

Simulation Based Knowledge Elicitation: Effect of Visual Representation and Model Parameters

Stewart Robinson ^{*1}

Operational Research and Management Sciences Group, Warwick Business School, University
of Warwick, Coventry, CV4 7AL, UNITED KINGDOM
stewart.robinson@warwick.ac.uk, Tel: 02476 522132, Fax: 02476 524539

*Ernie P.K. Lee*²

Operational Research and Management Sciences Group, Warwick Business School, University
of Warwick, Coventry, CV4 7AL, UNITED KINGDOM

John S. Edwards

Operations and Information Management Group, Aston Business School, Aston University,
Birmingham, B4 7ET, UNITED KINGDOM

* Corresponding author

¹ Present address: School of Business and Economics, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UNITED KINGDOM

² Transport Planning, London Underground Limited, London, SW1H 0BD, UNITED KINGDOM

Abstract

Since much knowledge is tacit, eliciting knowledge is a common bottleneck during the development of knowledge-based systems. Visual interactive simulation (VIS) has been proposed as a means for eliciting experts' decision-making by getting them to interact with a visual simulation of the real system in which they work. In order to explore the effectiveness and efficiency of VIS based knowledge elicitation, an experiment has been carried out with decision-makers in a Ford Motor Company engine assembly plant. The model properties under investigation were the level of visual representation (2-dimensional, 2½-dimensional and 3-dimensional) and the model parameter settings (unadjusted and adjusted to represent more uncommon and extreme situations). The conclusion from the experiment is that using a 2-dimensional representation with adjusted parameter settings provides the better simulation-based means for eliciting knowledge, at least for the case modeled.

Abbreviations

AI – Artificial Intelligence, ATD – After Test Dress, KBS- Knowledge Based System, UV – Ultra-Violet, VIS – Visual Interactive Simulation, VR – Virtual Reality

Key Words

Automobile industry, discrete-event simulation, knowledge-based systems, knowledge elicitation, visual fidelity, visual interactive simulation

Simulation Based Knowledge Elicitation: Effect of Visual Representation and Model Parameters

1. Introduction

Feigenbaum (1980) makes an empirical observation that a knowledge-based system (KBS) derives its power from the knowledge it possesses, not from the particular formalisms and inference schemes it employs. The formalisms and inference schemes only provide the mechanisms to use the power. In other words, an expert's knowledge is both necessary and nearly sufficient for developing a KBS. In the context of using automatic knowledge elicitation techniques (rule induction and pattern matching) to build a KBS, this empirical observation stresses the importance of collecting complete and accurate example cases to develop a powerful knowledge base. However, research on collecting such an informative set of example cases appears to be limited. Liang et al. (1992) explain that this could be attributed to the assumption that the training data are readily available. Unfortunately, this is often not true as some expert knowledge is difficult to obtain, particularly since it needs to be collected on a real-time basis. Deslandres and Pierreval (1997) discuss various approaches to eliciting expert knowledge for use in improving manufacturing quality control, but without considering the experts' time, availability or accuracy.

Where the process of obtaining training data has been investigated it is clear that it is a laborious and time-consuming task. Ong et al. (2006) analyse documents, and perform interviews and protocol analysis to elicit knowledge in an automotive manufacturing environment. They conclude that the 'data collection and analysis was extremely time consuming.' Similarly, Roy et al. (2004) describe how time and resources limited the extent of their knowledge elicitation efforts. Garthwaite et al. (2008) elicit expert knowledge about treatment pathways for bowel cancer. Although only four experts were involved, and this was only part of the process, its scale required that 'This elicitation process was supported by a team of statisticians' (Trueman et al., 2007).

One possible approach, which may help to alleviate these issues, is the use of simulation as a knowledge elicitation tool; and in particular, visual interactive simulation (VIS). VIS (Hurrion, 1986) is a widely-used variant of discrete-event simulation (DES) (Law, 2007), which is recognized as the most popular approach to dynamic simulation modeling in the field of operations research. In a VIS, a DES model drives a *visual* display that represents the dynamic workings of the simulation. In addition, the user is able to *interact* with the model in order to view statistics and to carry out different experiments to better understand and/or improve the real system that is being simulated. As such, VIS provides an environment in which experts could interact with a visual simulation of their system for the purpose of collecting data on their decision-making process.

In this paper we explore the efficacy of VIS in terms of its effectiveness and efficiency as a knowledge elicitation approach. First we review previous work that employs VIS for eliciting knowledge. In section 3 we outline the research undertaken. The work is based on a case study performed with Ford Europe. The background to the case is described in section 4 along with the VIS and the knowledge elicitation sessions. In section 5 the results are presented with respect to six hypotheses, and the findings are discussed in section 6. Lastly, the paper concludes with a brief summary of the work undertaken and a discussion on the potential for future research.

2. Visual Interactive Simulation (VIS) for Knowledge Elicitation

There has always been considerable interest given to artificial intelligence (AI) working with(in) simulation, and *vice versa*. To date, several taxonomies depicting such interest have been published, two of which are O’Keefe (1986) and Ören (1994). O’Keefe (1986) develops a taxonomy for combining KBS and simulation, listing seven possible combinations. Meanwhile, Ören (1994) lists two types of activities that use simulation in AI: use of simulation for applications of AI, such as, evaluating a KBS (Flitman and Hurrion, 1987; Shaw, 1989; Chryssolouris et al., 1991; Liang et al., 1992); and cognitive simulation, where systems with cognitive abilities are simulated. Such systems include humans and autonomous robots. Ören also lists two types of activities that use AI in simulation: AI-assisted simulation and AI-based

simulation. On the one hand, in AI-assisted simulation, AI techniques are used to provide computer assistance in areas such as formulating models and designing simulation experiments. For example, Paris et al. (2001) use evolutionary algorithms to develop possible designs for a flexible manufacturing system, while Chan and Chan (2005) use a rule-based system to design possible printed circuit board manufacturing systems. In each case the designs are then tested using a simulation. On the other hand, in AI-based simulation, AI techniques are used to generate model behaviour in simulation runs (Flitman and Hurrion, 1987; O'Keefe, 1989; Williams, 1996; Lyu and Gunasekaran, 1997; Robinson et al., 1998; Kunnathur et al., 2004). For instance, Moreno et al. (2001) use a simulation model to test a hospital management knowledge based system. Similarly, Yim et al. (2004) represent knowledge in a system dynamics model and simulate the impact of alternative decision-making strategies with the aim of improving strategic decision-making in a US-based telecommunications company.

The evidence to support the use of simulation for knowledge elicitation purposes stems mainly from the activities identified by Ören. Anecdotal evidence of using simulation to elicit expert knowledge is found in Flitman and Hurrion (1987), Shaw (1989), Pierreval and Ralambondrainy (1990), Chryssolouris et al. (1991), Hurrion (1991), Liang et al. (1992), Tan et al. (2000), Tan (2003) and Robinson et al. (2005). These are a few examples of what Ören (1994) classifies as using simulation for applications of AI.

In Flitman and Hurrion (1987) and Hurrion (1991) the authors build a VIS model for the operations of a simple coal-yard depot. Embracing a gaming mode, the user takes on the role of a depot manager and controls the depot's operations in the model. In the meantime, a KBS linked to the VIS model monitors and records all user actions. The data obtained are then used for machine learning. Also, Liang et al. (1992) employ VIS in a gaming mode to collect real-time scheduling decisions. These decisions are then used to facilitate learning in an automated knowledge acquisition process which integrates semi-Markov processes with neural network computing.

Shaw (1989) describes and demonstrates a 'learning by experimentation' methodology. A flexible manufacturing system is simulated and alternative scenarios employing different

scheduling rules are tested for each selected hypothetical state of the system. The scheduling rule that produces the best performance for a state becomes the rule to be deployed whenever the system assumes that state. As such, a collection of state-rule pairs can be generated as a training set for learning scheduling knowledge. A similar methodology is also applied by Pierreval and Ralambondrainy (1990) on a simplified flow shop example.

