service it must be tested in an operational setting. By using HILS the risk that the system will not meet its requirements, when tested operationally, should be reduced. Therefore, any evaluation performed in the HILS must reflect the conditions that the system will be exposed to operationally.

Traditional methods of selecting measures for use within each of the HFE evaluation stages are influenced by several factors such as practitioner experience, resources available, and criteria including reliability, validity, and sensitivity. Irrespective of the stage of evaluation there is a common approach to testing system improvements. This approach involves reproducing the tasks and activities that the human-system will typically perform, selecting measures that relate to that task, choosing a success/fail criterion for each measure, and then testing to see whether or not the system can perform the tasks against the success/fail criterion. For new systems, where there is little knowledge of what task and activities may be performed, synthesized tasks are developed that are then used as the basis for testing the system. However, this is a risky strategy because the tasks synthesized and the measures used to measure performance may not be the ones that reveal an important system performance deficit. Hence, a system change may be judged to be effective when it is not.

Although CWA has been used in a number of operational systems contexts over the last 20 years (e.g. Burns, Bryant & Chalmers, 2000) it has only been used to determine human performance measures within a microworld environment (Yu, Lau, Vicente & Carter, 2002) that is not readily applicable to system evaluation.

Because of this problem we are developing a CWA-based process for selecting measures that can be used in the HILS facility to inform the evaluation of future complex systems (Crone et al, 2003). To test whether a CWA-based process offers any advantages we are comparing two work analysis techniques: traditional task analysis and Cognitive Work Analysis. The techniques are being compared to each other in terms of the measures that they suggest and in particular whether the measures are sensitive to system change.

EXPERIMENT OVERVIEW

We designed several experiments to compare measures derived with the different work analysis techniques. A candidate system upgrade for procurement was selected which is an electronic warfare (EW) system mounted as part of the Black Hawk helicopter systems fitted for airmobile operations. As part of an EW "upgrade" (an EW system detects and provides information to the aircrew on radar systems. EW systems may also perform programmed actions to counter the radar systems) the range for which a threat could be detected was improved. A number of analytical products from the two work analysis techniques were developed, based on the EW system, and were used to produce candidate measures that should be sensitive to the upgrade.

To test whether the measures were sensitive to the upgrade, qualified air crew flew a number of missions in the HILS facility in the AOSC either with or without the system upgrade. During the missions data relating to the measures that were suggested by the work analysis techniques were collected. Following the mission flight, the crews were debriefed and interviewed using a modified critical decision-making methodology (CDM: Klein et al, 1989). Data for the measures collected were statistically analysed. Initial results indicate that the CWA-based technique produced a number of metrics that were sensitive to the EW system and were not suggested by the task-analytic technique. Analysis is continuing to assess whether the TA-based measures were sensitive also.

THE BLACK HAWK TACTICAL ENVIRONMENT: THE RINGS

Australian Black Hawk helicopters and crew typically take part in airmobile operations which involve the transportation of troops and/or equipment into areas where there may be hostile troops, over a variety of terrains and in various weather conditions. An operation is made up of three distinct phases.

1. During the first phase, mission planning, aircrews try to develop a picture of the area of operations. This includes developing a route that takes into account resources available (e.g. fuel and time), known and probable threat locations, and options available to them if aircraft systems fail or if they are engaged by hostile forces.
2. The second phase, mission flight, occurs while the aircraft is conducting the mission. In other words, the aircraft is flown along the planned route with the aircrew engaged in many tasks; for example, re-planning as a result of changed orders, engaging hostile troops or dealing with aircraft system failures.
3. The final phase, mission debriefing, is aimed at ensuring that information the aircrew may have gathered during the fight phase is disseminated to all concerned.

During the mission flight phase there may be periods of relatively benign flight and also periods when the aircraft is engaged by hostile troops (by small arms fire and/or by missiles). Typically, the only indication that an aircraft is being engaged by small arms fire will come from observation by the aircrew. If an aircraft is engaged by a missile then the EW system will detect and relay the information to the aircrew. Depending on the terrain, weather conditions and other factors (e.g. the aircraft position in the formation), the crew will manoeuvre the aircraft to defeat the threat. Clearly, the actual route taken throughout the mission flight phase is unique if one considers the variation in speed, altitude, heading, pitch, yaw and roll available to the aircraft. In addition, although many aircrew tasks and behaviours can be predicted, some may be unpredictable – that is, produced as a consequence of unforeseeable events or unique combinations of events.

The question for the design of the experiments was the following: how can we ensure that the behaviour (known and unknown) of aircrews and aircraft systems in such a dynamic environment is captured? Through semi-structured interviews with Black Hawk and fixed-wing aircrew subject matter experts (SME) we discovered that aircrew's tactical view of the world could be represented as a series of survivability rings or zones, in which certain system and human behavior could be located. (Note: The rings were originally developed by Defence Research and Development Canada (DRDC). However, this may be their first use for categorising human-system behaviour in order to define experimental requirements.) The interviews revealed that although Black Hawk aircrews were not explicitly trained to view the tactical world in this way, the framework was intuitive to them.

