University of Southampton
Faculty of Business and Law

An Artificial Intelligence Framework for Feedback and Assessment Mechanisms in Educational Simulations and Serious Games

by

James Stallwood

Thesis submitted in completion of a Doctor of Philosophy (PhD) award

July 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS AND LAW

Design

Doctor of Philosophy

AN ARTIFICIAL INTELLIGENCE FRAMEWORK FOR FEEDBACK AND ASSESSMENT MECHANISMS IN EDUCATIONAL SIMULATIONS AND SERIOUS GAMES

by

James Stallwood

Simulations and Serious Games are powerful e-learning tools that can be designed to provide learning opportunities that stimulate their participants. To achieve this goal, the design of Simulations and Serious Games will often include some balance of three factors: motivation, engagement, and flow. Whilst many frameworks and approaches for Simulation and Serious Game design do provide the means for addressing a combination of these factors to some degree, few address how those factors might be affected by the presence of an out-of-game tutor. It is the position of some researchers that the presence of real-world tutors in a Simulation or Serious Game experience can be shown to have a detrimental effect on motivation, engagement, and flow as a continuously changing state for the participant from in-game to out-of-game breaks immersion.

The focus of this study was to develop a framework for the design of Simulations and Serious Games that could provide the means to mitigate some of these identified negative effects of real-world tutor. The framework itself, referred to as the Wrongness Framework, uses artificial intelligence techniques and practices to provide internal feedback and assessment to the participant as a foundation for the creation of a rudimentary in-game tutor.

To achieve this goal it was necessary to develop the Wrongness Framework to include not only the findings of other scholars and researchers on the topic of feedback and assessment but also to introduce original refinements to existing artificial intelligence mechanisms. To test the abilities of the Wrongness Framework it was applied to two unique case studies each with a different purpose and scope. The first, the AdQuest case study, was a graphic design Serious Game scenario testing the ability of the Wrongness Framework's assessment mechanisms by having 102 postgraduate design students submit graphics for a luxury brand advertisement. These graphics were then assessed by the Wrongness Framework against expectations found in the Wrongness Framework's Intelligent System Knowledge Bank. The students were then surveyed for their responses to their assessments and individual rating scores for each design were taken. The second case study, Promasim, explored the possibilities of feedback tone and efficacy for non-player characters in a project management simulation. This was achieved with the use of expert interviews by both academics and working professionals to provide the information of experienced project managers to develop experiential interaction events for the Simulation.

Despite the results of these case studies a full case for the success of the Wrongness Framework could not be made. However, many of the identified challenges for the Wrongness Framework were met and, as such, a case can be made that an adequate foundation for the framework has been successful and has provided the case for further refinement.

Table of Contents

Table of Figures

page

Table of Symbols

Symbol	**Meaning**
$\forall$	“For all”
$\exists$	“There exists”
x	Element
$X = \{...\}$	Set
Xy	Statement
$X(y)$	“X of y”
$x \in X$	“x in X”
$x \notin X$	“x not in X”
$x \rightarrow y$	“If x, then y”
$x \Leftrightarrow y$	“If, and only if x, then y” / Equivalence
$\neg$	Negation
$x \wedge y$	“x and y”
$x \vee y$	“x or y”

Academic Thesis: Declaration Of Authorship

I, James Stallwood, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

An Artificial Intelligence Framework for Feedback and Assessment Mechanisms in Educational Simulations and Serious Games

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Either none of this work has been published before submission, or parts of this work have been published as: [please list references below]:

Stallwood, J., Ranchhod, A., and Wills, G., 2013, “Fostering Creativity in Young Adults through Game Development: A Preliminary Investigation”, *INTERREG*, France.

Stallwood, J., Ranchhod, A., and Wills, G., 2013, “A Model for in-Simulation Assessment”, *International Computer Assisted Assessment*, GB.

Signed: ……………………………………………………………………………

Date: July 2015

With special thanks to my supervisors Ashok Ranchhod and Gary Wills,
my fellow developers, and all those whose assistance made this possible.

Introduction

The purpose of this thesis is to propose a framework which can be used in the design of Simulations and Serious Games which will provide feedback to the participant and assess the participant's actions within the Simulation/Serious Game from a background mechanism that is present within the Simulation/Serious Game. By providing this framework, designers of Simulations and Serious Games are able to remove some of the need for a real-world instructor and instead give Simulations and Serious Games more autonomy in training their participants. It is the argument of this thesis that doing so will provide a number of distinct advantages to the various users of Simulations and Serious Games. Firstly, with the an integrated feedback and assessment mechanism, training solutions facilitated by Simulations and Serious Games using this framework will have further reaching potential for more participants; particularly those in areas where local traditional instructors may be few and far between. Secondly, that by centralising as much of the experience of the participant to what occurs within the Simulation or Serious Game as possible provides a better immersive experience and one that can be properly regulated by the context of the Simulation or Serious Game.

Whilst the far-reaching social or political implications of such a framework are important they are not the main focus of this research, rather the focus is more on the application of a framework and it's local implication for the design of Simulations and Serious Games. To this extent this should be read as a guide of how and why to apply the framework rather than the specific when to apply the framework.

* * *

Chapter One:
The State of Play in Simulations and Serious Games

The first chapter of this thesis is devoted to providing the background literature that will be used in order to establish a design framework which can be used when developing games-based learning software. The first, and most important question, to ask is whether a new framework as this need exist at all. Simulations and Serious Games have existed prior to this research and, no doubt, more still will have been designed and developed concurrently to it also. Wherein is the justification for such a framework to be used in Simulation and Serious Game design? To explore this question further it is necessary to first understand what is meant by Simulations and Serious Games as terms and define their latent characteristics.

It is assumed, for the purposes of this framework, that any reference to Simulations and Serious Games within this research be computer-based virtual environments. Whilst a long tradition of Serious Games and Simulations have taken place in the real-world, for example role-playing, environmental simulations, and political simulations (Shaffer, 2007a), these experiences already have an internal and regulating instructor position undertaken, ordinarily, by the person leading the Simulation or Serious Game. Similarly it is necessary to differentiate between Educational Simulations and simulations in general, as with general simulations there is no inherent requirement for any training or instruction to take place. Because of this it is necessary to differentiate between Educational Simulations, which are the topic of this work, and simulations of theoretical models or known physical phenomena (Aldrich, 2009b). Therefore, it should be understood within the content of this work that upper case S Simulations will be shorthand for Educational Simulations and lower case s simulations will refer to all other types.