Chryssolouris et al. (1991) begin the learning process for building a neural network by running several simulations of a job shop. In these simulations, the operational policy (weights of the decision-making criteria) and the workload (mix and volume of job types) are varied, and performance measures are collected at the end of each run. The performance measures plus workload parameter values constitute the input component of an input-output pair, whilst the policy parameter values constitute the output component. In this way, a set of input-output pairs can be collected from these simulations to learn the knowledge of selecting an operational policy.

Meanwhile, Tan et al. (2000) planned to investigate the validity and reliability of using an interactive, simulation-driven immersive virtual reality (VR) system for collecting data, which would then be used for learning human behavioural rules. Subsequently, Tan (2003) performed method-comparison studies between the data collected using the VR system with the data from direct observation, and concludes that there is some evidence supporting VR as a suitable technology for collecting data. However, the original plan to learn human behavioural rules was not carried out.

Robinson et al. (2005) develop the knowledge-based improvement methodology which uses VIS to elicit knowledge from decision-makers; in this case maintenance supervisors in a manufacturing plant. The data collected are then used to train an AI model which is subsequently linked to the VIS model to represent the human decision-makers. The linked models are used to assess the long-run performance of the decision-makers and to look for ways for improving their decision-making.

The above examples illustrate that simulation/VIS has been used to elicit episodic knowledge in the form of example cases from experts. They also show that the collected example cases have been used successfully for machine learning. Therefore, using simulation/VIS for knowledge elicitation is a tried and tested concept. It is not, however, without its limitations. Robinson et al. (2005) identify three specific problems with using VIS for knowledge elicitation:

- i. The VIS model needs to contain and report all the key attributes in the decision-making process, which could require a very detailed model that would be time consuming to develop. Obtaining the data required for such a model may also be problematic.
- ii. The human decision-makers may need to spend a long time in front of the model in order to collect sufficient example decisions to enable training of a robust KBS.
- iii. The human decision-makers may not take realistic decisions in a simulated environment given that there are no real consequences to their decisions.

The research reported in this paper focuses on the latter two issues. Our prime interest is in the impact of the level of visual representation and the nature of the simulated scenarios, as defined by the settings of the parameters for the simulation runs.

3. Outline of the Research Undertaken

The purpose of this research is to understand the effects of visual representation and model parameters on the efficiency and effectiveness of VIS for knowledge elicitation. To this end we consider three constructs for measuring elicitation effectiveness and one for elicitation efficiency as follows:

Elicitation Effectiveness

- *Decision Fidelity*: whether the decisions made in a VIS model bear close resemblance to the decisions made in the real world. This is clearly important if a valid dataset is to be generated from the VIS.

- *State Space*: whether the decision attributes cover a wide range of values for all attributes. This is important because ‘a training set (of example cases) must cover the full range of values for all inputs’ (Negnevitsky, 2005).
- *Case Quantity*: the total number of example cases recorded in a knowledge elicitation session. The importance of this is again described by Negnevitsky (2005): ‘the training set (of example cases) has to be sufficiently large.’

Elicitation Efficiency

- *Collection Rate*: the number of example cases recorded per unit of real-time in a knowledge elicitation session. Since VIS based knowledge elicitation can become time consuming and laborious, it is important to elicit knowledge as quickly as possible.

3.1 Proposed Effect of Visual Representation on Elicitation Effectiveness and Efficiency

Although it seems ideal to present a visual display with a high degree of realism in a VIS based knowledge elicitation exercise, the cost of generating such a display could be prohibitive. Such a degree of realism can be unnecessary given the needs of the application (Preece, 1994). For instance, in using a flight simulator it is less important to deceive the pilots into believing they are flying a real aircraft than to provide them with the necessary information, in the right format, for them to function as if they are flying an aeroplane. Under the visual representation dimension we investigate the effect of different levels of visual display on the effectiveness and efficiency of a VIS based knowledge elicitation exercise.

Based on the representations available in commercial simulation software packages, three levels of visual display are used: 2-dimensional (2D), 2½-dimensional (2½D) and 3-dimensional (3D). The 2D representation is the standard iconic display provided in VIS software since the late 1980s. A 2½D representation consists of 3D icons displayed on a 2D background with no perspective projection or photo-realism. It provides a higher degree of realism than the plain 2D representation, but at very little additional effort or cost. A 3D representation involves 3D photo-realistic icons displayed against a 3D photo-realistic background with perspective projection. In effect this is a non-immersive VR system. Such displays can require a great deal of additional effort and cost to produce.

There is a dearth of empirical studies that compare these visual representations. Dransch et al. (2010) use a wide range of visual displays to help experts judge the quality of environmental simulation models, but alternative dimensionality of a given representation was not one of the aspects they evaluated. In each case they reasoned what they thought would be the appropriate dimensionality for particular attributes, such as a 2D map replicating what the experts would see in a real situation, and used only that one. There are, however, some opinions expressed on the benefits, or otherwise, of 3D over 2D displays (Waller and Ladbrook, 2002). Barnes (1996) argues that a user develops a strong sense of involvement in a 3D world and that interactions with the model are likely to closely parallel those in the real world. As such, we would expect that *a higher dimension of visual representation would improve the fidelity of the example cases collected in a knowledge elicitation session* (decision fidelity construct).

Barnes (1996, 1997) also mentions that a 3D representation allows for better communication and visualisation of ideas. In a survey of simulation users, Akpan and Brooks (2005a) found that the majority of respondents agreed that a higher dimension of visual representation improved communication and understanding of the system. They also conclude that it is easier to uncover inaccuracies in the model. Based on these opinions, we would expect experts to more easily identify occasions during a knowledge elicitation session in which they need to intervene and interact with the model if a higher dimension of visual representation is used. This would lead to *a larger number of example cases being collected in a knowledge elicitation session* (case quantity construct).

Akpan and Brooks (2005a) note that a 3D model runs more slowly than a 2D model. In fact, the 3D model's run-speed is found to vary inversely with the level of photo-realism and resolution of the graphics (Rehn et al., 2004). As such, even if a larger quantity of cases is collected in a knowledge elicitation session, this may just be a function of the slower run-speed giving the expert more time to process the information from the simulation run. We expect, therefore, *the dimension of visual representation to affect the efficiency with which example cases are collected in a knowledge elicitation session* (collection rate construct). Because the benefits of a higher fidelity of display may be offset by a reduced run-speed, the direction of this effect is uncertain.

3.2 Proposed Effect of Model Parameters on Elicitation Effectiveness and Efficiency

Two levels of parameter settings are considered: unadjusted and adjusted. Unadjusted settings aim to model the system in its typical working state, while adjusted settings aim to develop more uncommon and extreme scenes during a model run. The aim of adjusting the parameter settings is *to increase the state space of the example cases collected in a knowledge elicitation session* (state space construct).

The benefit of presenting the expert with more uncommon and extreme scenes is that there is less probability of repeating the same scene many times, leading to expert boredom and fatigue (Robinson et al., 2005). The expert is also more likely to find more occasions on which to intervene and interact with the model. Hence, we would expect *adjusted parameters to increase the quantity of example cases collected in a knowledge elicitation session* (case quantity construct).

Lastly, as adjusted parameter settings are likely to lead to more unconventional and difficult decision situations, the expert may take longer to make a decision when interacting with the model. Therefore, we expect *different parameter settings to impact the efficiency with which the example cases are collected in a knowledge elicitation session* (collection rate construct).

However, because the effect of the additional time for decision-making might be under or over-compensated by the increase in the quantity of examples collected, the direction of the effect on elicitation efficiency is uncertain.

3.3 Research Hypotheses and Research Approach

The propositions specified in italics above lead to six research hypotheses which are listed in table 1. These hypotheses are tested via a case study in which a VIS model of a Ford engine test facility is used to elicit knowledge from the plant's decision-makers. These data are then analysed to determine whether they support, or otherwise, hypotheses 1 to 6. The case study,

VIS model and knowledge elicitation sessions are described in the next section. The analysis of the results and the findings are then discussed in the subsequent two sections.

