15th ICCRTS

“C2 and Agility”

Measuring the Impact of Situational Awareness on Digitised Force Effectiveness

Paper ID Number:

136

Topics:

Topic 8: C2 Assessment Tools and Metrics

Topic 5: Experimentation and Analysis

Topic 6: Modeling and Simulation

Authors:

Dr Susan Fellows, Dstl

Paul Pearce, Dstl

Professor James Moffat, Dstl

Point of Contact:

Dr Susan Fellows

Defence Science and Technology Laboratory (Dstl)

Land Battlespace Systems Department

Room 2, Building A3,

Dstl Fort Halstead

Sevenoaks

Kent

TN14 7BP

United Kingdom

Tel:

+44 (0)1959 897226

E-mail:

scfellows@dstl.gov.uk

Measuring the Impact of Situational Awareness on Digitised Force Effectiveness

Dr Susan Fellows[1], Paul Pearce[1], Professor James Moffat[1]
[1]*Defence Science and Technology Laboratory (Dstl), UK*

Abstract

The UK Command Interim Concept outlines agility, emphasising aspects such as responsiveness, adaptability, resilience and flexibility, as a key goal towards which the UK armed forces are heading over the next 20 years. It is expected that the introduction of digitised Communications and Information Systems (CIS) as part of a Network Enabled Capability (NEC) will be a key enabler for agility. To justify investment in these systems there is a need to demonstrate that CIS will result in an improvement in operational outcome and facilitate command agility. Previous operational analysis in support of CIS procurement has struggled to develop reliable and robust methods for quantifying the effects of CIS equipped forces on operational outcome. This study describes the development of a method for analytically assessing the benefits of CIS, through the development of appropriate measures of Situation Awareness (SA) based on the Endsley model. The study used a Command and Control (C2) based wargame and simulation model to assess the effect that SA has on operational outcome. Results show that the introduction of a digitised CIS does improve SA, and that improving SA has a strong positive correlation with improvements in force effectiveness.

1 Introduction

The UK Command Interim Concept (DCDC, 2007a) outlines agility[1], as a key goal towards which the UK armed forces are heading over the next 20 years. It emphasises aspects such as responsiveness, adaptability, resilience, flexibility, robustness and innovation, while accepting that these attributes are inter-dependent. It is expected that the deployment of digitised Communications and Information Systems (CIS) as part of a Network Enabled Capability (NEC) will be a key enabler for agility. In particular, it is anticipated that CIS will provide commanders with a better understanding of the situation so that agility can be exercised. CIS systems should facilitate flexibility in approach and allow more timely decisions to be made to counter opponent's actions.

To justify investment in the systems there is a requirement to demonstrate that CIS enabled command agility will result in an improvement in operational outcome. To date, Operational Analysis (OA) studies, in support of CIS procurement, have struggled to develop robust and repeatable assessment methods to quantify the benefits of CIS within a military scenario, thereby struggling to understand how the benefits of CIS relate to battle outcome. Analysis has generally concentrated on varying the underlying physical network structure, or the delays associated with message passing, to represent the connectivity offered by the CIS under examination; generally utilising timings based measures to differentiate between the

[1] Agility has five attributes: **Flexibility** is the ability to adjust rapidly to an unforeseen and dynamic situation. **Adaptability** embraces the need to learn quickly and adjust to new experiences and situations, in particular, to amend plans that, in the light of experience, seem unlikely to lead to a suitable outcome. **Resilience** is the degree to which people and their equipment remain effective under arduous conditions or in the face of hostile action. **Responsiveness** is a measure of not only speed of reaction, but also how quickly a commander seizes (or regains) the initiative. **Acuity** is sharpness of thought, characterised by intellectual and analytical rigour, enabling intuitive understanding of complex and changing circumstances. (DCDC, 2007b)

capabilities offered by the systems. As can be seen from the NEC benefits chain in Figure 1, this results in representing 'Appropriate Connectivity' and very little in between this and the final delivery of 'Timely and Appropriate Effects'. What is needed is an ability to look at the contributions from other factors (e.g. quality of shared awareness, adaptive C2 processes, etc) to combat effectiveness and to be able to look at trading within and between these factors.

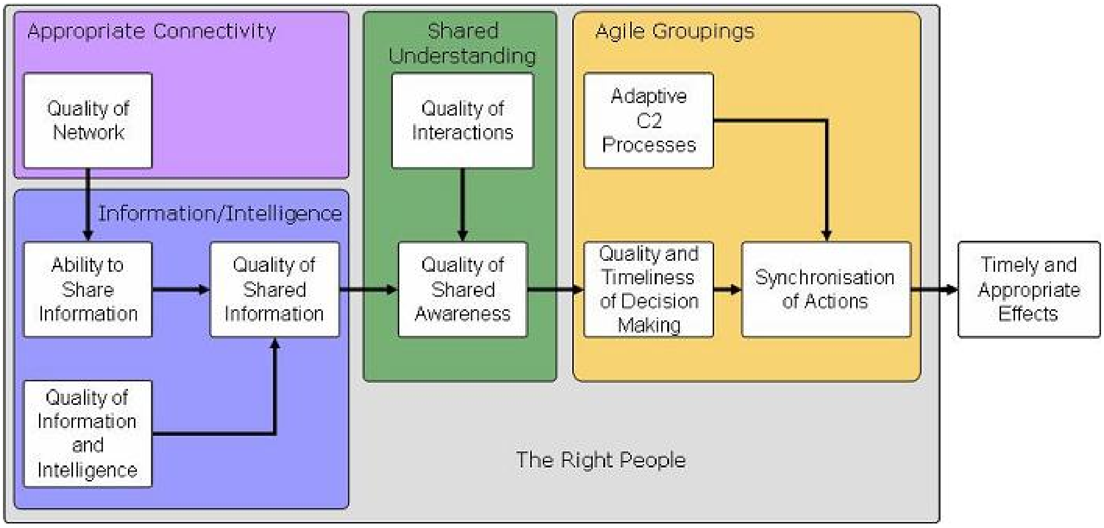

Figure 1: NEC Benefits Chain (taken from Boddington et al, 2007)

The NATO Code of Best Practice (COBP) for Command and Control (C2) Assessment (NATO, 2002) was developed to provide guidance on the analysis of C2 issues. The COBP recommends that a nested approach is taken to the development of measures. These should build from the dimensional parameters[2], through measures of performance (MoPs), measures of C2 effectiveness[3] (MoCEs), measures of force effectiveness[4] (MoFEs) to measures of policy effectiveness (MoPEs). A number of potential measures are presented in the COBP and these, along with the nested approach, have been used to inform this work.

The aim of the research was to "*develop and test a method for use on CIS studies and, to assess the suitability of the method by using it to quantify the effect of a digitised CIS on situation awareness and force effectiveness within a Battlegroup (BG) scenario.*"

In order to achieve the aim, a research hypothesis was defined that *'the digitisation of a BG HQ, subordinates and BG enablers would affect the timely delivery of appropriate effects, leading to changes in Blue force effectiveness'*. In the initial phase of the study (Fellows et al, 2008) no presumption was made that digitisation would improve force effectiveness (FE) therefore a two-tailed hypothesis was used. Results from the initial phase showed that as digitisation was improved so was FE; therefore for the latter phases of the work, as reported here, a one-tailed hypothesis has been used. The null hypothesis was that there would be no change in BG force effectiveness as a result of digitisation of a BG HQ, its subordinates and BG enablers.

This paper presents the method developed for assessing SA and, hence a method for analytically assessing the benefits of digitised CIS.

2 Experimental Design

2.1 Overview

[2] Measuring the properties inherent in the C2 system.
[3] Concentrating on the impact of C2 systems within an operational context.
[4] Considering how a force performs its mission or how well it meets it objectives.

To assess the utility of the method developed for measuring SA and FE the research hypothesis for the study was framed within the context of the overall NEC benefits chain. In particular, the work measured the SA (part of the Shared Understanding box), the outcome of the vignette (e.g. Timely and Appropriate Effects) and the impact of problems with the communications network (e.g. Appropriate Connectivity). It was anticipated that problems with the communications network would have a knock on effect through the Information/Intelligence Area to the Shared Understanding, and finally to the outcome in the experiment.

The study used a combat model, the Wargame Infrastructure and Simulation Environment (WISE)[5] (Pearce et al, 2003) as the source of data to calculate the measures. WISE was used both in wargame mode (i.e. human in the loop) and in constructive simulation mode.

2.2 SA Measures

For this work, SA was assumed to be '*knowing what is going on around you*' (Endsley, 2000). Endsley (1988) defined this more specifically as '*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future*'. The SA measures have been developed using Endsley's three level classification (Endsley, 1995): level 1 – perception of elements within the environment, level 2 – comprehension of current situation and level 3 – projection of future status. Previous research (Hone *et al*, 2006) proposed the application of the 3-Questions Model to enable the Endsley levels to be measured within an experimental framework. Hone et al.'s work proposed that quantitative level 1 SA measures should be derived by assessing 'who is where?', level 2 measures by assessing 'what are they doing' and level 3 measures by assessing 'what will they do?'.

To ensure that quantitative measures from the constructive simulation runs could be compared with the outputs of the wargaming activity, measures that could be directly derived from the simulation and compared with the subjective wargame results were required. Previous research (Salmon *et al*, 2007) has shown that self-rating techniques are quick and easy to administer when compared to freeze probe techniques and were selected for the work. Self rating quantitative measures of SA have been widely developed to measure warfighting experiments and are often combined with subjective measures, e.g. the Quantitative Analysis of Situational Awareness (QUASA) (McGuinness, 2004) technique. The objective measurement within QUASA uses a question based approach to test the truthfulness of statements regarding the recognised picture, known as a probe technique. Output from the probe technique can then be combined with subjective self rating methods. Obviously it is not possible within the constructive simulation approach adopted to ask questions and apply the probe technique used within QUASA. To overcome this limitation a means to directly measure 'probes' within the WISE constructive simulation based upon the 3-Questions Model was developed. To compare these 'probe' measures with the wargame experiments, a complementary self rating technique was administered during the WISE gaming.

Little previous research has been identified that attempts to mathematically derive measures of SA based on Endsley's levels or SA calculated directly from the combat processes undertaken within constructive simulations or wargames. Riese (2003) measured what was termed the system information, closely related to the 'who is where?' question, and was able to show how SA developed over time. A similar approach to that of Riese (2003) has been taken in this work, but the approach was extended to cover some cognitive aspects of SA. Measures for level 1 SA were calculated using the difference between the 'perception' held by an organisation in WISE and the ground truth view. Developments in other areas of the

[5] An overview of WISE can be found in Appendix 1.

defence research program (Hossain et al, 2008), to infer 'what are they doing?' and 'what will they do?'[6], were used to compute the level 2 and level 3 SA measures. Appendix 2 presents the mathematical descriptions of the quantitative measures developed. The complementary self rating technique developed, the Commander's Situation Awareness Questionnaire (COSAQ), is described in Appendix 3.