Figure 1 presents the rings – the rings are the spaces between the lines whereas the lines are changes in tactical status. The figure shows that an increasing level of threat results in a progression toward the central rings and reduces aircrew survivability. For example, during the mission flight phase the aircrew may use the available terrain and low level flight to hide their approach. If their EW system does not detect a threat then they are in the "Don't be seen" ring. If they are detected by a threat system, (change of tactical status to SEEN), and are aware of that, their behaviour may be considered to be characteristic of the "If seen don't be engaged" ring. Typical behaviours include deviation from their route and the EW system may perform some action to "break-lock" (this may be achieved by dispensing chaff packets that block the threat's radar waves). If that is not successful and the threat's missile system engages them (change of tactical status to ENGAGED), their aim is not to be hit – they are now in the "If engaged don't be hit" ring. They may perform sudden aircraft manoeuvres and the EW system may continue to perform actions to defeat the in-flight missile. If the above tactic fail to deter the missile and the aircraft is hit (change of tactical status to HIT), then they are now in the "If hit don't be killed" ring - landing safely or returning to base may the only options available to them.

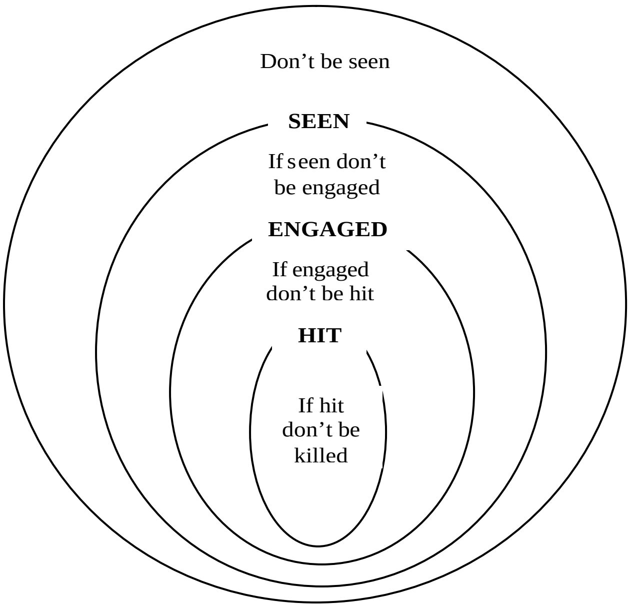


Figure 1 Survivability rings (adapted from DRDC)

Once the idea that the rings could be used to inform human and system behaviour was verified in a pilot study, the framework was used to help derive the main experiment requirements. During the main experiment, further evidence for the usefulness of the rings emerged as part of the post flight mission debriefs. Experienced Black Hawk EW operators would mentally map the current EW system behaviour onto separate rings during the mission flight phase. For example, if the EW system indicated that the threat system had changed mode from search to track then the crew knew that they had now "entered" the "If engaged don't be hit" ring and took action based on that assumption. The EW system behaviour seemed to "key" or cue them to appropriate behaviours. It seems that the EW system was being used as a tool to help the aircrew select the most appropriate behaviour to deal with the threat.

The relationship between the EW system and aircrew behaviour is illustrated in Table 1 and Figure 2 (taken from one mission flight). Figure 2 shows the actual route taken by the aircrew. The location of the threats is indicated by letters and significant data points are indicated by numbers. Figure 2 shows that the aircrew progressed along the planned route (dashed line) until data point 4 when they deviated from the route. The figure also shows data point 10 – where the aircrew aborted the mission in response to a threat system. Table 1 shows each of the ten data points and maps threat behaviour and EW system behaviour onto the appropriate rings with the resultant crew behaviours.

EXPERIMENT REQUIREMENTS

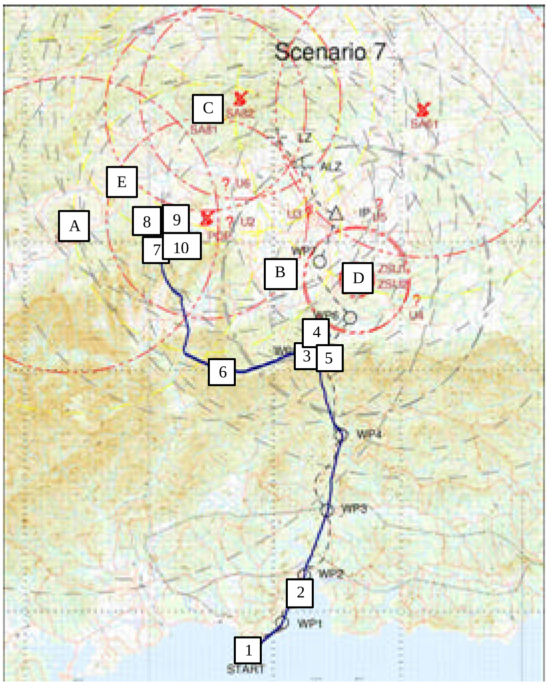
General Requirements

The experiments shared several high level requirements:

1. Complex human-system behaviour should be supported,
2. All airmobile operation stages should be included (mission briefing, flight and debriefing),
3. Aircrews should plan and fly airmobile missions in a tactically realistic environment using the Black Hawk simulator in the AOSC,
4. Aircrew should follow standard operating procedures (SOPs), including flight checks and threat avoidance,
5. Novel behaviour (but defensible by the aircrew) should be supported by the simulation environment,
6. Aircrew should perform their missions under normal operational mission constraints,
7. Several human and system performance measures should be recorded, as suggested by the work analysis techniques.
8. The experiments should support both known human-system behaviour (so supporting the task-analytic technique) and any new human-system behaviour (so supporting the CWA technique), i.e. the experiments should be "fair" to both work analysis techniques.

Application of the Rings to Generate Requirements

The observation that the EW system and aircrew behaviour could be placed in a rings framework had major implications for the design of the study—in particular design for the design of the trial missions, identification of simulation (including infrastructure) requirements, data capture and analysis, and the study protocol. For this paper only a small number of requirements will be listed, grouped under the headings of mission requirements, experimental design requirements and simulation requirements. In general, the rings informed events to be included in the mission; the events, in turn informed the type of requirement.



Data point number	Threat system identifier, threat location (clock code) and mode	Ring	EW system behaviour	Crew behaviour
0	Threats not in line of sight.	Don't be seen	No threat detected	Commence mission as planned
1	Threat A, 9 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them far away - no deviation from track necessary
2	Threat A, 11 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them far away - no deviation from track necessary
3	Threat B, 1 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them far away - no deviation from track necessary
4	Threat C, 11 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them near to track – deviation from planned route necessary.
5	Threat D, 1 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them near to track – deviation from planned route necessary
6	Threat E, 3 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – intelligence brief puts them near to track – deviation from planned route necessary
7	Threat E, 1 o'clock, Search	If seen don't be engaged	Search detected	Keep watch on threat – Situation awareness puts them far away - no deviation from re planned track necessary.
8	Threat E 11 o'clock Tracking	If engaged don't be hit	Tracking detecting	Treat high priority – use terrain making to defeat tracking – deviation from track necessary
9	Threat A, 11 o'clock Tracking	If engaged don't be hit	Tracking detecting	Treat high priority – use terrain making to defeat tracking - deviation from track necessary
10	Threat A, 6 o'clock Lock before launch	If engaged don't be hit	Launch detected	Very high priority threat – use terrain making to defeat lock – take evasive action - deviation from track necessary. Threat concentration too high – abort mission.

Figure 2 Aircraft route taken during mission scenario 7 and Table 1 Threat, EW system and aircrew behaviour mapped onto survivability rings

The relationship between requirements is complex; a requirement suggested by the rings may spawn a number of additional requirements or alter the effectiveness of other requirements. For example, one requirement, as indicated by the rings, was that a missile launch should occur during all missions – stimulating the "If engaged don't be hit" ring. However, during the pilot study it was noticed that if the missile was launched at the end of the mission (a missile launch could occur at the beginning, middle or end of a mission) the aircrew's ability to manoeuvre the aircraft was poor. Analysis of the aircraft flight data revealed that the poor aircrew performance was attributable to the amount of fuel remaining in the aircraft. Initially the mechanism used to limit the mission flight phase of the experimental run to 30 minutes was the amount of fuel "loaded" onto the aircraft (if the aircrew spent too much time loitering they would run out of fuel and they would be forced to land prematurely). However, it was clear, that toward the end of the mission the mass of the aircraft had changed so much that the flight characteristics had changed in a way that was unrepresentative of the real aircraft. The solution to this problem was not to start with a near empty tank, but to provide a full tank of fuel, and instruct the aircrew not to go lower than a predefined lower limit – something that they are familiar with. In this way the flight characteristics of the aircraft were predictable throughout the length of the mission flight phase and the aircrew behavioral data gained from a threat engagement at the end of the mission could be treated as valid.