The difference in the definitions of Simulations and Serious Games can seem quite small at first, particularly when given in general terms. Suppose we begin with a simple definition for Educational Simulations as:

A [virtual] environment, created to represent a real-world environment as closely as possible, which attempts to represent a scenario or number of scenarios within a predetermined and agreed level of complexity for the purposes of training the participant in some process or processes (Aldrich, 2009b; and Dondlinger, 2007).

Comparing this definition with one for Serious Games the similarities between the two become quite apparent:

A [virtual] environment, which attempts to represent a scenario or scenarios within a predetermined and agreed level of completing for the purposes of training the participant in some process or processes. This environment should also provide a compelling reason to promote engagement for the participant(s) to continue to use the Serious Game and progress further in their training (de Freitas, 2006, and Ulicsak and Wright, 2010).

These general definitions for Simulations and Serious Games represent traits that must be present for a games-based learning tool to qualify as either. However, these definitions do not prohibit design decisions which go further than the benchmarking of those definitions. For example, a Serious Game may have no need to present its environment within a realistic setting but there is nothing to stop a designer from developing a Serious Game which does so. However, certain characteristics are inherent and necessary. Because of this certain researchers (Aldrich, 2009b) have seen different e-learning software as being part of a greater spectrum of characteristics or approaches. This is a pleasing idea as it provides the means by which an example of e-learning software can be analysed by its effective or ineffective use of certain characteristics rather than by predefined archetype blueprints.

The matter of how effective these media are in delivering their instruction is not necessarily dependent on their defined characteristics. If this was the case then there would be no need for both Simulations or Serious Games to exist simultaneously – either one would suffice. The learning environments that both Simulations and Serious Games have the potential to present are without question of some educative value but this value can only judged in comparison to what might be achieved without using these tools.

The ability to engage with some hostile environment, in the case of some ecological disaster, without needing to be there is of great importance. The opportunity to have the very costly, such as a symphonic orchestra, or difficult, such as space travel, available at the click of a button is also useful. However, it is in the mundane that this case for Serious Games and Simulations must be made. Why should it be that any instructor would prefer to run a potentially costly piece of software, and one which is time-consuming to develop and worse still runs the risk of being constantly out of date with current thinking in the subject and will need continuing support? The only acceptable answer is that there is something or some characteristic of Serious Games or Simulations which make them preferable alternatives to more traditional types of learning.

1.1. Other terms used in this thesis

Engagement

Engagement is referred to by Jennett et al (2008) as the “first level of immersion”. Engagement is said to have happened when a participant has demonstrated enough capability to interact with the Simulation or Serious Game in a meaningful way. This 'meaningful' interaction requires that participant interaction with the Simulation or Serious Game remains on-task (Robertson and Howells, 2007).

Epistemic

This term is used by Shaffer (2007) to describe interaction within Simulations and Serious Games which is restricted to the confines of the real-world actions of a particular role performed by the participant. This means that the participant must not only act in a way that is expected of a certain role but that they are given access to the “tool set” of that role by the Serious Game/Simulation itself. Also, the reward structure of the Simulation or Serious Game should be a replica of what is to be expected by the role in real life. For example, if a participant is to assume the role of a journalist then the Serious Game or Simulation must allow them to conduct interviews, write articles, etc. The participant's skill in fulfilling this journalist's role must also be judged against expectations of a journalist's ability with certain tools and techniques.

Lastly, if the participant succeeds then their reward for success must match what a journalist would expect for that success, e.g. to be permitted to tackle 'bigger' stories. This is, of course, to be the same approach in the case of failure as well.

Finite State Machine (FSM)

A finite state machine is a building block in the development of simple artificial intelligence. Each state provides the description of some behaviour that may contain multiple actions or calculations that the artificial intelligence agent is to perform. Also present in finite state machines are descriptions of how the agent is to transition between one state and another and from where the information to make that decision is derived. An agent can exist only in one state at a time in a finite state machine.

Flow

Flow is described by both Killi (2005) and Connolly et al (2012) as a participant's total absorption within a task such that they are “so involved” with a Simulation or Serious Game task that it consumes their entire concentration. Flow is, therefore, used as a relative term in this thesis which can describe how involved a participant is with a task, scenario, or gaming experience as a whole. Factors, either in-game or out-of-game, can consequently be described as affecting flow either positively or negatively.

Immersion

Immersion is a term used in this thesis, which describes the amalgamation of **flow** and **engagement** in a participant using a Simulation or Serious Game. Though this thesis makes no attempt to quantify immersion, it might be useful to think of it as a description of how well a participant is performing against how optimized their interactions are with the Simulation or Serious Game.

Intelligent System

A computer system, which uses different artificial intelligence techniques to govern its processes (Russell and Norvig, 2010). Typically it is a reactive system.

Intelligent Tutoring System

Intelligent Tutoring Systems are artificial intelligence agents that guide their users through a number of questions in a given topic and, typically, assess their user's abilities based on the completeness of the user's answer (Van Eck, 2007). One vital aspect of Intelligent Tutoring Systems is that they should guide their users through questions of a complexity which is deemed suitable for that user. There are, consequently, many different assessment and feedback techniques which might be used to accomplish this goal.

Older

When referring to Simulation and Serious Games participants this term simply means "of adult age", which is to say from the ages of 18 and greater. Any further delineation of participants who fit this description is either given in addition to this comparative, e.g. "older students", or causes this term to be dropped altogether, e.g. "professionals".

Participant

This is a simple term, which refers specifically to users of a Simulation or Serious Game who might be described as the intended audience. Participant is preferred over another term, such as "user", as user might equally apply to a tutor or supervisor who is responsible for leading the Simulation or Serious Game experience. Participant is also preferred over "player" as player used in conjunction with the term game or games might have specific connotations (Aldrich, 2009b).

Rules-Based System

A type of **intelligent system**, a rules-based system is designed to respond to different scenarios of changing input variables based on strictly defined logical processes, or "rules". These rules dictate how and when a rules-based system responds and, consequently, input variables can be manipulated to achieve a specific reaction (Luger, 2005).

Scenario

In this thesis the term scenario refers to a self-contained interaction experience within a Simulation or Serious Game. This might be a single task, or a collection of tasks related to achieving a single goal.

Younger

School-aged children not including those who classify as being in further education. As with the term **older** any necessary delineation is given in addition to this term. Typically these will include categories associated with school periods, e.g. “primary”, “secondary”, “infant”.

1.2 The foundations of Aldrich's Slates of Educational Experience

Aldrich (2004) proposes that there are four slates of educational experience when using Simulations. Each of these slates represents some form of engagement with the participant and other users of the Simulation at different levels of interaction and complexity. An overview of these slates is provided below:

Slate One, “[T]he locker room”. This slate uses more conventional teaching techniques in which the participants are introduced to rudimentary knowledge about the subject (models, frameworks, etc.) and the goals of the programme of learning.