2.3 Experimental Design and Execution

The focus for the work was the BG command level, within the Land environment. The experiment was designed to compare a BG using an analogue CIS (referred to as the analogue case) and a BG using two alternative digitised CIS capabilities broadly related to the 2018 and 2025 timeframes (referred to as Epoch 3 and Epoch 4 respectively). Due to circumstances outside our control the experiment was conducted over a protracted period of time with the analogue and Epoch 4 cases wargamed approximately 12 months prior to the Epoch 3 game. During the analogue and Epoch 4 wargames the Situation Awareness Rating Technique (SART) (Taylor, 1990) was used as the self rating technique. Analysis of the previous results concluded that SART and the direct probe measures derived from the constructive simulation runs were not comparing the 'same' SA. Therefore, the COSAQ method was developed to enable a better match with the direct probe measures. For the Epoch 3 wargame both the COSAQ and SART self rating techniques were applied to enable a comparison of the techniques to be made.

The following measures of C2 and force effectiveness were used to test the hypothesis:

- Measurement of SA within the BG HQ, corresponding to a measure of C2 effectiveness;
- Measures of FE using casualties to Blue and Red and the overall Red:Blue loss exchange ratio.

The method developed, to generate the measures of C2 effectiveness and FE within the context of tactical land warfighting, used a number of tools and techniques. Gaming and simulation, using WISE, provided the core of the analysis method. The use of human in the loop games allowed courses of action resulting from the use of different CIS options to be examined. The games provided a 'vehicle' for discussion, where participants were encouraged to discuss the study question and the reasons behind their command decisions. The insights provided a valuable qualitative output to complement the quantitative analysis. Fifty replications of each case were completed as constructive simulations to examine the statistical variability around the games. Non-parametric statistical tests were used to compare the outputs from the cases. The Mann-Whitney U-Test was used to determine whether the force effectiveness measures from the cases were different at the 95% significance level and the Spearman rank correlation test was used to determine if the SA and FE measures were correlated.

Three closed wargames were played and all games had the same update rates for own side position and situation reporting between units. However, the games had different processing times for picture updating. For the analogue case an additional 90 second delay was included on the processing for all messages, to represent manual transcribing of coded messages. In addition, for the analogue case the BG HQ players worked from a Bird Table[7] and data was transmitted over a voice over internet protocol (VoIP) network. Watchkeepers[8] ensured that

[6] Indicators of Collective Behaviour.
[7] The Bird Table was a paper map situated on a central table which the players used to visualise their picture.
[8] The Watchkeeper is normally the first point of contact on the operations desk, dealing with incoming messages, actioning them as appropriate or passing them onto to the chief of staff for attention.

information on the Bird Table was updated. For the digitised cases no additional time delays were applied for processing messages and, in addition to the use of the Bird Table, the BG HQ picture from WISE was projected to represent the visualisation that would be expected from BG level CIS applications. The VoIP network was still used in the digitised cases to supplement the digitised picture. In the Epoch 3 digitised case the number of peer links within the force was limited. Within the Epoch 4 case the number of peer links was increased to represent a fully informed network. Details of the game design can be found in (Fellows & Pearce, 2009).

A BG Information Exchange Requirements (IER) set was used to provide a background level of voice and data traffic in WISE as generation of background traffic is not modelled explicitly within the model. An exercise was completed to determine appropriate message sizes and transmission times as related to key events, and the IERs were developed into a sequence appropriate to the scenario. The sequence was used, to determine the number of messages that would be passed between the organisations and the delay associated with them. These messages were injected into the experimental runs through the WISE master events list.

The context of the vignette was a BG action in a warfighting scenario. Within the scenario there was limited opportunity for ground manoeuvre so this was a stretching vignette for the assessment of the benefits of a digitised CIS. The vignette was devised carefully to ensure that the manoeuvring and actions of a BG would have an impact on the overall outcome of the scenario whilst still being credible. It should be noted that the scenario selected is ‘testing’ and significant Blue casualties were expected prior to the experimental runs. The key criteria for mission success in all of the cases were to (1) defeat the enemy lead elements, and (2) prevent them from establishing a hasty crossing site. This would, in the overall context of the scenario, have led to a significant delay on the enemy as a deliberate crossing would have been required. The scenario represents a number of BGs within a Brigade level, medium scale operation, however only one BG was explicitly controlled for the purpose of the experiment. The force effectiveness measure considers the casualties caused by the BG ‘in play’ on Red (through direct and indirect fire) and the casualties taken by this BG.

The quantitative SA measures were calculated at discrete times (every 15 minutes of model time) for the games and simulation replications. During the Epoch 3 game the SART and COSAQ questionnaires were deployed to the players at the same intervals.

3 Results

This section presents the results of the communications analysis, followed by the results of the SA and FE analysis. This enables a logical progression through the results and leads in to the discussion of the study outcomes in section 4.

3.1 Measures of C2 Effectiveness

The measures of SA have been used as substitute measures for C2 Effectiveness. The results derived from the constructive simulation replications were compared for the analogue, Epoch 3 and Epoch 4 cases. Table 1 presents a high level summary of the level 1 SA results for each experimental case, more detailed results can be found in Tables 3-1, 3-2 and 3-3 in Appendix 3. The level 2 and 3 results are not covered in this paper; further work is required to refine the Indicators of Collective Behaviour algorithm from which these measures are calculated.

Although SART was not used to compare SA with the direct probe measures from the constructive simulation runs for the Epoch 3 game, it was recorded in order to compare the results of SART with those of COSAQ. Figure 2 shows how SART compares with COSAQ.

	Scenario Time	07:01	07:15	07:30	07:45	08:00	08:15	08:30	08:45	
	Time since start of run	1	15	30	45	60	75	90	105	
Case										Overall
Analogue	Mean	0.890	0.774	0.408	0.466	0.479	0.692	0.727	0.714	0.471
	SD	0.000	0.003	0.009	0.014	0.023	0.107	0.086	0.078	0.013
Epoch 3	Mean	0.876	0.836	0.668	0.899	0.898	0.863	0.870	0.889	0.873
	SD	0.096	0.091	0.057	0.049	0.046	0.050	0.049	0.047	0.048
Epoch 4	Mean	0.876	0.768	0.572	0.744	0.806	0.709	0.767	0.796	0.733
	SD	0.096	0.080	0.068	0.053	0.049	0.056	0.050	0.054	0.048

Table 1: Summary of Level 1 SA calculated for the analogue, Epoch 3 and Epoch 4 cases

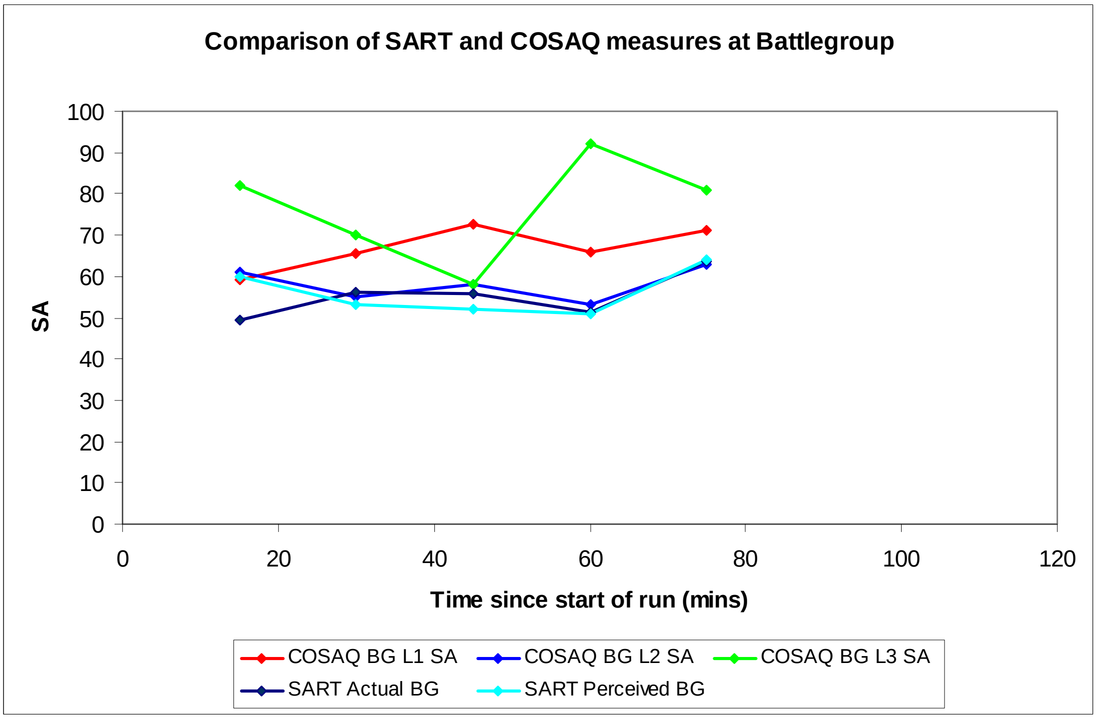

Figure 2: COSAQ and SART results for the Epoch 3 game

A validation exercise was undertaken, across all the command levels assessed using COSAQ, during which each player compared the COSAQ level 1, 2 and 3 scores with their recollections. The BG HQ values and trends shown in Figure 2, for level 1, 2 and 3 SA, were consistent with the BG commander's recollections of the battle and as such provide a useful benchmark for comparison. A comparison of SART with COSAQ shows that the SART actual and perceived scores are a relatively good match to the COSAQ level 2 score. In addition the SART actual score follows a very similar trend to the COSAQ level 1 score although it is consistently lower across all time steps. It is clear from Figure 2 that SART is not able to reflect the scores for level 3 SA. The use of SART actual and perceived measures is, therefore, not able to differentiate between the perception of elements in the environment, their comprehension and projection of future status whilst COSAQ is able to.

Figure 3 compares level 1 SA for the self rated COSAQ responses and the objective directly probed SA measure. The results show that the directly probed values are systematically higher than the COSAQ values. However, the current COSAQ results are based on a single players' assessment and therefore the interpretation of the results is, necessarily, limited. The increase in the COSAQ measure between 30 and 45 minutes is also evident in the directly probed measure. In addition there is a marginal drop in the directly probed level after 45 minutes in line with COSAQ. However, the trends in the directly probed values after 45 minutes are not as pronounced as those in the COSAQ results. The higher values observed from the directly probed measure need to be investigated further to confirm how sensitive the measure is to assumptions regarding acquisition priorities and composition.

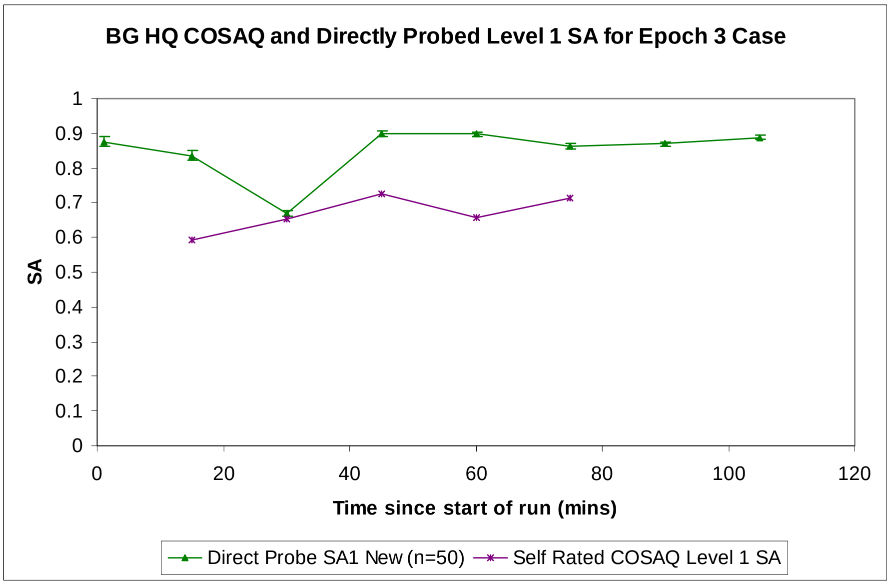

Figure 3: Comparison of Epoch 3 COSAQ and direct probe Level 1 SA measure

Figure 4 presents the mean level 1 SA measure from the simulation replications over time for the cases examined. The error bars reported in Figure 3 are the standard error of the results.