Table 1 *Research Hypotheses in Relation to Effectiveness and Efficiency Constructs and Level of Visual Display and Parameter Settings*

Construct	Factor	Hypothesis (0 – null, a – alternative)
Decision fidelity	Visual representation	$H_{1(0)}$: The degree of decision fidelity in the example cases collected in a knowledge elicitation session is not affected by the visual representation dimension used.
		$H_{1(a)}$: The degree of decision fidelity in the example cases collected in a knowledge elicitation session improves as a higher dimension of visual representation is used.
State space	Model parameters	$H_{2(0)}$: The size of state space occupied by the example cases collected in a knowledge elicitation session is not affected by the model parameters used.
		$H_{2(a)}$: The size of state space occupied by the example cases collected in a knowledge elicitation session increases as model parameters are adjusted to develop more uncommon and extreme scenes.
Case quantity	Visual representation	$H_{3(0)}$: The size of case quantity of a knowledge elicitation session is not affected by the visual representation dimension used.
		$H_{3(a)}$: The size of case quantity of a knowledge elicitation session increases as a higher dimension of visual representation is used.
	Model parameters	$H_{4(0)}$: The size of case quantity of a knowledge elicitation session is not affected by the model parameters used.
		$H_{4(a)}$: The size of case quantity of a knowledge elicitation session increases as model parameters are adjusted to develop more uncommon and extreme scenes.
Collection rate	Visual representation	$H_{5(0)}$: The collection rate in a knowledge elicitation session is not affected by the visual representation dimension used.
		$H_{5(a)}$: The collection rate in a knowledge elicitation session is affected by the visual representation dimension used.
	Model parameters	$H_{6(0)}$: The collection rate in a knowledge elicitation session is not affected by the model parameters used.
		$H_{6(a)}$: The collection rate in a knowledge elicitation session is affected by the model parameters used.

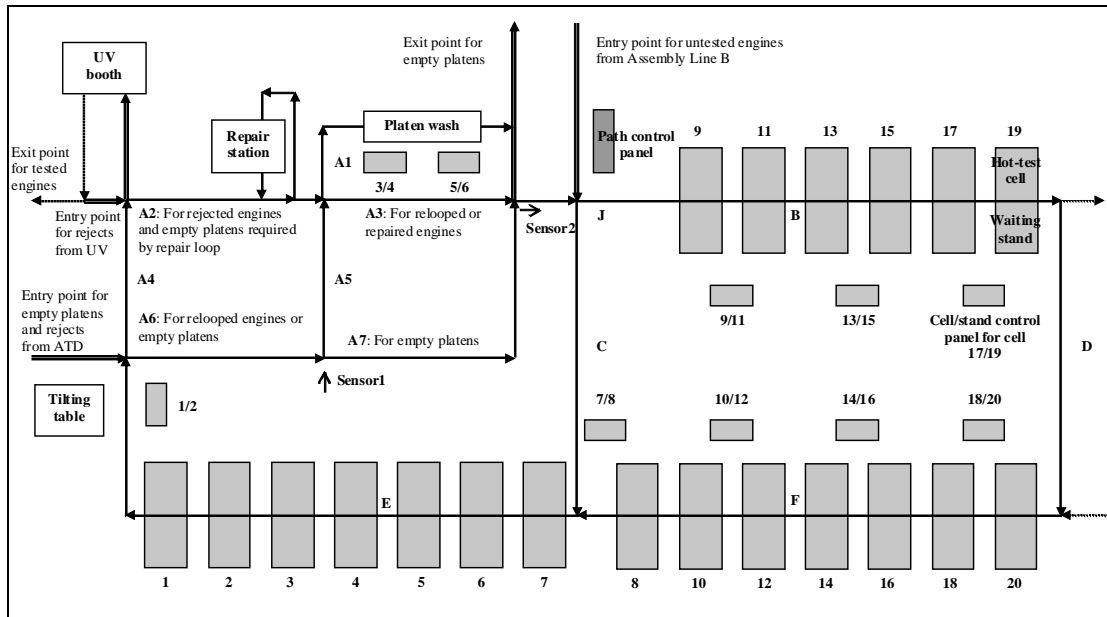
4. Knowledge Elicitation Case Study: Switch Operations in the Ford Puma Hot Test Facility

The subject of the case study is the hot test facility for the Ford Puma engine assembly plant in Dagenham, East London. In this facility assembled engines are tested by rigging them to a testing machine and running the engine for a period to monitor their performance. Engines that fail the test are sent to a repair station before retesting, and engines that pass the test are shipped to one of the automotive assembly plants. The subjects of this research are the team of experts ('switch operators') that monitor and manage the stream of engines passing through the hot test facility.

4.1 The Puma Hot Test Facility

Figure 1 shows a schematic of the hot test operation. Assembled engines arrive in the top centre of the schematic from the main assembly line (assembly line B). Two types of diesel engine are assembled in the plant: a 2 litre and a 2.4 litre. The engines are placed on a 'platen' (metallic pallet) and directed to one of 20 hot test cells. If the allocated hot test cell is currently in use, then the engine will wait on a stand opposite the cell. The engines move around the facility on roller conveyors, denoted by the arrowed lines. At junction J an engine can either be sent down conveyor B or conveyor C. By default the engine will then enter the first available hot test cell or waiting stand along the conveyor.

Figure 1 A Schematic of the Puma Hot Test Facility



Following the hot test, defective engines are sent to a repair station for rectification via conveyors A4 and A2. Engines that pass the test are sent to the after test dress (ATD) following a final inspection for leakages in the ultra-violet (UV) booth. Defective engines that have been repaired are sent along conveyor A3 to junction J for reassignment to a hot test cell.

The experts in this case study are the switch operators who are the key decision-makers. Their main objective is to maximize the number of engines tested and sent to ATD by maintaining an efficient and smooth flow of work. They also need to distribute the workload equitably among the hot test operators who are each responsible for rigging and de-rigging the engines in a pair of adjacent hot test cells. To aid them, the switch operators use 21 switches comprising of a path control switch (on the path control panel) and 20 cell/stand control switches (on the cell/stand control panels). The path control switch allows the expert to send engines along conveyor B or conveyor C. They can also opt for automated mode in which three engines are sent down conveyor B followed by one down conveyor C. The cell/stand control allows the switch operators to switch a hot test cell and its waiting stand on or off.

4.2 The Experts' Decision-Making Process

Table 2 summarizes the decisions (decision variables) that the switch operators can make and the decision attributes used in making those decisions. The latter were obtained through inspection of log sheets from the plant, informal interviews and observation interviews. It is necessary to identify both the types of decisions that can be taken and the decision attributes at this stage as these options and data (respectively) need to be represented in the simulation to be used for knowledge elicitation.

Table 2 *Switch Operators' Decision Variables and Decision Attributes for Engine Assignment*

Decision Variables	Decision Attributes
1. Set path option at junction J to conveyor B or C, or automatic	1. Quantity of engines on each section of the conveyor
2. Switch hot test cell/waiting stand on or off	2. Type of engine on each section of the conveyor (2l or 2.4l)
	3. Type of engine currently being/last tested in each hot test cell
	4. Type of engine parked on each waiting stand
	5. Operational status of each hot test cell (free, busy, broken)
	6. Quantity of engines tested by each hot test cell operator

The switch operator takes account of the quantity of engines on each section of the conveyor (attribute 1) in order to avoid bottlenecks and smooth the flow of work across the test cells. Similarly, attribute 6 is taken into account so as to even the work across the cell operators. The importance of engine type (attributes 2, 3 and 4) in making a decision is due to the time required to changeover the rigging in a test cell when changing the engine type. Minimising changeovers improves both throughput and reduces the workload on the hot test operators. The switch operator obviously considers the operational status of a test cell (attribute 5) when allocating work. Given the number of hot test cells and conveyors in the system, there is a total of 551 decision attributes.

When observing the switch operators it became clear that only they initiate interventions. The system itself is inactive in this respect, that is, it does not specifically request decisions from the switch operators. The experts initiate interventions on the basis of smoothing workflow and preventing bottlenecks. As a result, any simulation model for knowledge elicitation in this context has to be capable of being constantly monitored by the expert who can decide at any point to interrupt the run and change the settings of the system.