Slate Two, “[T]he shallow end of the pool”. Participants engage in a limited version of the full Simulation to acclimatize themselves with the interface of the Simulation and engage in simple interactions with the Simulation. Then the participants will be able to offer feedback on this preliminary interaction.

Slate Three, “[T]he deep end of the pool”. Participants engage, with instructor supervision, in more complex and open-ended Simulation scenarios. This is can best be achieved by having participants tackle Simulation tasks in teams, though there is no requirement each must provide direct input.

Slate three Simulations also have ways of demonstrating the “flow of the simulation process” and achievements of the participant so that, either synchronously or asynchronously, the participant or instructor can view at a glance the progress made by the participant in the Simulation.

Slate Four, “[F]ree swim”. Participants are left to engage with the simulation by themselves so that they might hone the skills they have been learning, or push the level of engagement with the Simulation. Aldrich believes it is necessary for a participant to spend a number of hours doing so so as to allow the skills being developed by the participant to be done so to an intuitive level.

For Aldrich (2004), these slates are the bedrock for the successful delivery of a Simulation. He goes so far as to claim that, whilst more precedence might be given to certain slates over others or the devotion to one slate over another might be contracted or expanded upon, each slate is sequentially necessary and none of them can be missed if a successful Simulation experience is to be had.

The research, and subsequent frameworks, developed by de Freitas and Jarvis (2009) (Fig.1.3), Amory (2007) (Fig.1.1.), and Kiili (2005) (Fig.1.2.) go some way toward validating the content of Aldrich's slates, if not necessarily confirming the order. Whilst each of these frameworks has a particular focus, taking a broad sweep through all of them provides a more composite picture of how effectively Aldrich's slates of educational experience can be integrated into the design of Serious Games and Simulations.

Amory's game object model version two, shown below (Fig. 1.1.), is fairly comprehensive in demonstrating each of the key activities undertaken by the participants of Aldrich's slates and the associated outcomes of those activities within a delineated hierarchy via object characteristics.

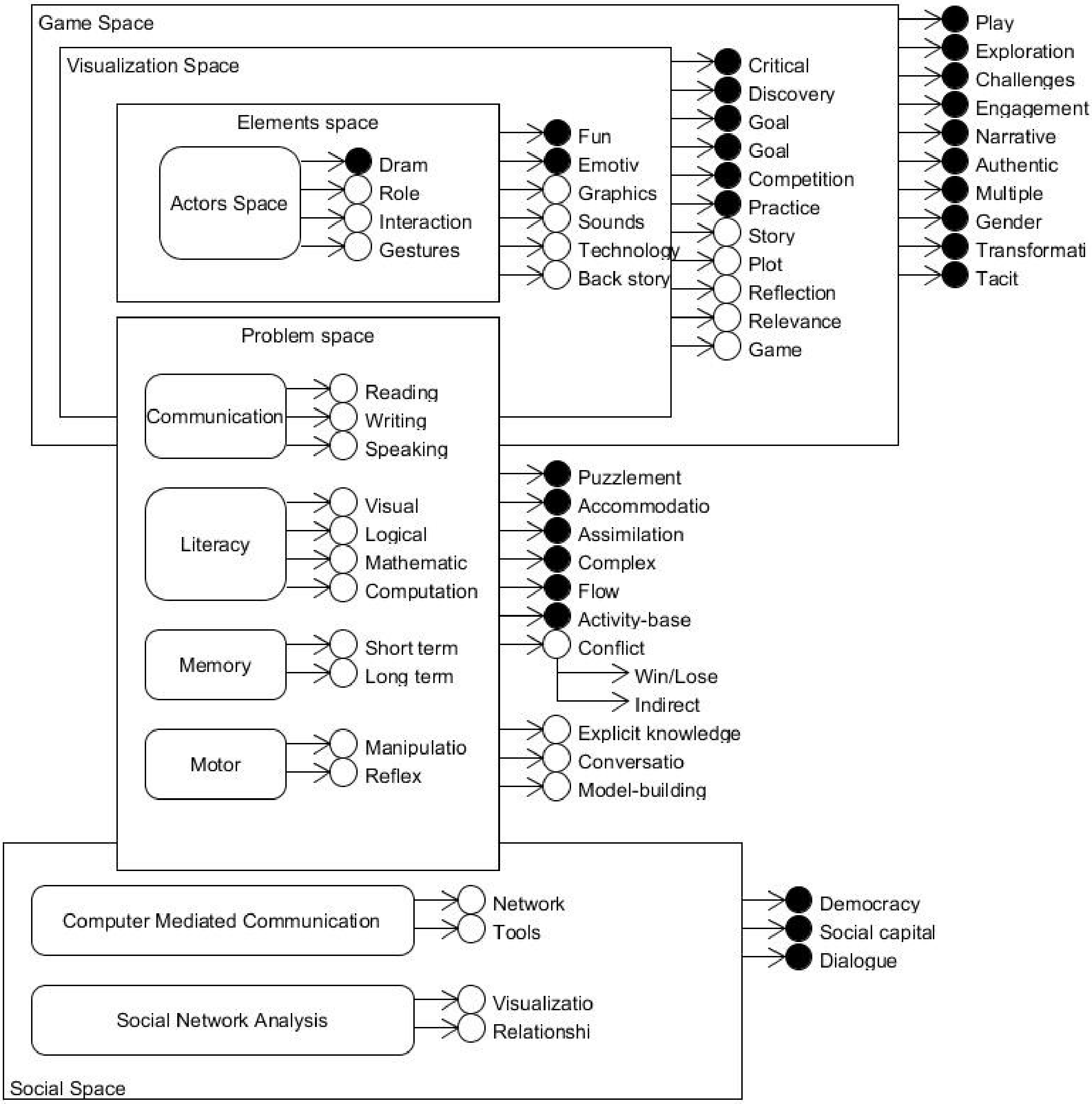


Fig. 1.1., *Game object model version II*, reproduced from Amory (2007)

This model, in and of itself, is not indicative of the process of learning through Simulations, of which GOMII has received criticism in the past (van Staalduin and de Freitas, 2011), but rather does provide points of reference, acting rather like an ordnance survey map of experience, whereby other models and frameworks may be situated to better explain Simulation and Serious Game process.