The shape of the curves for level 1 SA, in Figure 4, is consistent with key events in the games. The improved SA within Epoch 3 when compared to Epoch 4 was not expected and is related to difficulties in managing game variables[9]. These issues are explained in more detail in the discussion section. The results show that level 1 SA is better within both the digitised cases when compared with the analogue case. The analogue case approaches digitised levels only initially, when start states are similar, and towards the end of the simulation runs when the close battle is being fought.

---

[9] Steps were taken to control game variability, but in the Epoch 3 case played as part of the latter phase of experiment, a new set of players was involved. This led to significant challenges in repeating the game. In order to produce outputs that were comparable to the previous games (analogue / Epoch 4) it was necessary to script large portions of the game. However, despite this the new players used a different method of play to achieve the same objectives as before, and used Intelligence Surveillance Target Acquisition and Reconnaissance (ISTAR) assets in a more focused way.

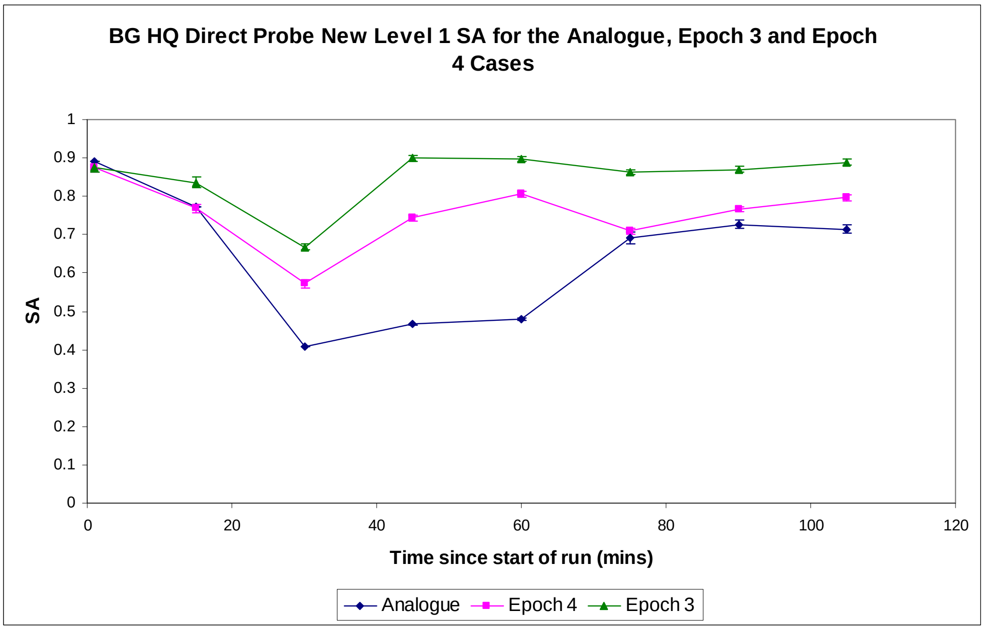

Figure 4: Comparison of Direct Probe level 1 SA for the analogue, Epoch 3 and Epoch 4 cases

The time based direct probe level 1 SA values have been combined into a single overall measure of SA (per replication) to allow a statistical comparison to be undertaken. The mechanism used to combine the results is documented in Appendix 2. Tests carried out on the distribution of results show that the distribution of SA results for the digitised (Epoch 4 and Epoch 3) cases do follow a normal distribution but those for the analogue case do not. As one of the samples is not normally distributed non-parametric statistical tests have been adopted to determine the significance of the results.

Figure 5 shows a box plot of the level 1 SA results for each of the cases examined. Both the Epoch 3 and Epoch 4 cases have an outlier which has been retained in the subsequent analysis.

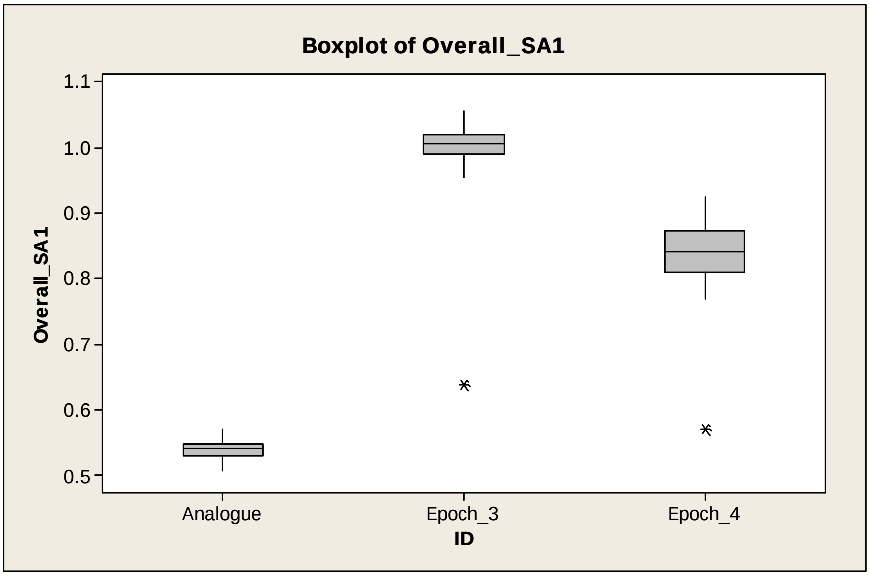

Figure 5: Boxplot of overall level 1 SA for the BG HQ for all the cases examined

The results for overall directly probed level 1 SA, as calculated from the BG HQ perception, have been compared using the Kruskal Wallis test. The Kruskal Wallis test statistic (H) has a p-value of <0.005, both unadjusted and adjusted for ties, indicating that the null hypothesis can be rejected in favour of the alternative hypothesis of at least one difference among the cases. A pair-wise comparison of the cases was completed using the Mann Whitney U test which showed a significant different between all cases as shown in Table 2.

	Analogue :Epoch 3	Analogue :Epoch 4	Epoch 3 :Epoch 4
Overall Level 1 SA	< 0.005	< 0.005	< 0.005

Table 2: Summary of Mann Whitney results for overall level 1 SA

3.2 Measures of Force Effectiveness

In the digitised cases (Epoch 3 and Epoch 4), players judged that mission success was achieved. The success was largely attributed to an improvement in SA, which allowed commanders to recognise the conditions required to start the counter-strike component of the operation earlier and in greater fidelity (through improved knowledge of enemy centre of masses and more accurate locations). The earlier H Hour, combined with more focused strikes in both ground manoeuvre and fires allowed enemy crossing sites and bridgeheads to be identified early and dealt with quickly. In the analogue case, reduced SA did not allow sufficient resolution or indeed warning of enemy actions, so crossing sites and bridgeheads were not dealt with in sufficient time or with sufficiently focussed combat power.

Figure 6 shows the force effectiveness results for personnel and vehicle losses to Red and Blue for the cases examined. A comparison of vehicle casualties for both the digitised cases (Epoch 3 and Epoch 4) with the analogue case using the Kruskal Wallis test showed the differences to be significant[10]. Table 3 presents results of a pairwise comparison using the Mann Whitney U test showing that in some instances the results are not significantly different. Of particular note is that the Epoch 4 Blue and Red personnel losses and Red vehicle losses are significantly different from the Epoch 3 case. The scale of the difference was unexpected and is likely to be due to the difference in the way in which the two games were played[11].

Pair-wise comparison			
Attribute	**Analogue :Epoch 3**	**Analogue :Epoch 4**	**Epoch 3 :Epoch 4**
Red Personnel Losses	0.211	<0.005	<0.005
Red Vehicle Losses	<0.005	<0.005	<0.005
Blue Personnel Losses	<0.005	0.140	<0.005
Blue Vehicle Losses	<0.005	<0.005	0.005

Table 3: Summary of Mann Whitney results for personnel and vehicle losses

Table 4 presents a summary of the results for force effectiveness, more detailed results can be found in Table 3-4 in Appendix 3.

[10] The Kruskal Wallis showed: the difference in Red vehicle losses between cases to be significantly different (p = 0.005, the difference in Red personnel losses between cases to be significantly different (p = 0.005), the difference in Blue vehicle losses between cases to be significantly different (p = 0.005), and the difference in Blue personnel losses between cases to be significantly different (p = 0.005).

[11] In particular the Blue lower controller used a tactic to avoid incoming Red indirect fire, when he was aware of it. This led to a significant reduction in Blue personnel casualties with respect to the other two cases where such a tactic was not employed.

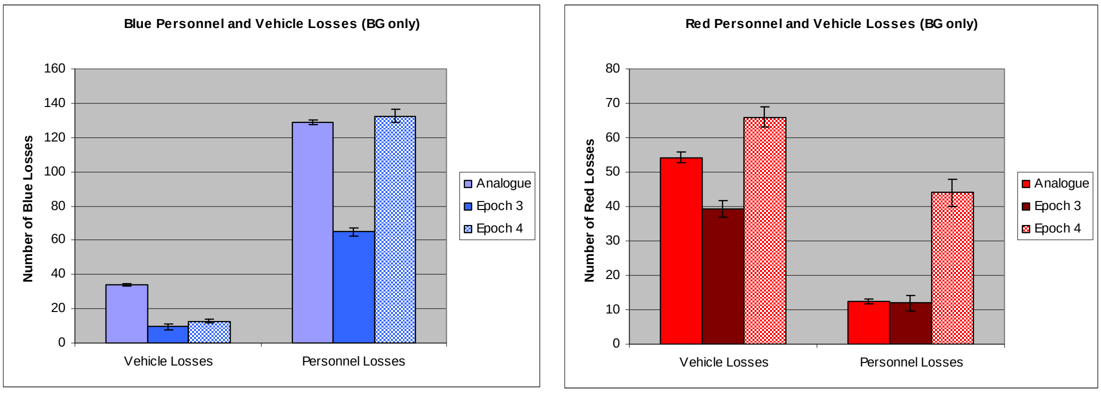

Figure 6: Blue and Red casualties and vehicle losses by case (casualties refers to personnel).