4.3 Profile of Experts

There are ten hot test personnel who are qualified to work as switch operators. This number is required to provide coverage across all three shifts. Eight of the switch operators were willing, and given management consent, to participate in the research. They are referred to as subjects A to H in order to maintain their anonymity.

All eight experts are in the age range 40 to 59 and with the exception of subjects B and G, they have between three and eight years experience in performing switch operations. Subject B has quite limited specific experience in switch operations, but as a team leader for the entire hot test operation, who stands in as a switch operator occasionally, he was deemed a suitable participant. Meanwhile, subject G was in his first year of working as a switch operator at the time of the knowledge elicitation exercise.

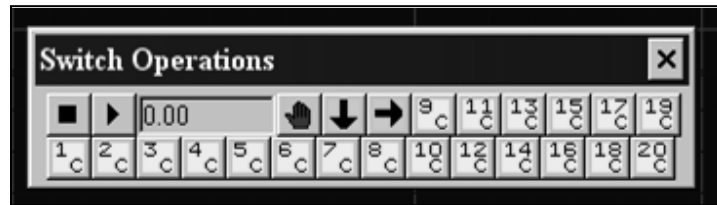
4.4 The Visual Interactive Simulation Model

The hot test model was adapted from an existing model of the complete Puma engine assembly and test facility written in the Witness simulation software (www.lanner.com accessed February 2011). The model of the assembly process was removed to improve run-speed. The arrival of engines from assembly was modeled by sampling from a frequency distribution generated by running the full model for a long period.

Figure 2 shows the control bar that was used by the experts during the knowledge elicitation sessions. The two buttons in the top left corner are to stop the model running (square) and to

continue the run (triangle). For the expert to make an intervention using the other buttons it is necessary to stop the model first. The numbers 0.00 represent the current time in the simulation run, in minutes.

Figure 2 *The Control Bar used to Facilitate Experts' Interventions and Interactions with the Model*



The remaining buttons allow the expert to input decisions as per table 2. The two arrow buttons allow the expert to route engines at junction J down along conveyor C or across to conveyor B. If both buttons are depressed then the allocation moves to automatic mode, that is, three engines to conveyor B followed by one to conveyor C. The twenty buttons labelled 1C to 20C represent the twenty hot test cells. These are used for switching the cells, and their corresponding waiting stands, on and off.

4.4.1 Visual Representations

In designing the display of the simulation special care was taken to ensure that all the decision attributes (table 2) were represented to the model user. Figures 3 to 5 show examples of the three levels of display (2D, 2½D and 3D respectively). Although the displays provide quite different representations, the information provided is exactly the same. For instance, the lights showing the operational status of a cell in the 3D version (figure 5) are represented by the colour of the cell icon in the 2D and 2½D versions.

Figure 3 Example of 2D Display of the Hot Test Model

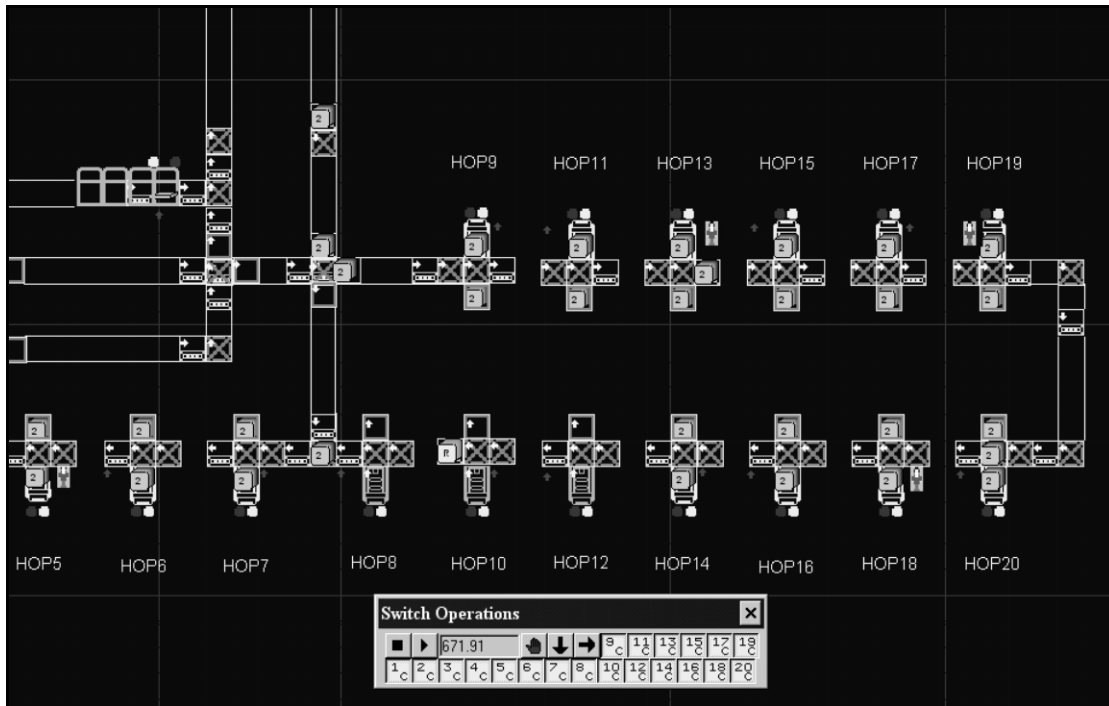


Figure 4 Example of 2½D Display of the Hot Test Model

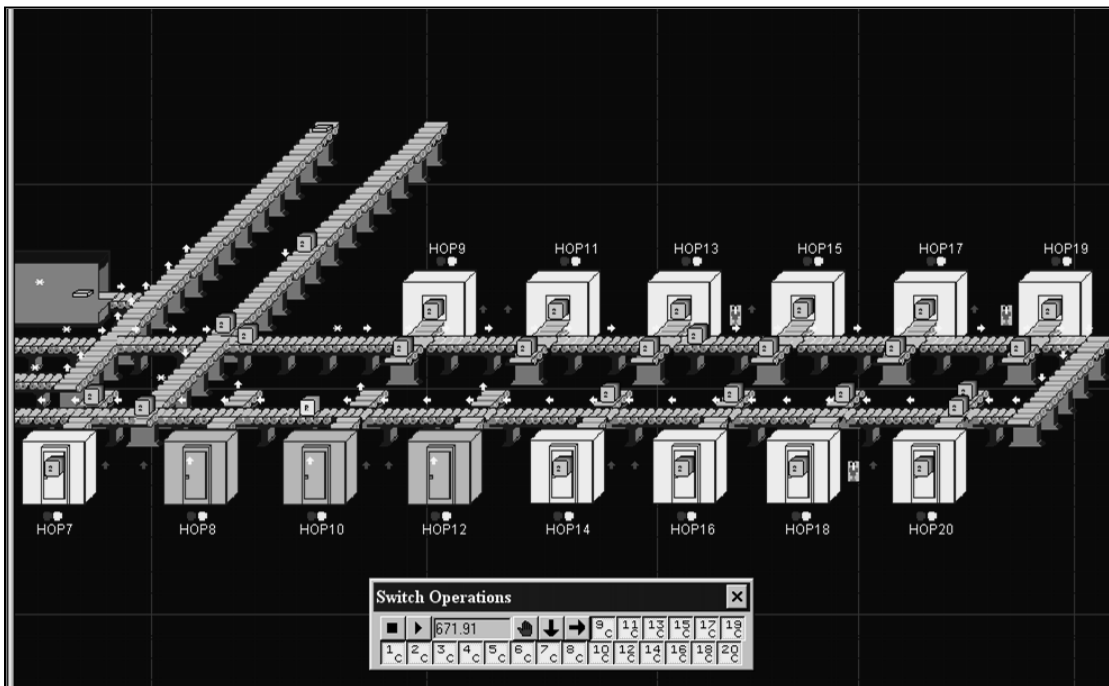
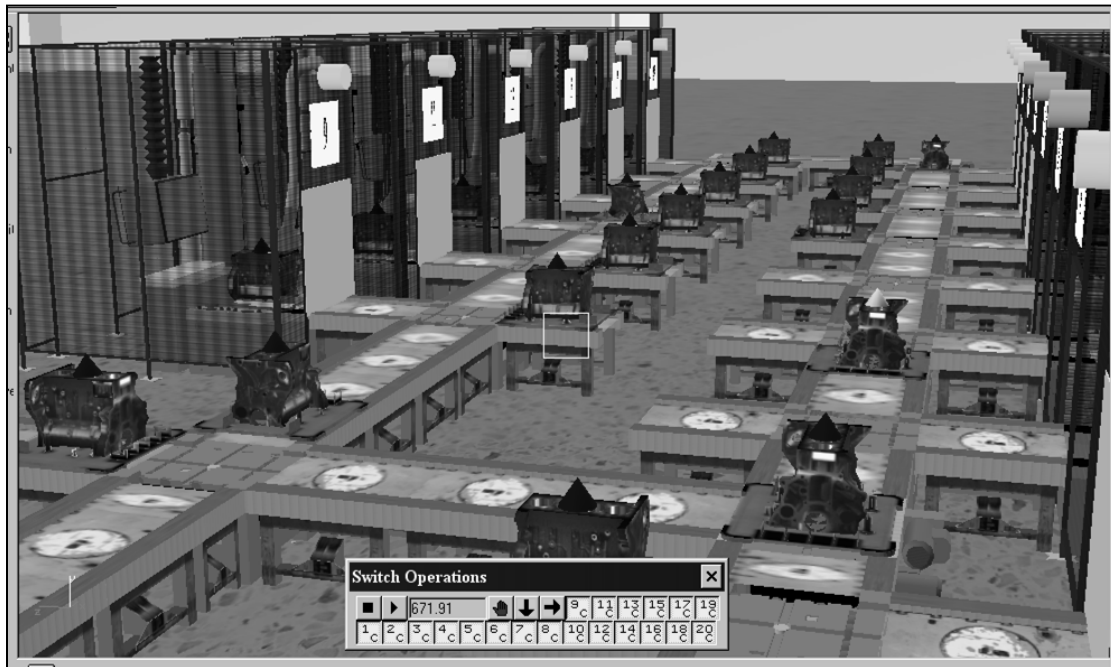