This is different from the design approach described in Kiili's (2005) experiential model (Fig. 1.2). The focus of the experiential model is based on participant interaction informing a journey through activity states rather than attempting to account for behaviours in an explicit and narrowly defined way. Though less complex as a higher-level model, this allows for more flexible designs of Simulations and Serious Games and provides the means for more experimental interaction between the participant and the Simulation/Serious Game itself. In this way flow between activity states and participant achievements are central not only to the design of Serious Games and Simulations but also in facilitating the experiences required in the third and fourth of Aldrich's slates – as the flow of active experimentation.

Though Kiili (2005) admits that this model is not concerned with the social aspect of gaming, self-evaluation and team reflection can easily be accommodated by intuitive interface design. To this end, the experiential model describes the clear expectation of participant behaviour which should be promoted in visualization and problem spaces of the GOMII model.

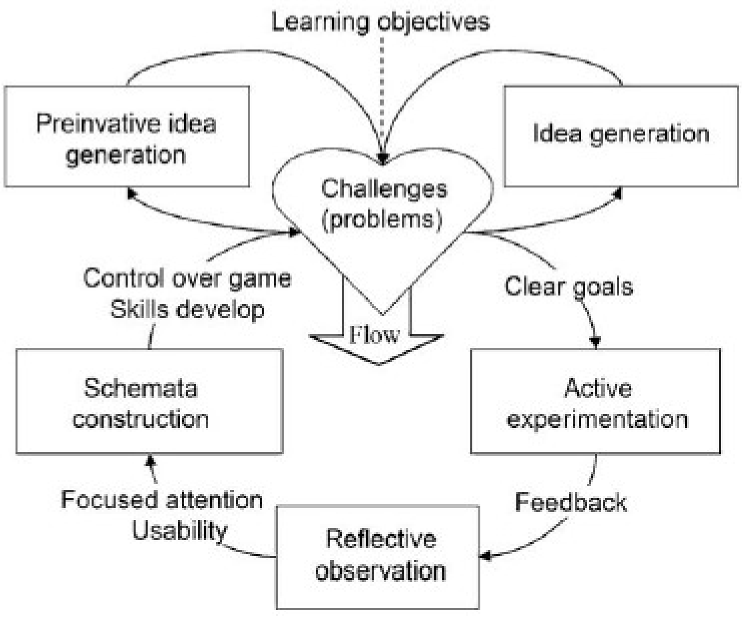


Fig. 1.2., *Experiential Model*, (Kiili, 2005)

The experiential model is not without its limits however. Kiili (2005) freely accepts that as a model for design, the experiential model (Fig.1.2.) does not fully encompass all of the design qualities that a good Serious Game or Simulation should include.

Instead the model acts rather as a bridge between pedagogical theory and game design integration. Design aspects such as narrative, participant advancement and achievements (levelling), and, quite extensively, cognitive load, are not accounted for however, the effects of those aspects and how they should be handled within the experience.

In addition to these design aspects the development approach for games-based learning shown in Fig. 1.3. (de Freitas and Jarvis, 2009) describes various techniques and means by which participants might engage with game-based learning tools more effectively by trying to increase their satisfaction with the games-based learning solution. This is an important aspect for consideration as all other models described previously make an assumption that the participant or participants are already convinced by or have some regard for the games-based learning activity they are to use. Ultimately this model is best suited towards assessing the validity of a games-based learning tool in the first place according to the needs of a client or stakeholder rather than solely focusing on the needs participant group themselves. Even within this limited scope however, the development approach of Fig. 1.3. does provide the basis for asking important design questions regarding the feasibility and need for a proposed game-based learning solution to exist. The overall effect is thus, to engender a feeling of familiarity and trust with the games-based approach to education for a given context.

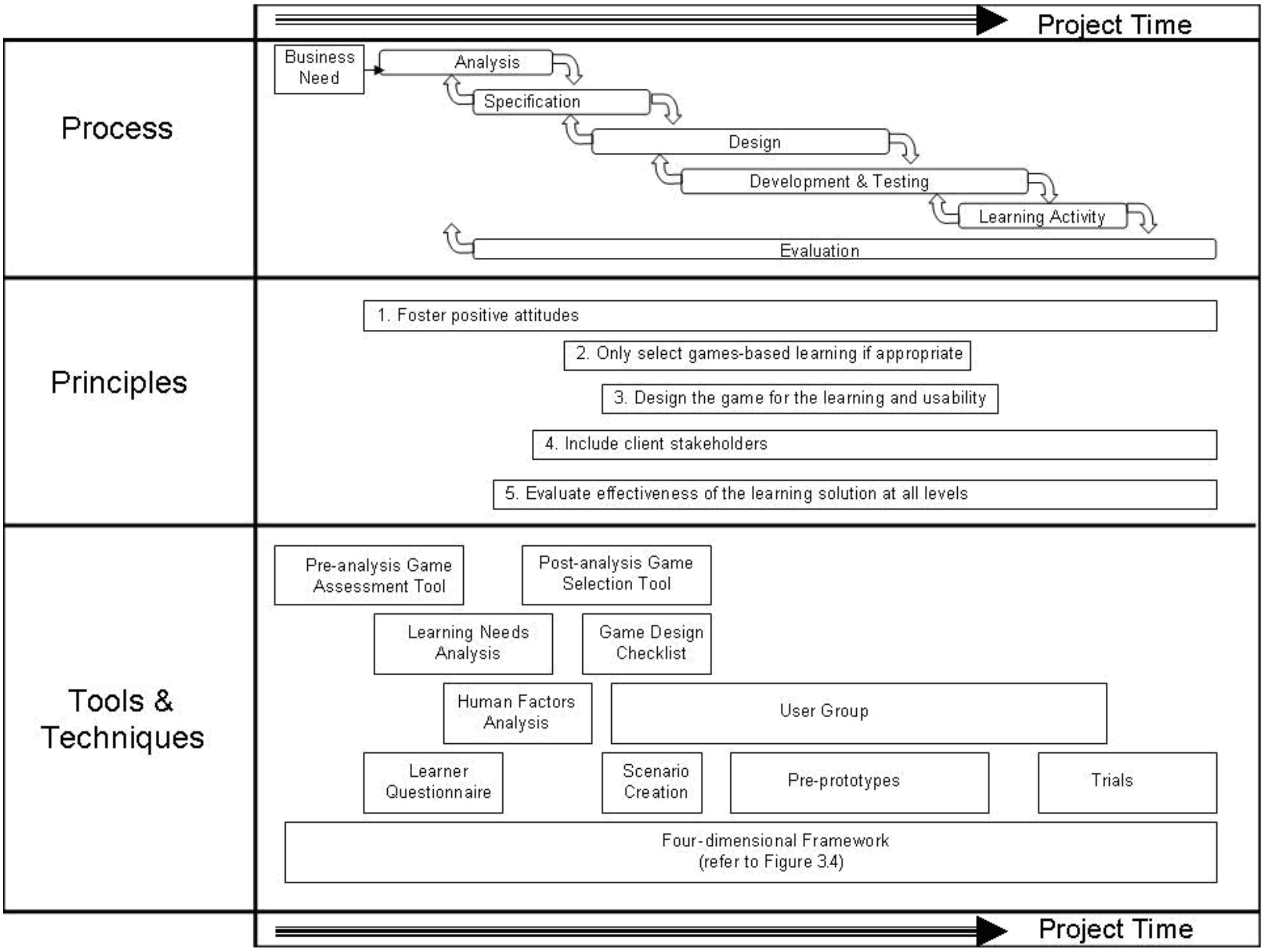


Fig. 1.3., *Development approach for game-based learning*, (de Freitas and Jarvis, 2009)