	ANALOGUE CASE			EPOCH 3 CASE			EPOCH 4 CASE		
	Total Red Losses	**Total Blue Losses**	**Red / Blue LER**	**Total Red Losses**	**Total Blue Losses**	**Red / Blue LER**	**Total Red Losses**	**Total Blue Losses**	**Red / Blue LER**
Mean	67.90	163.06	0.42	51.22	79.30	0.65	109.98	149.44	0.75
SD	11.77	13.28	0.08	12.75	11.29	0.17	18.84	14.91	0.17

Table 4: Summary of Red and Blue Losses for the analogue, Epoch 3 and Epoch 4 cases

Figure 7 shows the overall loss exchange ratio (Red losses:Blue losses). The mean Loss Exchange Ratio (LER) for the analogue case (M=0.41, SD=0.08) is less than the mean LER for the Epoch 3 case (M=0.65, SD=0.17), which is in turn less than the mean for the Epoch 4 case (M=0.75, SD=0.17). The Kruskal Wallis test statistic (H) has a p-value of <0.005, both unadjusted and adjusted for ties, indicating that the null hypothesis can be rejected in favour of the alternative hypothesis of at least one difference among the cases.

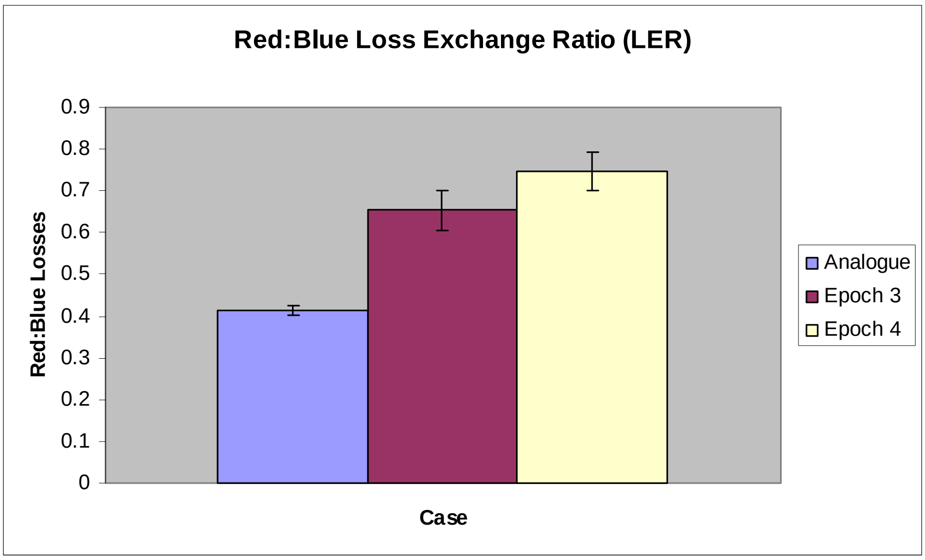

Figure 7: Loss Exchange Ratio for the analogue, Epoch 3 and Epoch 4 cases

Mann Whitney U tests were run between the cases. The results are summarised in Table 5. On the basis of these results the null hypothesis would be rejected for all combinations thereby supporting the hypothesis that the digitisation of a BG HQ, superiors, subordinates and BG enablers would improve the timely delivery of appropriate effects, leading to improvements in Blue force effectiveness.

Cases compared		p-value
Analogue	Epoch 3	<0.005
Analogue	Epoch 4	<0.005
Epoch 3	Epoch 4	0.006

Table 5: Mann Whitney U Test results for LER analysis

The LER presented in Figure 7 shows that despite improvements from digitisation the Blue losses are still higher than Red losses in both cases. As discussed earlier this is a testing scenario for Blue so this result is not unexpected.

Figures 8 and 9 show scatter plots of the overall directly probed level 1 SA against force effectiveness (Red:Blue LER) for the analogue and Epoch 3 cases and the analogue and Epoch 4 cases respectively. The results are based on fifty replications of each case. Figure 8 shows that there is a strong positive correlation between level 1 SA and force effectiveness for the analogue and Epoch 3 cases. A Spearman Rank test shows the correlation to be significant (r=0.611, n=50, p<0.005).

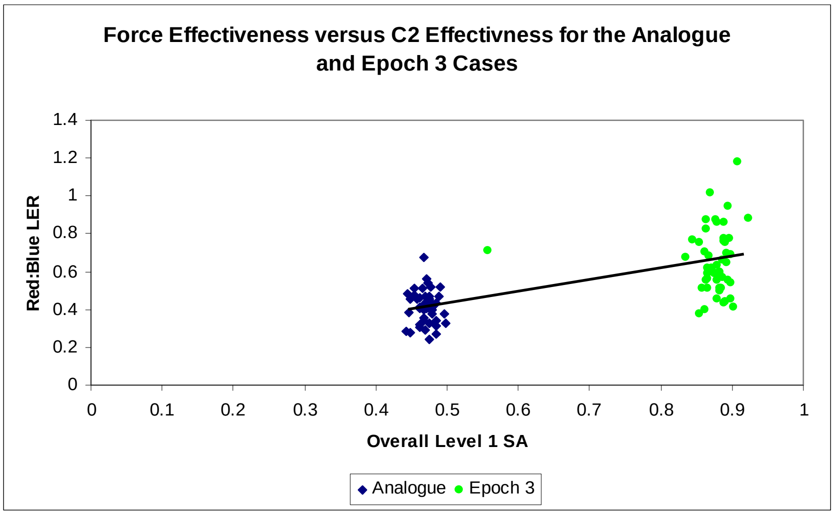

Figure 8: Plot of Red:Blue Loss Exchange Ratio against Level 1 SA for the analogue and Epoch 3 cases showing correlation

Figure 9 shows that there is a strong positive correlation between level 1 SA and force effectiveness for the analogue and Epoch 4 cases. A Spearman Rank test shows the correlation to be significant (r=0.700, n=50, p<0.005).

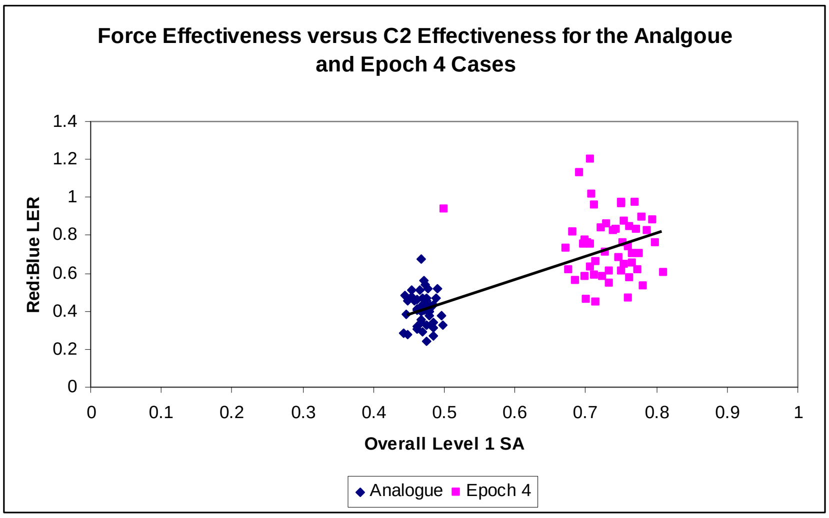

Figure 9: Plot of Red:Blue Loss Exchange Ratio against Level 1 SA or the analogue and Epoch 4 cases showuing correlation

Analysis of the distribution of LERs across all cases shows that the distribution of results in the analogue case is smaller than the distribution in both the digitised cases. The spread and shape of the distributions in the Epoch 3 and Epoch 4 cases is similar although with different mean values. The results show that, although SA is improved through the introduction of a digitised CIS and that force effectiveness improves, there is a greater variation in force effectiveness, and overall Level 1 SA, in the digitised cases when compared to the analogue.

4 Discussion

The overall method has been shown to be effective in testing the hypothesis that digitisation improves force effectiveness. Comparison of the Epoch 3 and Epoch 4 simulation results for force effectiveness shows that digitisation improves force effectiveness when compared to an analogue CIS case. Interestingly the spread of outcomes for force effectiveness is similar in the digitised cases but very different to that of the analogue showing that, although SA is improved through the introduction of a digitised CIS and that force effectiveness improves, there is significant variation in force effectiveness distributions for the samples. Further investigation will be required to determine if this is an artefact of the model, players, scenario, algorithms developed, or whether there is some inherent property of a digitised system that generates the differences.

The improved SA in the Epoch 3 CIS case when compared with the more networked Epoch 4 CIS capability did not generate the anticipated results. The force effectiveness results followed the expected trend but the SA results for Epoch 3 were better than those for Epoch 4. Exercising an appropriate degree of experimental control within the Epoch 3 game to enable sensible comparison with Epoch 4 was difficult. It is not possible, at this stage, to draw any meaningful conclusions regarding the sensitivity of the method for distinguishing between these two digitised cases. The difference between the games potentially skews the LER as well as future Blue actions and will, inevitably, force greater divergence as the game progresses. More detailed examination of the output from the games showed that there was significantly more ISTAR coverage over the main body of Red in the Epoch 3 game when compared to the Epoch 4 game, thereby resulting is an improvement in the Blue operational picture at the BG HQ in the Epoch 3 case, which is reflected in the higher SA values calculated.

Additionally, in the Epoch 3 game the Formation Reconnaissance units were not withdrawn whilst in the analogue and Epoch 4 games they were. Therefore, in the Epoch 3 game information continued to be fed into the BG by these units as Red advanced through their line. The significant difference in the use of ISTAR assets in the Epoch 3 and Epoch 4 games means that it is impractical to draw substantive conclusions of the SA from the two digitised games. However, it is encouraging that comparison of the analogue versus Epoch 3 and analogue versus Epoch 4 games (i.e. analogue versus digitised) shows increases in SA in the anticipated direction, i.e. the introduction of digitisation improves SA.

The decision not to withdraw the Formation Reconnaissance in the Epoch 3 game may also have affected the attrition on both sides. For example, by not withdrawing Formation Reconnaissance the Red advance may have slowed in response to the increased threat level. The threat level may have decreased when the Formation Reconnaissance had been eliminated or bypassed, thereby allowing Red to reach Blue killing areas before Blue had all forces into position. By contrast in the other cases the threat may have remained fairly stable as the Formation Reconnaissance was withdrawing in the face of the Red advance, resulting in a more drawn out Red advance. The difference in the Red advance rates would have given

Blue an opportunity to get into position to exploit the killing areas. The killing areas were only fully exploited in the Epoch 4 digitised case.

The comparison highlights the differences in play between all three games and in particular illustrates how the introduction of a digitised system, and effective use of force elements, can lead to different courses of action. While the gaming highlights a benefit of digitisation, a number of complex relationships have been introduced making it difficult to determine if the change in force effectiveness between Epochs 3 and 4 is a result of the change in the CIS. The difference observed may be related to the way in which the games were played rather than the change in the CIS. Despite these differences the results indicate that the method developed is robust in demonstrating that digitisation offers a significant improvement over analogue systems. It also shows that the means of measuring SA is sensitive to the capabilities of the ISTAR platforms and their employment.

When considering the enhanced measures of SA, the application of the COSAQ questionnaire has enabled a better understanding of the SA as it evolves over time in a scenario. In particular it is possible to compare the differences between levels 1, 2 and 3 SA and use these as a means of discussion during gaming plenary sessions. Comparison of the self rating techniques, SART and COSAQ, has shown that the SART trends were not unreasonable but SART was not able to adequately represent the levels of SA as scored by COSAQ. The COSAQ technique was validated by the players across the command levels played.

Assessment of the directly probed SA measures has shown that the method developed is a reasonable approximation when comparing level 1 SA to COSAQ. However, further refinement is required for level 2 and level 3 SA.