Figure 5 *Example of 3D Display of the Hot Test Model*



4.4.2 Model Parameters

The unadjusted parameter settings used data from the real plant to mimic normal working conditions in the hot test facility. Due to random variations and disturbances in these data, this led to a range of scenes as the model ran, ranging from standard (e.g. running relatively empty) to more extreme (e.g. full of untested engines).

The adjusted parameter settings aimed to create more extreme scenes during a model run. This was achieved by:

- Having a more even inter-arrival time for untested engines
- Running smaller batches of untested engines into the hot test area
- Having a higher proportion of defective engines
- Having a higher incidence of breakdowns in the hot test cells

As a result, the model user was faced with a series of more chaotic scenes as the model runs.

4.5 Knowledge Elicitation Sessions

Knowledge elicitation sessions were carried out with the eight experts over a period of 19 weeks. Given three levels of visual display, two levels of parameter settings and that each expert worked with every model in a repeated measures experimental design (Field and Hole, 2006), this entailed 48 sessions. The long elapsed time was a result of limited access to the switch operators who were required to continue their day-to-day duties in the plant.

A knowledge elicitation session consisted of a switch operator making decisions in the hot test model for a simulated shift. In all cases the experts first worked with the 2D model, then the 2½D model and finally the 3D model. For each visual representation they first worked with the model with unadjusted parameters and then with adjusted parameters. This meant that all the switch operators worked through the six model types in a fixed sequence. At least two days were given between sessions with the same expert to avoid fatigue and reduce potential learning effects being carried over from one session to another. At the beginning of the session, the switch operator was given brief training in the use of the model and the control bar (figure 2).

The model was run for one shift, without intervention, prior to the start of the knowledge elicitation session. This ensured that the model was in a realistic state at the start of the session with the switch operator. The same random number streams were used for each run of the model, although the interventions made by the switch operators meant that the model trajectories differed somewhat based on the decisions made. The speed of the model was controlled to ensure a reasonable compromise between faster than real time interaction and the ability of the decision-maker to identify circumstances that required intervention. However, the level of graphics in the 3D model slowed the simulation such that there was no need to slow the model further.

The knowledge elicitation sessions took between 20 minutes and 96 minutes. Average times for the 2D, 2½D and 3D models with unadjusted parameters were 39 minutes, 31 minutes and 75 minutes respectively. With the adjusted parameters the average session times were 37 minutes,

30 minutes and 89 minutes respectively. The shorter sessions for 2½D were no doubt a result of greater familiarity with running the model. The longer sessions for the 3D model were caused by the slower run-speed of the model.

On every occasion that the switch operator interrupted the model and made a decision to change the settings of the model, data were written out to a file. The file contained information on the status of the model (the decision attributes) and the decisions taken (decision variables). These data were then used as the basis for the analysis of the knowledge elicitation sessions.

5. Results

Results and analyses pertaining to each of the hypotheses are now discussed.

5.1 Analysis of Hypothesis 1: Decision Fidelity and Visual Representation

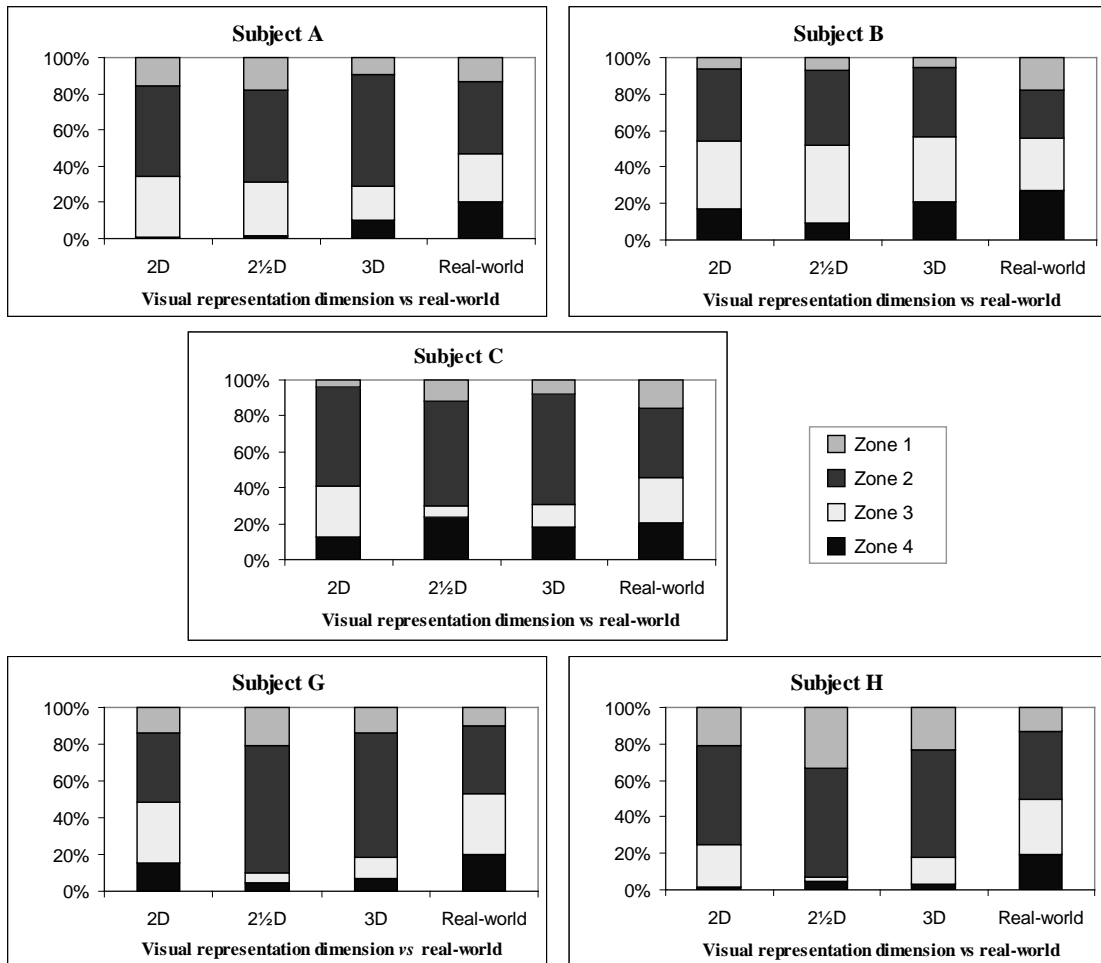
Hypothesis 1 states that decision fidelity is not affected by the level of visual representation. The alternative hypothesis is that decision fidelity improves with higher dimensions of visual representation. Decision fidelity is the resemblance of a decision in the model to the decision in the real system given the same decision attributes. However, since there is no recording of decisions or decision attributes in the real hot test system, it is not possible to directly test this hypothesis. Instead, and as an indicator of fidelity, the facility was split into four zones, the switch operators were shadowed for a period and the number of decisions made in each zone was counted. These counts were then compared to the corresponding count of decisions in each zone when working with the model.