If the models by Kiili (Fig. 1.2.) and Amory (Fig.1.1.) provide the means by which specific interactions and events within the Simulation/Serious Game experience can take place, this model establishes the groundwork which permits Aldrich's four slates of educational experience to be adopted in toto. The principles section of the model provides not only the means by which we can return to the original question of why we should want to bother with Simulations/Serious Games as a educative solution, but offers a means of validation with the various stakeholders of the solution after adoption; including, as it does, a level of quality control for the evaluative processes.

If these models can be said to have validated externally the slates proposed by Aldrich from a general, or high-level, viewpoint then it would be pertinent now to explore still further their specific interactions and that of the expectations of each of the slates in more specific and lower-level detail.

1.2.1. A critique of Aldrich's Slates with respect to specific case studies

Aldrich's (2004) slates describe four stages, which are considered to be an optimal Serious Game experience. As these slates are useful for describing necessary interactions between the participant, Simulation/Serious Game, and a tutor, it would be profitable to use them as some over-arching framework for understanding how experiences should be managed. To fully adopt these slates as a framing device for understanding the research of others it is important to ascertain that Aldrich's descriptions match what is commonly practised by researchers or users who run Simulations or Serious Games experiences. This would indicate that Aldrich's slates are a good indicator of common experiences of Simulations and Serious Games, supporting some of the current literature.

It must be understood, however, that Aldrich's slate descriptions are limited by the technologies of the time in which they were written. For example, Aldrich (2004: 17) refers to low-fidelity options for hosting Simulations online where no CD-ROM is available as a means of facilitating the fourth slate experimental phase for participants. When critiquing these slates, it is important to keep their historical position in mind. Thus if an experiment or experience is cited that was conducted within a similar time period as these slates were described it may be that technological limitations would have made facilitating a specific slate experience impractical or impossible for that researcher. The main issue is that the Aldrich provided a conceptual background for gaming experiences. These are timeless and will undoubtedly be affected by technological changes.

Aldrich's four slates can be described simply as four phases, or steps of experience. These are:

1. Slate One – participants are introduced to the goals of the Simulation/Serious Game experience and are given a foundation in the subject knowledge required to participate. This slate can take place within the virtual learning environment or in a real-world setting.

2. Slate Two – participants are guided through a basic run of the Simulation/Serious Game so that they might become familiar with in-game interactions or other features. This is an opportunity to present the subject knowledge relevant to the Simulation/Serious Game in an in-simulation scenario context developing the understanding developed in slate one.

3. Slate Three – participants interact with a more refined and subtle version of the Simulation/Serious Game and certain previous help features are removed. This is to stop an over-reliance of help features to make decisions. Feedback and assessment is given to the participant from both the virtual learning environment itself and real-world tutors if they are present. The purpose of this slate is that participants should have a full experience of the Simulation/Serious Game to make use of the formal subject knowledge they have learned in slates one and two in a controlled environment.

4. Slate Four – participants continue with the simulations alone and engage in their own experiences without a formal setting or direct interference from real-world tutors. Their pace in the Simulation/Serious Game is dictated by their own interactions. Participants may still communicate with each other but in less strict settings.

Aldrich (2004: 17) does not specify a time period for the experiences, noting that the full four phases might be experienced in days or, more probably, over a longer period of months or even years. Therefore, there is nothing to dictate how quickly these slates must be experienced by the participants, or that there should be a minimum number of sessions for that experience. The only condition is that the slates ought to be experienced sequentially to provide an enhanced Simulation experience (Aldrich, 2004: 14). This is an important condition to be aware of as it means that for the validation of a gaming experience the activities have to not only match the experiences with the slates but also the ordering of these activities have to be within the prescribed frame.

To begin it is not difficult to find examples of case studies of Simulations or Serious Games experiences that fit the description of Aldrich's slates broadly. Ulicsak and Wright (2010: 30) describe the play of a case study of the Serious Game “SimVenture”; a business simulation. Their description of the gameplay for a participant matches with the description of the first three of Aldrich's slates. The game initially describes the setting and provides subject knowledge to the participant – slate one. The participant then engages in a simple scenario with limited activity – slate two – and then progresses to more complex business situations and activities developing their learning through assessment and feedback – slate three. It would be difficult in some instances to demonstrate the case for slate four, however, Ulicsak and Wright (2010: 31) also provide feedback taken from an interview with a tutor that demonstrates some practice of it. In the interview it is detailed that the SimVenture is, in this instance, run over a five-week period and participants are encouraged to engage with the Serious Game without further tutor observation – slate four.

Whitton's description of a case study of the game “Marketplace” (Whitton, 2010: 171 – 172) shows a similar experience with that of SimVenture. One of the key differences between the Marketplace and SimVenture case studies was that, unlike SimVenture, the participants of Marketplace followed a much more structured cycle of integrating knew knowledge with formal lectures and tutorial sessions at the beginning and during the Serious Game experience. This kind of external support might be explained by the fact that the case study for Marketplace took place with undergraduate students as an elective module and SimVenture was used to supplement the learning of school-aged students. In order to meet specific standards, the delivery of Marketplace also had to contain other specific goals (educational) not related to the delivery of the Serious Game itself. However, despite this more rigid delivery approach the Marketplace experience that Whitton (2010) describes tallies with the description of delivery of Aldrich's slates.

It is possible to find validation for Aldrich's slates in other sources than direct Simulation/Serious Game case studies. Shelton's (2007) assessment of activity-alignment in the Serious Game “River City” draw parallels with the way in which activity and engagement should be paced according to Aldrich's slates. In it Shelton (2007: 109 - 110) describes what he refers to as the alignment of activities and gameplay features in such a way that does not hurt the flow of participant engagement by becoming “distracting”.