Calculating a quantifiable form of SA also enabled a correlation to be undertaken against FE. This showed that the two measures are correlated so as SA improves force effectiveness improves. Although the results show a correlation further research is required to definitively establish cause and effect.

5 Conclusions

In terms of measuring SA it is recognised that there are numerous definitions and theories (Salmon et al, 2008a; Salmon et al, 2008b; Hone 2008). However, Endsley's theory provides a level of structure within which SA could be measured and has good academic grounding and supporting evidence (Endsley, 1995). As such the theory was used to ensure expansion of the analytical understanding of SA by expressing C2 effectiveness in terms of the levels defined by Endsley.

The current research has successfully developed two complimentary methods to calculate SA in combat models. However, both will benefit from expansion to cater for more explicit representations of accuracy and completeness, comprehension and projection. In addition the difficulties in experimentally controlling gaming activity will require resolving and may be offset through the expansion of the directly probed method when undertaking constructive simulation studies. However, despite the gaming difficulty the methods have enabled an assessment of SA to be made and can be used to stimulate further discussion about objective SA measurement. The assessment of SA showed that there is a strong positive correlation between Level 1 SA and force effectiveness, i.e. as SA improves through the introduction of a digitised CIS force effectiveness improves, although it remains to be seen if cause and effect can be established.

In summary the method developed has successfully differentiated between an analogue and digitised BG showing that improvements in force effectiveness can be achieved with the introduction of a digitised CIS.

6 References

Boddington S et al (2007), *Quantitative Evidence of the Benefits and Risks of NEC (Fourth Iteration)*, internal report

DCDC (2007a), *Command Interim Concept*, UK Army Concept Paper

DCDC (2007b), Operate Interim Concept, UK Army Concept Paper

Endsley M (1988), *Design and Evaluation for Situation Awareness Enhancement*, in Proceedings of the Human Factors Society 32nd Annual Meeting, Vol 7, pp97-101

Endsley M (1995), *Towards a Theory of Situation Awareness in Dynamic Systems*, Human Factors, 37(1), pp 32-64

Endsley M (2000), *Theoretical Underpinnings of Situation Awareness: A critical review*, in Situation Awareness Analysis and Measurement, eds Endsley, M, Garland, D, Lawrence Erlbaum Associates, ISBN 0-8058-2133-3

Fellows S, Dandy C, Russell C (2008), *Tactical Communications Implications Study (TCIS): Assessment of method developed to measure digitised force effectiveness*, internal report

Fellows S, Pearce P (2009), *Tactical Communications Implications Study (TCIS): Enhancement to method developed to measure digitised force effectiveness*, internal report

Hone G (2008), *Guest Editor's Introductory Remarks*, International C2 Journal, Vol 2 No 1, Special Issue 'Awareness' (http://www.dodccrp.org/html4/journal_v2n1_00.html)

Hone G, Martin L, Ayres R (2006), *Awareness – does the acronym "SA" still have practical value?*, 11th International Command and Control Research and Technology Symposium

Hossain A, Walmsley N, Pearce P (2008), *A Minimum Spanning Tree Approach to Identifying Collective Bahviour and Inferring Intent for Combat Models*; International C2 Journal; Vol 2 No 2 (http://www.dodccrp.org/files/IC2J_v2n2_01_Hossain.pdf)

McGuinness B (2004), *Quantitative Analysis of Situational Awareness (QUASA): Applying Signal Detection Theory to True/False Probes and Self-Ratings*, Command and Control Research and Technology Symposium

Moffat J, 2002, *Command and Control in the Information Age: Representing its* Impact, The Stationary Office, London, UK

NATO (2002), *NATO Code of Best Practice for C2 Assessment*, CCRP, ISBN 1-893723-09-7

Parsaye K, Chignell M (1998), *Expert Systems for Experts*, Wiley

Pearce P,, Robinson A., Wright S, (2003), *The Wargame Infrastructure and Simulation Environment (WISE)*, Knowledge Based Intelligent Information and Engineering Systems, 7th International Conference, KES 2003, Oxford, UK

Robinson,A (2000), *The MOD Modelling and Simulation Strategy*, internal report.

Roberts N (2008), *Bayesian Inferencing*, internal report

Riese LTC SR (2003), *Quantifying Information Availability for Situation Awareness*, MORS Workshop on Decision Aids/Support to Joint Operations Planning

Salmon PM., Stanton, NA, Ladva D, Jenkins DP, Walker GH, Rafferty L (2007) *Measuring Situation Awareness During Command and Control Activity: A Comparison of Measures Study*, HFIDTC/2/1.2.5/3, (version 2)

Salmon P et al (2008a), *What really is going on? Review of Situation Awareness Models for Individuals and Teams*, Theoretical Issues in Ergonomics Science, Vol 9, No 4, pp 297-323

Salmon P et al (2008b), *Measuring Situation Awareness in Complex Systems: Comparison of Measures Study*, submitted to International Journal of Industrial Ergonomics

Taylor R.M (1990), *Situational awareness: aircrew constructs for subject estimation*; IAM-R-670;

Acknowledgements

The authors would like to acknowledge the significant contributions made by Dr Catherine Dandy and Major Colin Russell in relation to the experimental work carried out. We would also like to thank Nick Walmsley, Nick Roberts and Michael Richardson for the many useful discussions on SA measures and for their extensive assistance in developing the mathematical representations.

Appendix 1 Overview of WISE

WISE is a stochastic, event driven model that allows decisions to be made by players, software algorithms or a combination of both. It is a Land focussed model (with a representation of air and maritime support to Land operations) at the system level, which can represent warfighting, peace support, or stabilisation operations. By using players rather than software algorithms to make decisions, WISE is played as a Synthetic Environment (SE), whereas with software representing human decision-making, WISE is run as a closed form, constructive simulation. Architecturally, WISE is a Personal Computer (PC) based system, written in C++ utilising a number of open source software products under the RedHat Enterprise Linux operating system, with a modelling approach centred on the use of software agents within a distributed network. The use of agents allows a 'loose coupling' between WISE system components which in turn enhances the capability of WISE to represent various approaches to command decision-making.

The flexibility of the modelling approach allows for scenarios to range in scope from Army Divisional level to Company level. When used in SE, or 'wargaming' mode, WISE is typically used to facilitate discussion on a topic of interest and as such is able to provide insight into a number of 'lines of development ' such as equipment, training and organisation. Orders and decisions produced during a particular wargaming experiment can be captured and used as input to the closed form, constructive simulation mode. The use of the wargaming mode allows a rich and detailed exploration of a particular scenario. This is complemented by the closed form simulation mode which allows for a number of excursions to be made to investigate the robustness of the results to perturbations in input parameters.

The architecture of WISE ensures that it can represent a sufficiently rich variety of Command and Control (C2) capabilities as stated within the MOD model strategy (Robinson 2000), which required:

- The ability to represent C2 at all levels: the spectrum of operations means political leaders will take a more active role in the decision-making process. Models should allow high level policy decisions to impact on the use of assets even at the tactical level since in non-warfighting operations individual tactical assets may have strategic importance.
- The need to represent the C2 structure flexibly: flexible structures are required to cope with new organisational force structures, ad hoc force packages and ad hoc coalitions.
- The need to represent joint C2: increasingly operations are of a joint or multi-national nature and the inter-relationships between services, coalition partners and non-governmental organisations need to be considered.

Rather than giving a detailed description of all of the physical models in WISE (i.e. the processes that cover direct fire, indirect fire, surveillance and target acquisition, etc.) the remainder of this section concentrates on detailing how the C2 has been designed and implemented.

WISE avoids the explicit representation of equipment types and force elements by considering force elements as organisations that have roles and resources. As such a scenario could consist of a number of organisations, associated decision-making roles and physical resources. These characteristics can then be used to describe any force element or equipment type to be represented within a scenario. For example, it is possible to consider an organisation that represents Challenger 2 tanks as an aggregated group of physical resources and with an associated "squadron commander" role, or an organisation that represents an individual Challenger 2 tank as a physical resource, e.g. "Tank1" with associated roles of "commander", "Gunner", "Loader/Signaller" and "Driver". Each role can be assigned an appropriate model for its required human behavioural representation.

An organisation can be assigned any number of roles. Each role within the organisation undertakes processing on receipt of information to update the organisation's operational picture. The picture is based on a perception of the environment and is used to drive its decision-making. Roles can also be assigned to physical resources allowing attacks on physical resources to directly affect the processing capabilities of an organisation. For example, if a force commander is making decisions from a command vehicle that is destroyed through enemy action, an important decision-making function is removed from the organisation.

Information arriving at an organisation is processed and stored within an organisation's perception. Domain objects are used as a means of representing data about the domain of operation within an organisation's perception. Domain objects are organised into inheritance hierarchies to reduce the need for repetition, and are an example of frame-based knowledge structuring (Parsaye & Chignell 1998). All instances that are represented in the scenario are described by a domain object, and this provides a means by which any type of object can be considered, whether it is the United Nations, an individual tank or a sensor.

Relationships can occur between domain objects of the same or different types. Through such relationships the analyst has the ability to logically link domain objects to allow the creation of any type of C2 structure. These relationships can be used to provide a link to any other type of organisation object and effectively define the C2 structure. Because each organisation descriptor contains the relationships, it is possible to define multiple hierarchies for a single organisation. For example an organisation designated as "1_UK_Brigade" could have separate organisation role descriptors for artillery and for the C2 of companies and squadrons, and hence represent a number of different command chains.

The C2 links between organisations define how information flows during a scenario. It is possible for links to be broken and re-assigned dynamically during a wargame or simulation giving the ability to represent the formation and break up of self–synchronising, agile task organised groups. Each C2 link has delays associated with the passage of information which can be used to represent physical communications delays or delays as a result of C2 staff processes or functions.

The cycle of processing within an organisation is representative of a situation assessment process leading to an end state that represents the organisation's situation awareness. A role within an organisation can be represented by any appropriate human decision-making model or by a human player. When multiple roles are defined for an organisation, WISE is implicitly representing shared situation awareness within the organisation.

In order to distinguish between levels of situation awareness within an organisation, WISE expresses situation awareness within the context of the model of situation awareness proposed by Endsley (Endsley 1995). All decision-making within WISE is based on an organisation's perception of the environment, which is essentially a representation of Level 1 situation awareness within Endsley's model. Players within the wargame are presented with a fused organisational picture from which they would undertake Level 2 and Level 3 situation awareness. Within a constructive simulation run this would be undertaken by the Rapid or Deliberate Planners (Moffat 2002).

Appendix 2 SA Measures

This appendix describes how the quantitative values of situation awareness have been derived based on the levels defined within the Endsley model (Endsley, 1995). Each of the calculations is covered below. The overall SA values have been used in the correlation analysis with force effectiveness.

LEVEL 1 SA

All of the analysis was conducted at a unit level. It considered the friendly units, in the direct command chain of the organisation under examination, as well as the enemy / neutral acquisitions in the perception of the organisation. Acquisitions of individual platforms were removed from the analysis as in WISE these always represent platforms. A filter was used to remove all 'old' acquisitions. In addition, the analysis removed duplicate acquisitions (retaining the most recent) and used a 30km by 30km assessment box.