The four zones were:

- Zone 1: junction J
- Zone 2: hot test cells 9-19 (odd numbers)
- Zone 3: hot test cells 8-20 (even numbers)
- Zone 4: hot test cells 1-7

The latter three correspond to conveyors B, F and E respectively in figure 1.

Figure 6 *A Comparison of the Percentage of Decisions for Subjects A, B, C, G and H in each Zone*



Five of the switch operators were shadowed for between 260 minutes and 1,341 minutes. Figure 6 shows the proportion of decisions made in each zone for the three visual representations (running the model with unadjusted parameters) and the proportion in each zone in the real system. Initial inspection suggests some level of correspondence between the proportion of decisions made in each zone for each expert. Certainly zone 2 is the zone in which most

decisions are consistently made. This is unsurprising as this represents the most highly utilized part of the system.

Chi-squared goodness-of-fit tests are performed to compare the number of decisions in each zone between the model representations and the real system. The results in table 3 show that there are significant differences in every case but for switch operator G working with the 2D representation. This is the only case where we can say that the proportion of decisions made in the model might be similar to those in the real system. This suggests that the VIS used has limited ability in obtaining realistic decisions from the switch operators, at least in terms of the zones in which the decisions are made.

Table 3 Summary of χ^2 Test Statistics: Comparison between Model and Real System Decisions

Subject	VIS model with unadjusted model parameters		
	2D	2½D	3D
A	24.82	16.01	36.93
B	82.33	99.28	95.76
C	30.28	20.71	45.30
G	4.31*	63.09	81.69
H	15.43	33.64	29.09
Average	38.22	46.55	57.75

* Statistically insignificant at 5%

The average χ^2 test statistic values for each column suggest that the 2D representation has the highest fidelity, followed by the 2½D and then the 3D. This is in direct contradiction to the alternative hypothesis that decision fidelity improves with higher dimensions of visual representation.

5.2 Analysis of Hypothesis 2: State Space and Model Parameters

Hypothesis 2 states that state space is not affected by the model parameters used. The alternative hypothesis is that the size of state space will increase as model parameters producing more uncommon and extreme scenes are used. To test this hypothesis data on the decision attributes

were analysed. These data show the state of the model at the point at which a decision was entered into the model.

Prior to the analysis the data were cleaned, recoded and rescaled. Cleaning entailed removing redundant attributes; those whose highest and lowest values are zero. Binary variates which were components of various attributes were then organized and recoded to convert them into nominal composite variates. The effect of cleaning and recoding was to reduce the total variates from 551 to 184. Following this, the data were rescaled to ensure all values were on a scale of 0-1. Finally, Andrews plots were used to identify and remove a handful of outliers (Krzanowski, 2005).

The state space is measured as follows:

$$s_d^* = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n \sum_k d_{ijk}^{*2}}{2n^2}} \quad (1)$$

The standard distance, s_d^* , is a dispersion measure for n cases of mixed multivariate data, where d_{ijk}^* represents the Euclidean distance component between two cases i and j for the k^{th} variate.

The computed standard distances are given in table 4.

Table 4 *Computed Standard Distances by Subject and Model Type*

Subject	2D		2½D		3D	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
A	4.908	5.800	4.804	5.788	4.968	5.943
B	4.746	5.747	4.788	5.771	4.736	5.713
C	4.909	5.881	4.817	5.782	4.908	5.758
D	4.821	5.815	4.959	5.853	4.928	5.916
E	4.882	5.828	4.911	5.881	4.911	5.785
F	4.904	5.948	4.859	6.014	5.053	5.877
G	4.856	5.821	4.997	5.720	5.133	5.883
H	4.973	5.846	4.923	5.866	5.056	5.891
Average	4.875	5.836	4.882	5.834	4.962	5.846

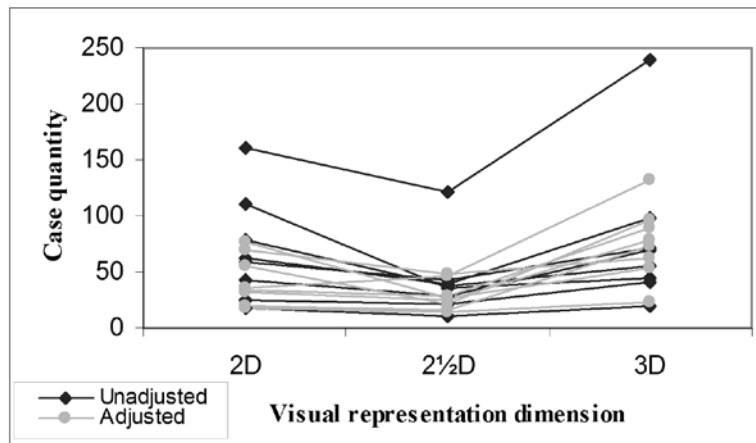
Inspection of the data in the table suggests a clear effect on state space from using adjusted parameters. Hypothesis testing with a two-way repeated measures ANOVA shows firstly that there is no significant effect of visual representation on state space. There is, however, a significant effect from the type of model parameters used. Hypothesis 2 is therefore rejected at a 5% level of significance. Hence, it can be concluded that the size of state space occupied by the example cases collected in a knowledge elicitation session increases as model parameters are adjusted to develop more uncommon and extreme scenes. Also, it can be concluded that using the adjusted set of parameters over the unadjusted set increases the size of the state space, as measured by s_d^* , by nearly 20% on average.

5.3 Analysis of Hypotheses 3 and 4: Case Quantity, Visual Representation and Model Parameters

Hypotheses 3 and 4 state that case quantity is not affected by the visual representation or the model parameters used respectively. The alternative hypotheses are that the size of case quantity will increase with a higher level of visual display and with adjusted model parameters respectively.

Figure 7 shows the case quantity obtained for all 48 knowledge elicitation sessions against the visual representation used. There is a clear pattern with a reduction in case quantity for the 2½D display. The case quantities appear to be similar for the 2D and 3D representations. Also, ignoring the outlier, there seems to be little effect from the use of different parameters. These observations are confirmed through a two-way repeated measures ANOVA. There is sufficient evidence to reject hypothesis 3 at 5% significance. The alternative hypothesis, however, cannot be accepted as stated, since whilst there is a significant difference between the case quantity with a 2D or 3D display over a 2½D display, there is no significant difference between the 2D and 3D displays. Meanwhile, hypothesis 4 cannot be rejected at 5% significance. No interaction effect was found between visual representation and model parameters.