Shelton's advocacy is for an introductory phase of activities to acclimatise the participant followed by the introduction of increasingly more complex activities and subject knowledge over time. What is more telling is that Shelton's description of instructional support matches an ideal that after a certain point of establishing a goal-oriented activity driven experience it will be possible to taper off that instructional support (Shelton, 2007: 110). In this it is possible to draw direct parallels between Shelton's ideal experience description and the one proposed in Aldrich's slates two, three, and four.

A similar description of an idealised delivery can be found in Frank's (2007) writings on challenge as a “focus” of Serious Game design. Like Shelton, Frank (2007) espouses the gradual increase in complexity that begins with a simple initial stage in which participants are exposed to basic subject knowledge and simple problem solving activities to demonstrate their competence before moving on to more challenging problems. Unlike Shelton, however, Frank (2007: 570) considers the implications of activity-flow for the single- and multi-player experience. For the single-player experience Frank (2007) argues that challenges must be set and managed, i.e. assessment, within the Serious Game itself. This leads Frank to conclude that in multi-player instances those challenges can afford to be more dynamic and complex as they can account for externally set objectives as well as those found in the Serious Game itself. Frank (2007: 569) establishes these conclusions on a description of how the goals of the Serious Game should be established and fixed in the minds of the participant before deeper exploration of the Serious Game in a way that is very similar to the process advocated in Aldrich's slate one. From this basis, Frank argues, the exploratory flow of activities can be established, linking up with the descriptions of slates two and three and Shelton’s conclusions (2007: 110). Whilst Frank's description of ideal participant experience does not account for activities described in Aldrich's fourth slate, his description does tally in part with Shelton's and Aldrich's three slates adding further validity to the idea that Aldrich's slates can be relied upon to describe an ideal Serious Game/Simulation experience.

Dunwell et al (2011) consider not only the increasing complexity of activities in a Serious Game as Shelton (2007) and Frank (2007) do, but also elaborate on the factor of timing. This is important, because often previous experiences in the delivery of educational goals through activities point to a sequential approach as being optimal, other options for a tutor can be made available.

Dunwell et al. (2011), for instance, consider timing from the perspective of when to provide feedback to the user but also touch on feedback timing as part of a sequential delivery; as Aldrich's slates. Here, in the example of the game “Triage Trainer” Dunwell et al (2011) consider feedback and feedback of what kind ought to take place to assist the participant. They clearly note that certain activities and certain feedback should be offered to the participant at different stages of the Serious Game experience and, decide, that early supportive feedback for simple initial tasks – similar to what is described in Aldrich's second slate – is more beneficial at that stage. Conversely, Dunwell et al (2011) clearly separate the Serious Game experience in a way that is analogous to Aldrich's slates by suggesting later stages should be more complex and make use of delayed feedback to put the onus of learning on the participant – as Aldrich's third slate.

The research presented in this critique does appear to generally support the description of Aldrich's slates for an optimized Simulation/Serious Game experience, however there are procedural difficulties. Chiefly among them is how one should go about ensuring that participants continue with their experience alone as described in slate four. This is an issue of engagement and motivation rather than a direct challenge to the concept of the Aldrich's slates of experience. Another difficulty relates to the support participants receive when no real-world tutor is available or if no community exists within the Simulation/Serious Game. The design and refinement of a Simulation/Serious Game needs to accommodate these needs. It should thus be stressed that challenges or difficulties that might arise in providing an experience following the guidelines of Aldrich's slates in no way invalidate them. Rather they should be seen as a framework to be improved upon to control a Simulation/Serious Game experience. In the next few passages of the kinds titled *Facilitating the X,* the groundwork for understanding these challenges is given by examining further literature that can be used to support those slates. At this stage, however, it is enough to say that Aldrich's slates have been independently demonstrated to represent common practices for providing Simulation/Serious Games experiences and can thus be considered to have been validated as a guide for practice.

1.3. Facilitating the locker room

The presentation of knowledge and the way in which it is handled to promote the maximum possible absorption by participants of Simulations/Serious Games, is a subject on which there is continued debate. Earlier studies undertaken by researchers like Malone (1980) suggest that one such way is to use the cognitive curiosity of the participant; to which Malone goes further to suggest two methods to achieve this cognitive curiosity. The first such method might be to seemingly undermine what the participant actually knows or believes to be true, Malone gives the example of establishing the process of photosynthesis and then pointing out that certain plants can thrive in darkness. The premise of this approach then is to challenge what a participant thinks they know to an extent that having such a thing challenged will reinforce incremental learning.

Arguably, a by-product of this undermining principle would be that the participant would be so intrigued by having their views challenged that they would provide a self-motivation for continuing with the e-learning (Salen and Zimmerman, 2003). Another alternative suggested by Malone is to provide only incomplete subject knowledge for the participant such that by taking part in the activities of the Simulation/Serious Game they might, through experimentation, fill any gaps in their knowledge, and having derived this knowledge by themselves it is more readily remembered – a technique which thrives within Kiili's experiential model.

Malone's summation of cognitive curiosity is validated by the rather more recent proposal of intrinsic integration by Habgood and Ainsworth (2011). As a design framework, intrinsic integration, as defined by Habgood and Ainsworth, suggests, rather as Malone's research does, that participants of Simulations or Serious Games can find reward in simply accomplishing a task and do not necessarily require external stimuli to compel them to continue with their learning.

In addition to this definition, Habgood and Ainsworth propose two central characteristics to intrinsic integration which are that:

- The delivery of subject material and the engagement of learning should take place within the flow of the Serious Game/Simulation and, preferably, at moments in which the game play is particularly entertaining.

- That the environment created within the chosen e-learning archetype and the participant's interaction with that environment be fundamentally based on the subject material.

The second of these characteristics is similar to the findings of Shaffer (2007a) and Gee (2003) regarding epistemic games and engagement and motivation factors respectively. However, the first of the characteristics, in conjunction with Malone's findings, provides an interesting design query regarding the aims of Aldrich's first slate. How should the initial presentation of subject knowledge take place? How to make this knowledge restricted or undermined to promote Malone's idea of cognitive curiosity and restrict it in such a way that the restriction itself is not undermined by the necessary mechanics of the first characteristic of intrinsic integration? Would it even be fruitful to attempt this anyway? These questions are largely dependent on the type of participant likely to engage with the e-learning.

Habgood and Ainsworth (2011) believe that there are limitations to the effectiveness of their framework to be overcome with regards to how successful e-learning software designed with this framework would be for younger participants. The reasons for their belief in these limitations are based upon the pedagogical research about the principle of concrete learning; which is to say younger people have greater difficulty in transferring knowledge and how to apply that knowledge in new, seemingly unrelated, scenarios (Habgood and Ainsworth, 2011). de Freitas (2006), in her review, promotes the power of immersive environments to assist in the transferral of skills from the digital, or “rehearsed”, to real-world scenarios – particularly for, as she terms them, “underserved learners”.