To calculate level 1 SA a utility function was constructed to produce a score for the acquisition based on the error in location[12]. Although it would be possible to define the utility function based on the type of organisation, error in location and range from the acquiring organisation, for this work a single utility function scoring error in location was used (see Figure 2-1).

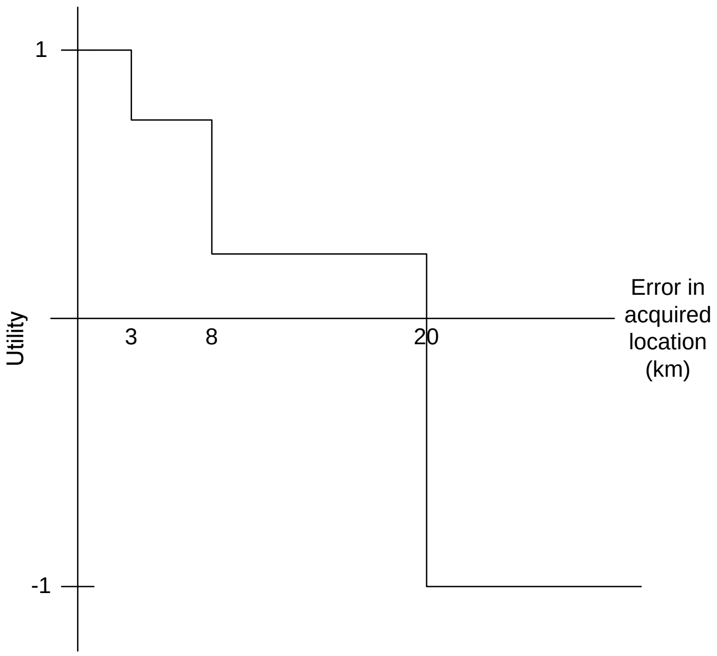

Figure 2-1: Utility function used in the original calculations

Table 2-1 demonstrates how the utility functions have been used to generate a score for the unit or acquisition being assessed.

[12] This is calculated as the modulus of the perceived location minus the ground truth location of the unit.

Unit	Seen	Range (km)	Error in Location (km)	Score (Range[13], Error in Location)
T_1	1	α	A	x
T_2	0	β	B	y
T_3	1	α	C	z
…	…		…	…
T_n	1	β	B	Y
Count of all units (N)	**Count of units in perception (n)**			**$U(R,\Delta_{Location})$**

Table 2-1: Information required for Level 1 SA calculation

In previous calculations of level 1 SA (Fellows et al, 2008), for TCIS, no account was taken of the composition of the unit. In the calculation for this phase of work Bayesian inferencing (Roberts, 2008), implemented in WISE to support the Indirect Fire Precision Attack (IFPA) procurement, was utilised to provide a convenient mechanism to account for the contribution that unit composition makes to level 1 SA. Level 1 SA is now defined as in Equation 1.

$$SA1 = \frac{\sum_{i=1}^{N} \delta_i U_i \left(R, \Delta_{location}\right) F_i(C_{perceived})}{\sum_{i=1}^{N} F_i(C_{GroundTruth})}$$

Equation 1

Where δi={0,1} and indicates whether a unit exists in the perception, i represents the units over which the summation is completed with N covering all of the friendly, enemy and neutral units, Ui(R,ΔLocation) is the utility score for the unit based on range from the assessing unit and error in its known location.

Fi(Cperceived) represents a function defining the contribution of the composition of the unit based on the acquisitions (Cperceived) achieved of its resource groups, in the perception being assessed. The score will reflect the priority of the most likely unit type and the confidence in the inference, along with a contribution from the unit type being assessed as UNKNOWN. Fi(CGroundTruth) represents a similar function but based on the actual unit type, which will simplify down to the priority of the unit type.

It is assumed that Fi(C) will have a form of f(proposed type, confidence, UNKNOWN, 1-confidence) for the perceived items and f(actual type, 1, UNKNOWN, 0) for the ground truth component. It is expected that it will be calculated using:

$$F_i\left(C_{perceived}\right) = P\left(proposed\right) \times Confidence + P\left(UNKNOWN\right) \times \left(1 - Confidence\right)$$

Equation 2

$$F_i\left(C_{GroundTruth}\right) = P(actual)$$

Equation 3

Where P(proposed) is the priority of the proposed (assessed) unit type and P(actual) is the priority of the actual unit type.

[13] For the current work, range was not assessed as a variable.

LEVEL 2 SA

The collective behaviours algorithm developed under the OA Domain (Hossain et al, 2008) was used to infer clusters of acquisitions and to represent a level of comprehension. Level 2 SA was calculated as follows:

a. Let the number of clusters at time step i be ni. Let cluster j have mj members and cj is the confidence that the members form a cluster.

b. If there are Ei known entities at time step i, then the proportion assigned to a cluster is:

$$Assigned_i = \frac{\sum_{j=1}^{n_i} m_j}{E_i}$$

Equation 4

c. c.Let the overall confidence at time step i be:

$$Confidence_i = \frac{\sum_{j=1}^{n_i} m_j c_j}{\sum_{j=1}^{n_i} m_j}$$

Equation 5

d. The final level 2 SA measure is calculated as in equation 6:

$$SA2 = \frac{Assigned_i * Confidence_i}{2}$$

Equation 6

LEVEL 3 SA

The calculation of level 3 SA is based on the clusters that are inferred and represents a measure of intent with reference to set of defined objectives. Rankings were generated from the collective behaviours algorithm to be used for this measure.

Consider the ranking of a cluster j moving towards a set of objectives {o1, o2, …, on}. Let rj,k be the ranking score for cluster j moving towards objective k. In the calculations, m is the number of clusters and n is the number of objectives. For each cluster j we assume that:

$$\sum_{k=1}^{n} r_{j,k} = 1$$

Equation7

The mean ranking score for cluster j is defined as:

$$\bar{r}_j = \frac{1}{n}$$

Equation 8

Consider a single cluster that could move towards any of n different objectives. Perfect situation awareness would equate to a ranking score of one objective being one and that for all of the other objectives being 0. The worst situation awareness state would be one in which the ranking of all of the objectives was the same i.e. 1/n.

The standard deviation (SD) given n objective rankings (ri) is:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n}}$$

Equation 9

The minimum value of the SD is 0 (when all of the rankings are equal. The maximum value of the SD is:

$$SD_{\max} = \sqrt{\frac{(1-\bar{r})^2 + (n-1)\bar{r}^2}{n}}$$

Equation 10

Therefore the expression for the total level 3 SA would be:

$$SA3 = \frac{\sum_{j=1}^{m} m_j \left(\max\{r_{j,k}\}_{k=1}^{k=n} - \bar{r}_j \Big/ SD_{j,\max} \right)}{\sum_{j=1}^{m} m_j}$$

Equation 11

Where mj is the number of members in cluster j and SDj,max is:

$$SD_{j,\max} = \sqrt{\frac{(1-\bar{r}_j)^2 + (n-1)\bar{r}_j^2}{n}} = \frac{\sqrt{n-1}}{n}$$

Equation 12

OVERALL SA

To compare SA at each of the time steps, within the war game or within a replication, the SA measures were computed as the (arithmetic) mean of the values for each replication.

When deriving a value across all times for a given war game or replication decision time were considered. For each case (digitised versus analogue), there was a time point at which Blue must make the decision to commit his forces; let this time be t_d[14]. Information available to the Blue commander after t_d is worth less than information available before this time, because the major decision has (or should have) been made. It is further assumed that information available immediately before td is worth more than information available a long time before t_d.

[14] As part of the gaming protocols there will be a requirement to formally record the decision point. In instances where multiple decision points are identified then these should all be recorded so that we can consider the most appropriate weightings to apply.

To represent the latter effect, the SA measures at each time step (t) prior to t_d are weighted according to a set of geometric weights: given n time steps, each weight, w_i, is defined as:

$$\omega_i = \omega_{t_d} r^{(|t_d - i|)\delta t} (1 \le i \le t_n)$$

Equation 13

given a is w_1, and δt is the interval between the time steps. For simplicity, we set w_1 to be 1.0, to give weights relative to the most recent value.

The sum of n terms is given by:

$$S_n = \frac{(1 - r^n)}{(1 - r)} \text{, for } r \neq 1.$$

Equation 14

Therefore, assuming the time step immediately prior to td makes the most important contribution to the overall SA picture and the ith weight represents the time step (i-1) steps before that, the weight for each SA value as a function of time is:

$$\hat{w}_i = \frac{w_i}{S_n}$$

Equation 15

The ratio, r, should be set to the (average) time step as a proportion of the overall time span. For simplicity, we assume that the contribution of SA available after the key decision time is not as valuable as SA just prior to the decision time.

The SA measure across all of the time steps examined for a replication for a particular case was defined to be:

$$SA(case, replication) = \sum_{i=1}^{n} \omega_i SA_i$$

Equation 16

Where SAi is the value calculated for the SA (covering Levels 1, 2 and 3) within the time step i.

Appendix 3 Commander's Situation Awareness Questionnaire

A Commander's Situation Awareness Score questionnaire was used for TCIS (see Figure 3-1). The questions relate to the quantitative measures that are derived from the WISE gaming and simulation output. A similar scoring approach has been adopted as that used for SART.

Commander's Situation Awareness Score (version 1.0)

Command level:	**Side:**	**Scenario:**
Command role:	**Game time:**	**Game identifier:**
Organisation:	**Date:**	

		0 (low) … 50 … 100 (high)
1.	How accurate do you consider the positions of friendly forces in your area of interest to be? Very accurate (high) or very inaccurate (low)	low … high
2.	How aware are you of all possible friendly forces in your area of interest? Very aware (high) or not very aware (low)	low … high
3.	How accurate do you consider the positions of enemy forces in your area of interest to be? Very accurate (high) or very inaccurate (low)	low … high
4.	How aware are you of all possible enemy forces in your area of interest? Very aware (high) or not very aware (low)	low … high
5.	How close to ground truth do you think your current operational picture is with respect to awareness of the number of enemy forces within your area of interest? Very close (high) or not very close (low)	low … high
6.	How close to ground truth do you think your current operational picture is with respect to awareness of the number of friendly forces within your area of interest? Very close (high) or not very close (low)	low … high
7.	How aware are you of the current situation or status of friendly forces, e.g. force strengths? Very aware (high) or not very aware (low)	low … high
8.	How aware are you of the current enemy disposition in your area of interest, e.g. formations? Very aware (high) or not very aware (low)	low … high
9.	How confident are you that you have an adequate awareness of the current composition, e.g. size and type, of enemy units and sub-units? Very confident (high) or not very confident (low)	low … high
10.	How concerned are you about the proximity of the current enemy threat to your own forces centre of mass? Very concerned (high) or not very concerned (low)	low … high
11.	If appropriate, how accurate do you consider the current speed and heading of the enemy forces to be? Very accurate (high) or not very accurate (low)	low … high
12.	How confident are you in your understanding of the intent of friendly forces? Very confident (high) or little confidence (low)	low … high
13.	How confident are you in your understanding of the intent of the enemy? Very confident (high) or not very confident (low)	low … high
14.	How would you rate the quality of your operational picture? High quality (high) or low quality (low).	low … high
15.	In your estimate what would be an acceptable level of quality for the operational picture? High quality (high) or low quality (low).	low … high

1. Please mark a point on the scale and label it with the appropriate value

Figure 3-1: Commander's situation awareness questionnaire

In order to relate the data captured from the questionnaire to the measures derived from WISE a set of simple equations have been produced. Table 3-1 shows how each question related to the type of attributes and level of SA that it is measuring.