Mission requirements. One of the mission requirements was to ensure that known and unknown aircrew-system behaviour was supported. This, together with the requirement that the missions should support behaviour from the "If seen don't be engaged" rings, resulted in a mission requirement that aircraft should be able to fly at crew-defined altitudes over various types of terrain. By applying this requirement a whole range of behaviours were observed. For example, over flat terrain the typical response to being seen by a threat system would be to turn away from it and reduce altitude, over undulating terrain the response may be to continue on the predetermined route using the terrain to hide from it and again reduce altitude. Both of these behaviours are known (they are SOPs) and were identified in the TA artifacts. However, by providing very mountainous terrain new behaviour (i.e. not identified in the TA artifacts) was observed. For example, during a mission the crews were observed significantly gaining altitude so that they could gain positional information on the threat – they knew that if they were seen (their intention) they could still hide from the threat if it was close enough to engage them. This was a significant departure from standard behaviour, but was a tactically valid solution to the problem of fixing the radar system's location. (Because the EW system provided bearing and not range information to the aircrew they used several bearings to the radar system to fix its location.)

Experimental design requirements. It was also important for statistical testing that trial mission runs were balanced in terms of the tactical conditions the crews were exposed to because we were investigating whether the measures were sensitive over the range of possible human-system behaviour. By considering each mission in terms of the rings (i.e. opportunities for events) rather than specific behaviours the missions could be evaluated for equivalence. For example, Figure 2 indicates that even though the crew diverted from the planned route, the event requirement for that mission (that the crew should be engaged by a threat system) was still achieved. In this way the data collected from a scenario from the "upgrade" condition could be compared to the data from other scenarios in the "not upgraded" condition. This is very different to other experiments using the HILS where crews are required to repeat the same mission profile and behaviour for the purpose of statistical protocols.

Simulation requirements. One of the more interesting visual and auditory simulation requirements concerned the usefulness of showing missile smoke and the explosion from detonating missile. This was particularly interesting because in previous simulation trials using the AOSC facility the requirement for showing these events was considered a very low priority – the cost to benefit ratio was too high. However, in the context of the rings the benefit may be clearly articulated - if the aircrew cannot see missile smoke and the explosion from a detonating missile that had missed the aircraft, they may be unaware that they are now moving from the "If engaged don't be hit" ring back out to the "If seen don't be engaged" ring and their behaviour may be affected accordingly. Hence, missile smoke and the explosion are important environmental cues that should be included to ensure the representativeness of the experiment and data validity. One again the rings informed the type of event, which informed the simulation requirement.

DISCUSSION

The aim of this paper was to outline the benefit of using a representation of aircrews' tactical environment to derive HILS requirements. By developing a complete set of simulation requirements the goal of providing an environment in which to evaluate whether work analysis techniques could be used to produce measures sensitive to a system change seems to have been achieved. From an initial analysis of the results it seems that a number of CWA- based measures are sensitive to EW system changes. Further analysis is under way to examine whether the TA –based measures are also sensitive to the system change. The use of the rings to frame the experiment was beneficial in a number of areas including operator (participant) involvement, validity of the system and operator behaviour observed, and completeness of the study. For example, the aircrew commented that the level of simulation fidelity and types of missions flown were similar (though harder) to normal operations. They found that all their normal and unique behaviours could be supported by the simulation environment and this was evidenced by them following SOPs and developing behaviour for novel (but tactically sound) situations. This is different to most simulations in which only a subset of known crew behaviour is studied. The rings also ensured that there was a mechanism to enable missions containing many different variables (e.g. distance to the threat when it is first observed, number of threats in the missions, route taken by the aircrew, etc.) to be evaluated in terms of equivalence.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support given by the following, Dr Simon Parker for sponsoring this work, the members of the AOSC for helping to define and implementing the requirements for this study and to the members of the Australian Army Black Hawk aircrew community who readily gave up their time to participate in the experiments.

REFERENCES

Burns, C. M., Bryant, D., & Chalmers, B (2000). A work domain model to support naval command and control. *Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics*, 2228-2233.

Crone, D. J., Sanderson, P. M. and Naikar, N. (2003) Using Cognitive Work Analysis to Develop a Capability for the Evaluation of Future Systems. *Proceedings of the 47th Annual Meeting Human Factors and Ergonomics Society*. Denver, CO.

Kirwan, B and Ainsworth, L.K (1992) *A Guide to Task Analysis*. Taylor and Francis. London.

Klein, G. A., Calderwood, R. and MacGregor, D. (1989) Critical Decision Method for Eliciting Knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, Vol 19, 3*, May/ June.

Miller, C. A. and Vicente, K. J (1999) Task "Versus" Work Domain Analysis Techniques: A Comparative Analysis. *Proceedings of the 43rd Annual Meeting Human Factors and Ergonomics Society*.

Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P. (1994) *Cognitive Systems Engineering. John Wiley and Sons, Inc. New York*

Vicente, K. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. LEA. Mahweh NJ.

Yu, X., Lau, E., Vicente, K. J. & Carter., M. W. (2002). Toward theory-driven, quantitative performance measurement in ergonomics science: the abstraction hierarchy as a framework for data analysis. *TIES, 3 (2)*,124-142.