Figure 7 *A Comparison of Case Quantities under Different Visual Representations and Model Parameters*

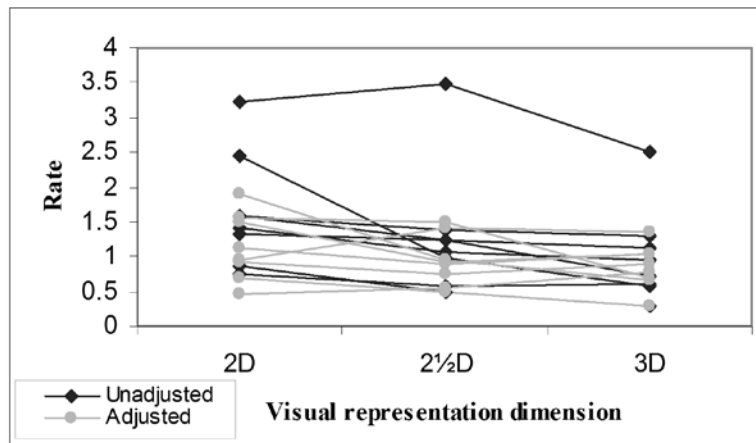


5.4 Analysis of Hypotheses 5 and 6: Collection Rate, Visual Representation and Model Parameters

Hypotheses 5 and 6 state that collection rate is unaffected by visual representation and model parameters respectively. Collection rate is defined as the number of example cases recorded per minute of real time in a knowledge elicitation session.

Figure 8 shows the results for the 48 knowledge elicitation sessions. There does appear to be some reduction in collection rate from increasing the dimension of visual representation, while there is no clear pattern from changing the parameters (especially if the outlier is ignored). The results of a two-way repeated measures ANOVA confirm these to be the case. There is sufficient evidence to reject hypothesis 5 at 7.3% significance and to conclude that a 2D representation results in a higher collection rate than a 3D display. It should be noted, however, that no significant difference could be found between the 2D and 2½D displays, or between the 2½D and 3D representations. There is also insufficient evidence to reject hypothesis 6. No interaction effect was found between visual representation and model parameters.

Figure 8 *A Comparison of Collection Rate under Different Visual Representations and Model Parameters*



6. Discussion of Findings

Section 3 presented a series of proposed effects on the effectiveness and efficiency of VIS as a knowledge elicitation tool. These effects are now discussed in the light of the results presented above.

6.1 The Effect of Visual Representation on Elicitation Effectiveness and Efficiency

Hypotheses 1, 3 and 5 pertain to the effect of visual representation on elicitation effectiveness and efficiency.

Decision Fidelity

The evidence collected in relation to hypothesis 1 suggests that decision fidelity improves with lower levels of visual representation; this contradicts the expected result. Hence, it seems this finding goes against the belief that if objects perceived in a real world environment are simulated faithfully, then experts can apply the same mental processes in the model that they apply in engaging with the real world (Barnes, 1996).

Further to this, and possibly more importantly, the apparent absence of high-fidelity decision-making in the knowledge elicitation sessions implies that VIS may have limited utility as a knowledge elicitation tool. Two potential reasons why there was a divergence between the decision-making in the model and the decision-making in the real system are the representation of information and the motivation of the experts. Both of these need to be addressed if VIS is to be an effective knowledge elicitation tool.

There are pieces of information that are difficult to represent in a VIS. For instance, the switch operator sometimes determines whether an engine has been tested or not by touching it. If the engine is hot then he knows it has been tested. The switch operator may also be subjected to verbal abuse if too much work is sent to one cell operator. The feel of an object and the anger of an operator are difficult to represent in a VIS, although efforts could be made to shade the engine and operator icons depending on their current state.

The motivation of the experts when using the VIS may not be well aligned with their motivation in managing the real world process. A particular issue arises when there are no consequences from making poor decisions in the model (Robinson et al., 2005). To be more effective in knowledge elicitation, then incentives in the model need to be aligned more closely with the real system. For instance, ways of mimicking the benefit of meeting throughput targets, such as cash payments, need to be resolved.

Case Quantity

There is sufficient evidence to reject hypothesis 3 and to conclude that the level of visual display does affect the case quantity collected in a knowledge elicitation session. However, it cannot be concluded that higher dimensions of visual representation increase the case quantity, since there is no significant difference between the case quantity with a 2D and a 3D representation. The significant difference occurs when a 2½D display is used, resulting in a lower case quantity.

Although opinion and surveys suggest that higher dimensional visual representations improve communication and provide more understanding of the system (Barnes 1996, 1997; Akpan and

Brooks, 2005a), these results suggest that higher dimensions do not improve the knowledge elicitation capability of a VIS. The result also contrasts with Akpan and Brooks' (2005b) finding that it is easier to identify model inaccuracies in a 2½D representation than a 2D one.

Collection Rate

From the case study there is evidence to suggest that a lower dimension of visual representation leads to a greater collection rate (hypothesis 5). In particular, the collection rate from the 2D model is higher than the collection rate from the 3D model, albeit that there were no significant differences between these two models and the 2½D variant. This suggests that the slower run-speed of the 3D display is reducing the collection rate. The collection rate in figure 8, however, is not reduced by anywhere near the reduction in run-speed. In section 4.5 we see that the knowledge elicitation sessions with the 3D model took up to 2½ times longer on average. This suggests that the slower run-speed increases the number of elicited cases per simulated time unit as the expert has more time to process the information being presented by the VIS.

6.2 The Effect of Model Parameters on Elicitation Effectiveness and Efficiency

As expected, the state space did increase when the model parameters were adjusted so that the experts were presented with more uncommon and extreme scenes. Indeed, the effect was to increase the state space by nearly 20% on average (table 4). This means that the training set obtained from a knowledge elicitation session will have a wider range of situations in which decisions are made. This should be beneficial for improving the training of a knowledge based system.

Meanwhile, there is insufficient evidence to conclude that the case quantity and collection rate are affected by the parameter settings (hypotheses 4 and 6). These findings show that although the uncommon and extreme scenes provided in the simulation might offer the experts more interesting situations that would retain their attention throughout the knowledge elicitation sessions, they did not materialise in more responses being elicited from the experts. Also, the

experts did not find the unconventional decision-making required by these uncommon and extreme scenes more difficult, as they did not take more time to make their decisions.

6.3 Overall Findings

In the case study the 2D visual display provides the most effective and efficient means of eliciting knowledge, when compared to the 2½D and 3D representations that were used. The 2D display resulted in the highest fidelity decisions and the highest collection rate. Although the case quantity was no different to that obtained from the 3D model, it was higher than the 2½D version.

In terms of model parameters, adjusting these to create more uncommon and extreme scenes had little effect, except to generate a wider state space. This, in itself, may be sufficient reason to recommend adjusted parameter settings as this would have a direct benefit in giving a wider range of situations in a training set obtained from a knowledge elicitation session.

7. Conclusion

The research reported in this paper focuses on the effectiveness and efficiency of simulation based knowledge elicitation in a manufacturing environment. In particular, the impact of alternative visual representations and parameter settings have been explored. The results and findings pertain, of course, to the particular case of the Ford Puma hot test facility that was studied. Quite different results might be obtained if the experiment were repeated in a different circumstance.

The apparent lack of fidelity in the decisions taken by the experts when interacting with the simulation models is disappointing. This may not be entirely unexpected given the difficulty of reporting some decision attributes in the simulated environment and the lack of incentive alignment between the simulation and the real system.

Given the benefits VIS offers for knowledge elicitation in terms of elicitation effectiveness (state space and case quantity) and elicitation efficiency, it seems to be worth pursuing the approach further. Future work should concentrate on improving the fidelity of VIS as a knowledge elicitation approach. Based on this research, the key areas to focus on are: identifying and modeling key decision attributes; and on aligning incentives in the simulated environment with those in the real system. The impact of the increasing use of 3D animation in everyday life on the effectiveness and efficiency of different levels of simulation visual representation should also be studied. Finally, repeating these experiments in different decision-making environments would also improve our understanding of the effectiveness and efficiency of VIS for knowledge elicitation.

Acknowledgements

This work was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) through the University of Warwick Innovative Manufacturing Research Centre. The authors also acknowledge the support of Ford of Europe and the Lanner Group in providing access to plant staff and software resources respectively.