Question No	Question	Attribute Mapping	Attribute Name
1	How accurate do you consider the positions of friendly forces in your area of interest to be? Very accurate (high) or very inaccurate (low).	Δ location	δLOC_{OWN}
2	How aware are you of all positive friendly forces in your area of interest? Very aware (high) or not very aware (low).	N	N_{OWN}
3	How accurate do you consider the positions of enemy forces in your area of interest to be? Very accurate (high) or very inaccurate (low).	Δ location	δLOC_{ENEMY}
4	How aware are you of all possible enemy forces in your area of interest? Very aware (high) or not very aware (low).	N	N_{ENEMY}
5	How close to ground truth do you think your current operational picture is with respect to awareness of the number of enemy forces within your area of interest? Very close (high) or not very close (low).	F_i(GT)	GT_1
6	How close to ground truth do you think your current operational picture is with respect to awareness of the number of friendly forces within your area of interest? Very close (high) or not very close (low).	F_i(GT)	GT_2
7	How aware are you of the current situation or status of friendly forces, e.g. force strengths? Very aware (high) or not very aware (low).	F_i(Perceived)	$COMP_{OWN}$
8	How aware are you of the current enemy disposition in your area of interest, e.g. formations? Very aware (high) or not very aware (low).	Level 2	$FORM_{ENEMY}$
9	How confident are you that you have an adequate awareness of the current composition, e.g. size and type, of enemy units and sub-units? Very confident (high) or not very confident (low).	F_i(Perceived)	$COMP_{ENEMY}$
10	How concerned are you about the proximity of the current enemy threat to your own forces centre of mass? Very concerned (high) or not very concerned (low).	Range	R
11	If appropriate, how accurate do you consider the current speed and heading of the enemy forces to be? Very accurate (high) or not very accurate (low).	Level 2	$MOVE_{ENEMY}$
12	How confident are you in your understanding of the intent of friendly forces? Very confident (high) or little confidence (low).	Level 3	$INTENT_{OWN}$
13	How confident are you in your understanding of the intent of the enemy? Very confident (high) or not very confident (low).	Level 3	$INTENT_{ENEMY}$
14	How would you rate the quality of your operational picture? High quality (high) or low quality (low).		
15	In your estimate what would be an acceptable level of quality for the operational picture? High quality (high) or low quality (low).		

Table 3-1: Commander's situation awareness questionnaire scoring framework

The COSAQ Level 1 SA was calculated based on the responses to the questionnaire as defined by equation 17.

$$SA_1 = \frac{(\delta LOC_{OWN} \times COMP_{OWN}) + (\delta LOC_{ENEMY} \times COMP_{ENEMY})}{(GT_1 + GT_2)}$$

Equation 17

The COSAQ Level 2 SA was calculated based on the responses to the questionnaire as defined by equation 18.

$$SA_2 = \frac{COMP_{ENEMY} + MOVE_{ENEMY}}{2}$$

Equation 18

Level 3 SA is simply the value from question 13 or $INTENT_{ENEMY}$.

Appendix 3 Detailed Results

Scenario Time	07:01	07:15	07:30	07:45	08:00	08:15	08:30	08:45	
Time since start of run	1	15	30	45	60	75	90	105	
Replication									Overall
1	0.890	0.773	0.409	0.460	0.476	0.776	0.779	0.757	0.468
2	0.890	0.773	0.422	0.467	0.492	0.749	0.768	0.753	0.475
3	0.890	0.773	0.410	0.455	0.458	0.786	0.800	0.775	0.462
4	0.890	0.773	0.410	0.467	0.496	0.776	0.778	0.757	0.475
5	0.890	0.773	0.407	0.476	0.466	0.788	0.799	0.787	0.479
6	0.890	0.773	0.404	0.460	0.487	0.762	0.789	0.767	0.468
7	0.890	0.773	0.401	0.486	0.494	0.572	0.604	0.598	0.485
8	0.890	0.773	0.391	0.439	0.454	0.489	0.559	0.558	0.443
9	0.890	0.773	0.423	0.477	0.464	0.712	0.736	0.743	0.479
10	0.890	0.773	0.396	0.450	0.473	0.729	0.764	0.747	0.458
11	0.890	0.773	0.396	0.456	0.465	0.783	0.815	0.794	0.462
12	0.890	0.773	0.412	0.434	0.447	0.768	0.772	0.757	0.446
13	0.890	0.773	0.403	0.478	0.576	0.728	0.739	0.731	0.491
14	0.890	0.773	0.409	0.437	0.494	0.507	0.825	0.810	0.448
15	0.890	0.773	0.419	0.475	0.472	0.722	0.731	0.712	0.479
16	0.890	0.773	0.428	0.474	0.464	0.531	0.790	0.763	0.475
17	0.890	0.773	0.415	0.480	0.470	0.583	0.604	0.610	0.479
18	0.890	0.773	0.410	0.460	0.480	0.769	0.786	0.777	0.468
19	0.890	0.773	0.423	0.481	0.512	0.499	0.763	0.751	0.484
20	0.890	0.773	0.405	0.477	0.475	0.575	0.588	0.578	0.477
21	0.890	0.773	0.425	0.430	0.446	0.786	0.783	0.754	0.444
22	0.890	0.773	0.405	0.460	0.456	0.537	0.583	0.581	0.461
23	0.890	0.773	0.399	0.466	0.483	0.752	0.761	0.738	0.471
24	0.890	0.773	0.409	0.493	0.513	0.774	0.773	0.762	0.496
25	0.890	0.773	0.411	0.471	0.484	0.759	0.777	0.757	0.477
26	0.890	0.773	0.404	0.473	0.482	0.800	0.798	0.776	0.478
27	0.890	0.773	0.413	0.469	0.480	0.772	0.792	0.769	0.475
28	0.890	0.773	0.421	0.472	0.471	0.539	0.575	0.585	0.473
29	0.890	0.773	0.393	0.468	0.468	0.767	0.782	0.772	0.471
30	0.890	0.773	0.413	0.458	0.484	0.754	0.745	0.723	0.467
31	0.890	0.773	0.406	0.477	0.465	0.563	0.586	0.586	0.475
32	0.890	0.773	0.398	0.461	0.480	0.777	0.800	0.783	0.468
33	0.890	0.773	0.407	0.466	0.466	0.768	0.772	0.760	0.471
34	0.890	0.773	0.421	0.497	0.502	0.734	0.739	0.717	0.498
35	0.890	0.773	0.399	0.447	0.449	0.505	0.578	0.590	0.449
36	0.890	0.773	0.411	0.449	0.457	0.520	0.584	0.571	0.453
37	0.890	0.773	0.423	0.479	0.480	0.765	0.766	0.745	0.484
38	0.890	0.773	0.410	0.486	0.530	0.529	0.592	0.594	0.489
39	0.890	0.773	0.408	0.465	0.489	0.743	0.738	0.708	0.472
40	0.890	0.797	0.407	0.472	0.488	0.568	0.581	0.578	0.474
41	0.890	0.773	0.399	0.464	0.473	0.755	0.756	0.733	0.469
42	0.890	0.773	0.398	0.449	0.446	0.755	0.782	0.763	0.455
43	0.890	0.773	0.411	0.460	0.462	0.772	0.776	0.751	0.466
44	0.890	0.773	0.405	0.460	0.493	0.765	0.769	0.759	0.469
45	0.890	0.773	0.403	0.458	0.472	0.552	0.602	0.603	0.461
46	0.890	0.773	0.413	0.458	0.456	0.609	0.616	0.610	0.461
47	0.890	0.773	0.407	0.470	0.464	0.756	0.769	0.757	0.474
48	0.890	0.773	0.399	0.475	0.502	0.787	0.813	0.796	0.481
49	0.890	0.773	0.401	0.478	0.471	0.760	0.791	0.764	0.479
50	0.890	0.773	0.419	0.479	0.499	0.773	0.797	0.783	0.485
Mean	0.890	0.774	0.408	0.466	0.479	0.692	0.727	0.714	0.471
SD	0.000	0.003	0.009	0.014	0.023	0.107	0.086	0.078	0.013

Table 3-1: Details of Level 1 SA calculated for the replications of the analogue case

Scenario Time	07:01	07:15	07:30	07:45	08:00	08:15	08:30	08:45	
Time since start of run	**1**	**15**	**30**	**45**	**60**	**75**	**90**	**105**	
Replication									**Overall**
1	0.890	0.850	0.688	0.901	0.900	0.871	0.884	0.888	0.877
2	0.890	0.825	0.764	0.880	0.880	0.852	0.872	0.882	0.866
3	0.890	0.825	0.689	0.920	0.925	0.869	0.856	0.886	0.894
4	0.890	0.825	0.692	0.910	0.894	0.849	0.873	0.895	0.883
5	0.890	0.850	0.699	0.912	0.937	0.886	0.882	0.919	0.890
6	0.890	0.850	0.672	0.894	0.909	0.882	0.865	0.899	0.871
7	0.890	0.825	0.689	0.884	0.903	0.767	0.769	0.859	0.862
8	0.890	0.825	0.727	0.923	0.905	0.883	0.874	0.909	0.898
9	0.890	0.850	0.700	0.912	0.894	0.862	0.888	0.916	0.886
10	0.890	0.850	0.671	0.910	0.885	0.875	0.865	0.893	0.880
11	0.890	0.876	0.672	0.917	0.930	0.916	0.924	0.935	0.892
12	0.890	0.875	0.643	0.912	0.922	0.877	0.879	0.888	0.883
13	0.890	0.875	0.703	0.914	0.906	0.863	0.885	0.885	0.889
14	0.890	0.850	0.673	0.902	0.929	0.910	0.907	0.927	0.880
15	0.890	0.850	0.710	0.913	0.908	0.883	0.906	0.925	0.889
16	0.890	0.825	0.692	0.942	0.907	0.875	0.873	0.876	0.909
17	0.890	0.850	0.635	0.893	0.900	0.862	0.855	0.874	0.865
18	0.890	0.825	0.642	0.877	0.906	0.874	0.852	0.894	0.854
19	0.890	0.850	0.641	0.894	0.901	0.867	0.876	0.880	0.866
20	0.890	0.850	0.680	0.894	0.888	0.878	0.896	0.917	0.870
21	0.890	0.850	0.640	0.914	0.910	0.890	0.871	0.905	0.882
22	0.890	0.850	0.662	0.915	0.917	0.889	0.887	0.907	0.887
23	0.890	0.825	0.685	0.886	0.895	0.901	0.906	0.905	0.865
24	0.890	0.875	0.677	0.931	0.899	0.809	0.855	0.855	0.898
25	0.890	0.825	0.641	0.913	0.891	0.867	0.855	0.878	0.879
26	0.890	0.880	0.686	0.926	0.918	0.857	0.856	0.895	0.897
27	0.890	0.825	0.752	0.949	0.922	0.912	0.921	0.928	0.923
28	0.890	0.825	0.643	0.895	0.885	0.856	0.870	0.883	0.865
29	0.890	0.875	0.691	0.889	0.897	0.846	0.832	0.855	0.868
30	0.890	0.825	0.637	0.940	0.882	0.824	0.846	0.878	0.898
31	0.890	0.850	0.577	0.923	0.915	0.897	0.893	0.900	0.883
32	0.890	0.850	0.708	0.900	0.898	0.881	0.882	0.900	0.878
33	0.890	0.825	0.637	0.902	0.900	0.871	0.886	0.898	0.872
34	0.890	0.875	0.639	0.878	0.894	0.859	0.899	0.902	0.854
35	0.890	0.825	0.673	0.937	0.900	0.861	0.887	0.893	0.902
36	0.890	0.850	0.684	0.885	0.902	0.902	0.909	0.928	0.865
37	0.890	0.825	0.703	0.911	0.908	0.869	0.873	0.879	0.887
38	0.890	0.918	0.687	0.880	0.903	0.860	0.897	0.894	0.862
39	0.890	0.850	0.630	0.924	0.910	0.918	0.898	0.910	0.890
40	0.890	0.825	0.710	0.901	0.878	0.837	0.855	0.875	0.876
41	0.890	0.825	0.647	0.907	0.923	0.898	0.893	0.915	0.879
42	0.890	0.842	0.681	0.915	0.925	0.848	0.845	0.881	0.889
43	0.890	0.900	0.633	0.882	0.912	0.885	0.894	0.894	0.859
44	0.890	0.825	0.715	0.918	0.907	0.875	0.874	0.902	0.893
45	0.890	0.850	0.637	0.855	0.884	0.890	0.897	0.904	0.835
46	0.890	0.825	0.672	0.930	0.896	0.855	0.864	0.895	0.895
47	0.890	0.919	0.711	0.854	0.898	0.883	0.903	0.913	0.845
48	0.890	0.850	0.617	0.896	0.895	0.861	0.877	0.905	0.865
49	0.890	0.875	0.705	0.925	0.894	0.785	0.796	0.813	0.895
50	0.213	0.225	0.359	0.588	0.596	0.583	0.594	0.597	0.558
Mean	0.876	0.836	0.668	0.899	0.898	0.863	0.870	0.889	0.873
SD	0.096	0.091	0.057	0.049	0.046	0.050	0.049	0.047	0.048