Shaffer (2007b), however, proposes a different perspective on the issue. The argument of epistemic design would be to say that the need to provide an opportunity for participants to demonstrate skills transfer from one scenario to another is of lesser importance than the need for a Simulation/Serious Game scenario to require precise knowledge for problem solving as a matter of design. In epistemic Simulations the context for learning is clear and well defined, and the related subject knowledge needed for participant interaction within the Simulation is demonstrated in any number of scenarios which would show transferred knowledge (Shaffer, 2007b). Because of this, an inappropriate setting will remain an inappropriate setting no matter what alterations we might make to the design. Thus, this forces us once more, within the guidelines of the model given by de Freitas and Jarvis (2009) in Fig. 1.3 above, to question whether or not a Serious Game or Simulation solution is even appropriate for any subject matter which would have these kinds of knowledge transfer issues.

Older participants of Serious Games and Simulations share many similar if not identical pedagogical needs as with younger participants. As with offering an introduction to the subject matter, there is a very real need to offer a context for the learning; in accordance with the expectations of the first of Aldrich's slates. Whitton (2010), makes a clear case for meaningful and clear objectives for adult participants to assist with their motivation, which, often is more self-serving and self-reflective than the motivation of younger people. However, she also outlines the importance of providing a structure, reminiscent of Serious Games and Simulations designed for younger people, to assist older participants in their own experiential learning. Indeed, little, beyond the learned ability of adults to apply knowledge in different situations other than those presented to them, need be different for the design of a Simulation or Serious Game using the first educational experience slate.

Both adult and young learners learn best when left to explore their surroundings empirically (de Freitas and Jarvis, 2009; de Freitas, 2006; Whitton, 2010; Ainsworth and Habgood, 2011; and Kiili, 2005) and Serious Games and Simulations offer safe, enclosed, and engaging environments in which experiential learning can take place.

The matter of how a Simulation or Serious Game can be designed such that it can assist the participant in this experiential engagement is, fundamentally, a question of engagement, flow, and providing a compelling reason for them to begin to do so or continue to do so.

1.4. Facilitating the shallow end of the pool

At this stage of the Simulation matters of flow and engagement are of the utmost importance as this will be where the participant first experiences the Simulation proper. How then can a Simulation/Serious Game be designed such that it is engaging enough that the participant will wish to continue with it, or indeed, even take it seriously as a learning experience by the users? Likewise, the question of how to balance the flow of the experiences within the Simulation/Serious Game is equally important as it will influence informed feedback, assessment, the completion of goals, and thus, ultimately, the motivation of the participant. As should be clear, engagement and flow are therefore interlinked through motivation; particularly as motivation is often gained through challenge (Robertson and Howells, 2008).

Motivation, as a product of engagement and flow, is best achieved in endogenous fantasies (Garris et. al., 2002; and Shaffer, 2007a); which is to say situations which represent likely real-world scenarios. In real-world scenarios a clear definition of what is to be expected (learning objectives) and what can be expected by the participant is more easily achieved and, thus, the quantifiable assessment of these objectives. In their seminal review, Garris, Ahlers, and Driskell (2002) propose the Input-Process-Outcome model for promoting an engaging and educative experience with Simulations and Serious Games (or rather games with an educational context) which uses feedback and reflective methods to aid participants in achieving learning objectives.

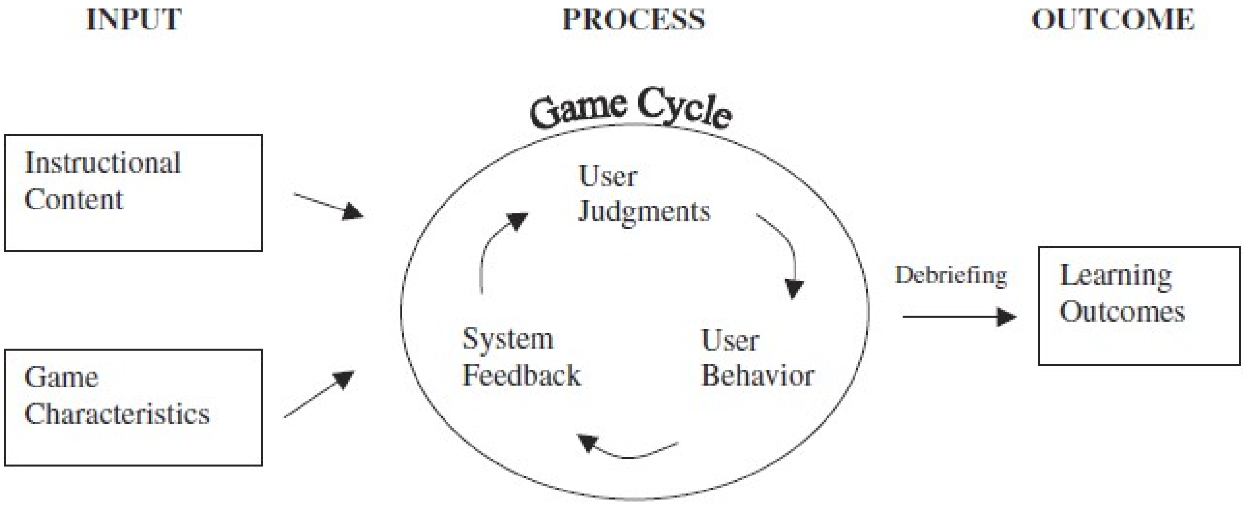


Fig. 1.4., *Input-Process-Outcome Model*, (Garris et. al., 2002)

Participant engagement is key to the Input-Process-Outcome model and goes far beyond simply interacting with the Simulation or Serious Game itself as an entity. In order to promote further learning the engagement for the participant has to be wider reaching than that. To this end, motivation can be said to take place in two ways: motivating participants to learn more about the subject of the Simulation externally to the Simulation, and motivating participants to explore further the possibilities afforded to them by the Simulation and thus promote experiential learning within wider and more complex circumstances.

The effectiveness of games to promote continued learning of a subject is an area of some contention. In their study of how games-based learning motivated students to further engage with mathematics, Kebritchi, Hirumi, and Bai (2010) found that there was no significant increase in motivation for the participants of the study, although actual scholastic achievements by the participants was seen to increase across the group. Their explanation for such results suggest that, though the games adherence to the experiential model of learning was well defined, their control group may have been influenced unfairly by the instructors leading the game-based learning and, most importantly, the game which was tested simply could not qualify, by definition, as a Serious Game or Simulation.