References

- Akpan, J.I. & Brooks, R.J. (2005a). Practitioners' perception of the impacts of virtual reality on discrete-event simulation. In Kuhl, M.E., Steiger, N.M., Armstrong, F.B. and Joines, J.A. (eds), *Proceedings of the 2005 Winter Simulation Conference*, 1976-1984.
- Akpan, J.I. & Brooks, R.J. (2005b). Experimental investigation of the impacts of virtual reality on discrete-event simulation. In Kuhl, M.E., Steiger, N.M., Armstrong, F.B. & Joines, J.A. (eds), *Proceedings of the 2005 Winter Simulation Conference*, 1968-1975.
- Barnes, M. (1996). Virtual reality and simulation. In Charnes, J.M., Morrice, D.J., Brunner, D.T. & Swain, J.J. (eds), *Proceedings of the 1996 Winter Simulation Conference*, 101-110.
- Barnes, M.R. (1997). An introduction to QUEST. In Andradóttir, S., Healy, K.J., Withers, D.H. & Nelson, B.L. (eds), *Proceedings of the 1997 Winter Simulation Conference*, 619-623.
- Chan, F.T.S. & Chan, H.K. (2005). Design of a PCB plant with expert system and simulation approach. *Expert Systems with Applications*, 28 (3), 409-423.
- Chryssolouris, G., Lee, M., Domroese, M. (1991). The use of neural networks in determining operational policies for manufacturing systems. *Journal of Manufacturing Systems*, 10 (2), 166-175.
- Deslandres, V. & Pierreval, H. (1997). Knowledge acquisition issues in the design of decision support systems in quality control. *European Journal of Operational Research*, 103 (2), 296-311.
- Dransch, D., Kothur, P., Schulte, S., Klemann, V. & Dobsław, H. (2010). Assessing the quality of geoscientific simulation models with visual analytics methods - a design study. *International Journal of Geographical Information Science*, 24 (10), 1459-1479.

- Feigenbaum, E.A. (1980). *Knowledge Engineering in the 1980's*. Stanford, CA: Dept. of Computer Science, Stanford University,.
- Field, A. & Hole, G. (2006). *How to Design and Report Experiments*, Sage Publications.
- Flitman, A.M. & Hurrion, R.D. (1987). Linking discrete event simulation models with expert systems. *Journal of the Operational Research Society*, 38 (8), 723-733.
- Garthwaite, P.H., Chilcott, J.B., Jenkinson, D.J. & Tappenden, P. (2008). Use of expert knowledge in evaluating costs and benefits of alternative service provisions: a case study. *International Journal of Technology Assessment in Health Care*, 24 (3), 350-357.
- Hurrion, R.D. (1986). Visual interactive modelling. *European Journal of Operational Research*, 23, 281-287.
- Hurrion, R.D. (1991). Intelligent visual interactive modelling. *European Journal of Operational Research*, 54, 349-356.
- Krzanowski, W.J. (2005). *Principles of Multivariate Analysis*. Oxford University Press.
- Kunnathur, A.S., Sundararaghavan, P.S. & Sampath, S. (2004). Dynamic rescheduling using a simulation-based expert system. *Journal of Manufacturing Technology Management*, 15 (2), 199-212.
- Law, A.M. (2007). *Simulation Modeling and Analysis (4th ed)*. New Yoek: McGraw-Hill.
- Liang, T.P., Moskowitz, H. & Yih, Y. (1992). Integrating neural networks and semi-markov processes for automated knowledge acquisition: an application to real-time scheduling. *Decision Sciences*, 23 (6), 1297-1313.

- Lyu, J. & Gunasekaran, A. (1997). An intelligent simulation model to evaluate scheduling strategies in a steel company. *International Journal of Systems Science*, 28 (6), 611-616.
- Moreno, L., Aguilar, R.M., Piñeiro, J.D., Estévez, J.I., Sigut, J.F. & González, C. (2001). Using KADS methodology in a simulation assisted knowledge based system: application to hospital management. *Expert Systems with Applications*, 20, 235-249
- Negnevitsky, M. (2005). *Artificial Intelligence: A Guide to Intelligent Systems (2nd edn)*. Addison Wesley.
- O'Keefe, R.M. (1986). Simulation and expert systems – a taxonomy and some examples. *Simulation*, 46 (1), 10-16.
- O'Keefe, R.M. (1989). The role of artificial intelligence in discrete-event simulation. In Widman, L.E., Loparo, K.A. & Neilsen, N.R. (eds), *Artificial Intelligence, Simulation and Modelling*, Wiley, 359-379.
- Ong, M.H., West, A.A., Lee, S.M. & Harrison, R. (2006). A structured approach to evaluating the impact of implementing a component-based system in the automotive engine manufacturing domain. *International Journal of Production Research*, 44 (13), 2645-2670.
- Ören, T.I. (1994). Artificial intelligence in simulation. *Annals of Operations Research*, 53, 287-319.
- Paris, J.L., Tautou, L. & Pierreval, H. (2001). Dealing with design options in the optimization of manufacturing systems: an evolutionary approach. *International Journal of Production Research*. 39 (6), 1081-1094.
- Pierreval, H. & Ralambondrainy, H. (1990). A simulation and learning technique for generating knowledge about manufacturing systems behaviour. *Journal of the Operational Research Society*, 41 (6), 461-474.

Preece, J. (1994). *Human-Computer Interaction*. Addison-Wesley.

Rehn, G.D., Lemessi, M., Vance, J.M. & Dorozhkin, D.V. (2004). Integrating operations simulation results with an immersive virtual reality environment. In Ingalls, R.G., Rossetti, M.D., Smith, J.S. & Peters, B.A. (eds), *Proceedings of the 2004 Winter Simulation Conference*, 1713-1719.

Robinson, S., Edwards, J.S. & Yongfa, W. (1998). An expert systems approach to simulating the human decision maker. In Medeiros, D.J., Watson, E.F., Manivannan, M. & Carson, J. (eds), *Proceedings of the 1998 Winter Simulation Conference*, San Diego, CA: The Society for Computer Simulation, 1541-1545.

Robinson, S., Alifantis, T., Edwards, J.S., Ladbrook, J. & Waller, T. (2005). Knowledge based improvement: simulation and artificial intelligence for identifying and improving human decision-making in an operations system. *Journal of the Operational Research Society*, 56 (8), 912-921.

Roy, R., Adesola, B.A. & Thornton, S. (2004). Development and a knowledge model for managing scheduling disturbance in steel-making. *International Journal of Production Research*, 42 (18), 3975-3994.

Shaw, M.J. (1989). A pattern-directed approach to flexible manufacturing: a framework for intelligent scheduling, learning, and control. *International Journal of Flexible Manufacturing Systems*, 2, 121-144.

Tan, A.A.W. (2003). *The Reliability and Validity of Interactive Virtual Reality Computer Experiments*. Eindhoven, The Netherlands: PhD Thesis, Technische Universiteit Eindhoven.

- Tan, A.A.W., Timmermans, H.J.P. & de Vries, B. (2000). Investigation of human behaviour in a virtual environment. In *Proceedings of the VWSim'00 Conference, 2000 Virtual Worlds and Simulation Conference*, San Diego, California.
- Trueman P., Lawson K., Bending M., Ganderton M., Chaplin S., Wright D., Duffy S., Saxby R., Chilcott J., Tappenden P., Pilgrim H., Shorthouse A. & Tappenden J. (2007). *Bowel cancer services: costs and benefits*. Final Report to the Department of Health.
<http://www.shef.ac.uk/scharr/sections/heds/modelling/colorectal-cancers.html>. Accessed June 2011.
- Waller, A.P. & Ladbroke, J. (2002). Experiencing virtual factories of the future. In Yücesan, E., Chen, C.H., Snowdon, J.L. & Charnes, J.M. (eds), *Proceedings of the 2002 Winter Simulation Conference*, 513-517.
- Williams, T. (1996). Simulating the man-in-the-loop. *OR Insight*, 9 (4), 17-21.
- Yima, N-H, Kima, S-H, Kimb, H-W & Kwahkc, K-Y. (2004). Knowledge based decision making on higher level strategic concerns: system dynamics approach. *Expert Systems with Applications*, 27, 143–158.