Table 3-2: Details of Level 1 SA calculated for the Epoch 3 case

Scenario Time	**07:01**	**07:15**	**07:30**	**07:45**	**08:00**	**08:15**	**08:30**	**08:45**	
Time since start of run	**1**	**15**	**30**	**45**	**60**	**75**	**90**	**105**	
Replication									**Overall**
1	0.890	0.777	0.461	0.744	0.843	0.688	0.777	0.769	0.724
2	0.890	0.801	0.597	0.681	0.775	0.766	0.816	0.811	0.686
3	0.890	0.777	0.452	0.798	0.822	0.675	0.732	0.796	0.762
4	0.890	0.777	0.609	0.789	0.840	0.684	0.764	0.784	0.774
5	0.890	0.777	0.609	0.789	0.801	0.726	0.754	0.812	0.770
6	0.890	0.777	0.627	0.673	0.775	0.722	0.740	0.771	0.682
7	0.213	0.213	0.337	0.527	0.527	0.432	0.528	0.572	0.500
8	0.890	0.777	0.480	0.736	0.797	0.664	0.769	0.783	0.715
9	0.890	0.801	0.607	0.802	0.824	0.723	0.761	0.817	0.782
10	0.890	0.777	0.472	0.786	0.858	0.753	0.806	0.845	0.760
11	0.890	0.777	0.653	0.784	0.838	0.747	0.806	0.827	0.775
12	0.890	0.777	0.644	0.711	0.794	0.743	0.783	0.825	0.715
13	0.890	0.777	0.603	0.748	0.808	0.682	0.798	0.795	0.739
14	0.890	0.777	0.586	0.666	0.804	0.744	0.756	0.759	0.676
15	0.890	0.777	0.618	0.736	0.821	0.773	0.830	0.881	0.734
16	0.890	0.777	0.568	0.756	0.813	0.729	0.809	0.865	0.743
17	0.890	0.777	0.615	0.708	0.781	0.707	0.785	0.809	0.708
18	0.890	0.777	0.614	0.786	0.796	0.704	0.770	0.836	0.767
19	0.890	0.777	0.532	0.768	0.847	0.716	0.807	0.852	0.751
20	0.890	0.777	0.527	0.772	0.852	0.710	0.752	0.800	0.754
21	0.890	0.777	0.556	0.707	0.780	0.717	0.781	0.828	0.701
22	0.890	0.777	0.605	0.821	0.823	0.657	0.772	0.782	0.795
23	0.890	0.777	0.475	0.708	0.772	0.704	0.792	0.824	0.692
24	0.890	0.777	0.650	0.815	0.835	0.770	0.814	0.858	0.798
25	0.890	0.777	0.608	0.700	0.753	0.720	0.740	0.727	0.698
26	0.890	0.801	0.592	0.767	0.812	0.711	0.733	0.739	0.753
27	0.890	0.777	0.487	0.808	0.818	0.687	0.792	0.870	0.772
28	0.890	0.777	0.568	0.812	0.847	0.703	0.774	0.792	0.788
29	0.890	0.777	0.561	0.699	0.806	0.825	0.848	0.862	0.700
30	0.890	0.777	0.598	0.695	0.809	0.738	0.779	0.783	0.699
31	0.890	0.777	0.609	0.766	0.805	0.606	0.655	0.720	0.751
32	0.890	0.777	0.632	0.697	0.788	0.751	0.767	0.762	0.703
33	0.890	0.777	0.667	0.728	0.783	0.668	0.757	0.760	0.728
34	0.890	0.777	0.469	0.736	0.797	0.649	0.735	0.741	0.713
35	0.890	0.777	0.615	0.770	0.839	0.705	0.738	0.756	0.760
36	0.890	0.801	0.602	0.777	0.819	0.726	0.777	0.803	0.763
37	0.890	0.777	0.606	0.664	0.768	0.688	0.773	0.769	0.672
38	0.890	0.777	0.668	0.825	0.854	0.749	0.794	0.863	0.809
39	0.890	0.777	0.595	0.714	0.763	0.708	0.729	0.730	0.708
40	0.890	0.777	0.655	0.789	0.846	0.726	0.821	0.852	0.780
41	0.890	0.777	0.579	0.765	0.835	0.717	0.760	0.772	0.752
42	0.890	0.777	0.598	0.734	0.843	0.796	0.846	0.863	0.733
43	0.890	0.777	0.576	0.787	0.818	0.693	0.691	0.714	0.766
44	0.890	0.777	0.478	0.772	0.834	0.724	0.772	0.817	0.746
45	0.890	0.777	0.618	0.704	0.786	0.756	0.799	0.849	0.706
46	0.890	0.777	0.488	0.785	0.823	0.682	0.745	0.805	0.756
47	0.890	0.777	0.547	0.746	0.789	0.717	0.789	0.828	0.730
48	0.890	0.777	0.548	0.729	0.762	0.640	0.747	0.778	0.713
49	0.890	0.777	0.595	0.709	0.804	0.688	0.726	0.749	0.708
50	0.890	0.777	0.526	0.731	0.851	0.722	0.783	0.786	0.723
Mean	0.876	0.768	0.572	0.744	0.806	0.709	0.767	0.796	0.733
SD	0.096	0.080	0.068	0.053	0.049	0.056	0.050	0.054	0.048

Table 3-3: Details of Level 1 SA calculated for the replications of the Epoch 4 case

Replication	ANALOGUE CASE			EPOCH 3 CASE			EPOCH 4 CASE		
	Total Red Losses	Total Blue Losses	Red / Blue LER	Total Red Losses	Total Blue Losses	Red / Blue LER	Total Red Losses	Total Blue Losses	Red / Blue LER
1	57	166	0.34	47	80	0.59	88	151	0.58
2	77	195	0.39	47	76	0.62	103	183	0.56
3	63	163	0.39	48	74	0.65	126	149	0.85
4	74	164	0.45	49	99	0.49	101	164	0.62
5	61	164	0.37	59	78	0.76	135	139	0.97
6	74	158	0.47	41	67	0.61	114	139	0.82
7	53	161	0.33	38	95	0.40	144	154	0.94
8	59	149	0.40	51	94	0.54	90	136	0.66
9	60	167	0.36	45	88	0.51	82	153	0.54
10	82	152	0.54	31	68	0.46	68	145	0.47
11	63	149	0.42	41	93	0.44	109	155	0.70
12	57	158	0.36	44	74	0.59	62	138	0.45
13	88	168	0.52	61	80	0.76	111	135	0.82
14	63	152	0.41	41	65	0.63	93	150	0.62
15	67	144	0.47	62	72	0.86	87	160	0.54
16	93	159	0.58	79	67	1.18	104	125	0.83
17	72	170	0.42	48	87	0.55	117	156	0.75
18	78	164	0.48	30	79	0.38	108	154	0.70
19	89	172	0.52	33	59	0.56	98	160	0.61
20	48	156	0.31	70	69	1.01	112	174	0.64
21	70	144	0.49	39	76	0.51	78	168	0.46
22	50	159	0.31	42	74	0.57	119	135	0.88
23	70	176	0.40	64	73	0.88	141	125	1.13
24	57	148	0.39	37	81	0.46	103	135	0.76
25	67	188	0.36	45	81	0.56	121	160	0.76
26	63	157	0.40	52	67	0.78	124	163	0.76
27	71	182	0.39	66	75	0.88	125	150	0.83
28	86	173	0.50	39	76	0.51	111	135	0.82
29	48	188	0.26	58	85	0.68	110	142	0.77
30	86	169	0.51	63	91	0.69	84	144	0.58
31	80	157	0.51	47	80	0.59	132	137	0.96
32	67	150	0.45	57	65	0.88	109	144	0.76
33	83	142	0.58	39	63	0.62	119	168	0.71
34	66	155	0.43	66	88	0.75	95	161	0.59
35	65	183	0.36	37	90	0.41	119	161	0.74
36	74	164	0.45	50	85	0.59	99	173	0.57
37	62	145	0.43	58	88	0.66	110	151	0.73
38	52	181	0.29	67	95	0.71	104	172	0.60
39	64	168	0.38	71	92	0.77	111	175	0.63
40	72	139	0.52	50	85	0.59	124	138	0.90
41	58	185	0.31	68	79	0.86	143	147	0.97
42	80	161	0.50	32	74	0.43	91	149	0.61
43	72	177	0.41	43	84	0.51	112	171	0.65
44	58	161	0.36	44	63	0.70	96	140	0.69
45	53	162	0.33	76	112	0.68	143	119	1.20
46	53	153	0.35	41	74	0.55	128	146	0.88
47	63	152	0.41	66	86	0.77	123	143	0.86
48	71	179	0.40	74	90	0.82	122	127	0.96
49	64	173	0.37	53	56	0.95	127	125	1.02
50	92	151	0.61	52	73	0.71	124	148	0.84
Mean	67.90	163.06	0.42	51.22	79.30	0.65	109.98	149.44	0.75
SD	11.77	13.28	0.08	12.75	11.29	0.17	18.84	14.91	0.17

Table 3-4: Details of Red and Blue Losses for the analogue, Epoch 3 and Epoch 4 cases