The last of these points may seem arbitrary, if not a little petty. How might one design a Simulation or Serious Game to test mathematical knowledge without imposing some kind of superficial environment on top of it? For a subject with such broad applications as mathematics surely narrowing down the environment to only a few specific instances where that knowledge may be of immediate relevance is going to isolate learning even further?

This is a valid objection but it does strike at the heart of the original question; does a Simulation or Serious Game for this subject need to exist or, indeed, should it exist in this way? Written word problems of the sort *"A car [A] is travelling from London to Birmingham at a rate of 45 miles per hour. A second car [B] is travelling from Birmingham to London at a rate of 70 miles per hour, etc..."* are already an endogenic fantasy and aside from the usual arguments pertaining to different kinds of learners and even the socio-political criticisms of word problems such as these, how are they universally improved by animation? This question may appear to be difficult to answer as a justification, but if taken as a general guideline it becomes easier to clarify what exactly one should expect from a Serious Game or Simulation in terms of both motivation and the virtual environment.

The Input-Process-Outcome model proposed by Garris, Ahler, and Driskell (2002) (Fig.1.4.) establishes the importance of effective feedback in facilitating engagement and flow. However, where this model breaks down is that it fails to account for the need for participants to offer their own feedback. Other such studies have shown that effective engagement with experiential learning is related directly to the participants being able to share ideas with their fellows (Robertson and Howells, 2007; and Ke, 2010) though each conclude on the importance of instructor's support for the learner in this process.

Yaman et al (2008) note that for most participants, learning new skills is difficult without instructional support. On this topic of instructional support, Yaman et al (2008) conclude that it is in providing external instructional support that participants are often able to achieve more experiential and independent learning. However, they also realise that to reduce the complexity of a Simulation to better facilitate this external instructional support may relegate the Simulation to a mere instructional aid.

Robertson and Howells (2007) take the stance such that instructors in such circumstances are no longer chiefly concerned with actively teaching, or rather instructing, but instead are responsible only for providing the environment by which a participant can engage with the Simulation/Serious Game fully. In terms of the second slate of educational experiences that means allowing the Simulation/Serious Game to engage the participant and provide flow. So as a matter of design this means creating an environment where engagement and flow, through feedback on experience, is structured for the participant by the instructor but not necessarily directed by them (Egenfeldt-Nielsen, 2004). This balance is key to avoiding the control issues faced by the participants in the study by Kebritchi, Hirumi, and Bai (2010).

The roles of instructors in Simulations and Serious Games may seem knotty and complex but, as has been established earlier, the way in which participants engage with whichever e-learning tool is being used and the flow of that engagement is central to providing a motivational experience. So far all the examples which have been cited have come from external instructor sources. To do this is to make feedback and sharing for participants an alien process to the Simulation or Serious Game itself as it removes participant engagement from the e-learning environment. The question then becomes how, if at all, can such processes be integrated into the virtual experience itself.

In order for the environment of the Simulation or Serious Game to provide feedback it is necessary to create an interaction method which not only is natural to the context of the subject and the fantasy, so as not to break immersion, but also one which can communicate the participant's intentions as unambiguously as possible. In order to make such queries it would be necessary to make this interaction as simple as possible, i.e. a few key command words or actions, which would lead necessarily to ambiguities, or, to make the commands so relativistic and precise in each circumstance in which they are likely to occur that a large expert system would be needed for even the simplest of e-learning software.

In the research undertaken by Charles et al (2011) in games-based feedback the core design tenet for Virtual Learning Landscapes (VLLs), which is to say the environment of the Simulation or Serious Game, was a participant-centric design and interface which would provide basic feedback based on the interactions of the participant.

Whilst the design of this VLL provided simple feedback to the participants on as to whether or not challenges they were attempting to undertake were feasible based on acquired prior knowledge about the participant, it gleaned this information on a reward system which cannot be said to be endogenous. Indeed even the environment in which the participants found themselves was not endogenic to the subject of the game. This is not a bad thing of course, but, rather, leaves open the question as to whether similar measures could take place in Simulations or Serious Games and how feasible it would be to do so.

There is of course more required than just this. All users of the Simulation or Serious Game would have to have the means of interacting with each other in such a way that did not break immersion, could be mediated by the e-learning tool itself, and also one which would provide the means by which the feedback could be two-way and complex enough to be circumstantially useful.

1.5. Facilitating the deep end of the pool

In many regards the leap between the second and third of Aldrich's slates is the smallest. In this slate, in addition to more complex Simulation and Serious Game scenarios, the participant should be expected to interact with their fellow users in a much more refined and deliberately conscientious way; that refinement of course comes by means of prior experience from earlier Simulation interactions. This need for multi-user input for the Simulation/Serious Game is key to the experiential approach; indeed, if somewhat unsurprisingly, researchers have found that e-learning in isolation is much less effective than undertaken as a group activity (Robertson and Howells, 2007; Ke, 2010; and Yaman et al, 2008).

Maintaining the continuing interest in these interpersonal Simulation/Serious Games relationships for the users goes beyond what has already been discussed regarding motivation as a product of engagement and flow – though these elements are important. Jennett et al (2008) make a very interesting case for immersion arguing that, rather than being a by-product of e-learning tool which has been engineered to achieve goals of engagement and flow, immersion itself can be something quite separate.

In their review, Jennett et al (2008) make the usual connections between engagement and flow and include in addition the roles of cognitive absorption and presence. Immersion, they argue, is not necessarily connected to flow. Rather Jennett et al (2008) take the position that to design for flow is to design for an optimal experience and immersive experiences are not required to be optimal. Similarly cognitive absorption, which is described in like terms to Malone's (1980) cognitive curiosity, is not necessarily immersive. Indeed, as their argument goes (Jennett et al, 2008: 643), it is possible for someone to be highly absorbed with the act of acquiring knowledge but not in the setting of the Simulation or the means by which that knowledge is presented. Lastly, that immersion and presence are different using the example that the video game Tetris can be highly immersive but in no way would the player believe that they are actually present and moving the coloured blocks themselves.

The definition, and by inference the desired design requirement, that can be taken from such arguments is that immersion is the process of holding the participant's attention against possible distractions from the Simulation/Serious Game. The authors go further to suggest that immersion will also lead to a lack of distraction by issues of time, "awareness of the real-world", and and will lead to increased involvement with the virtual environment and its tasks. It is important then to suggest that whilst immersion, as argued by Jennett et al (2008: 643), is separate from the three areas of flow, cognitive absorption, and presence, they are not unrelated but, rather, that they are distinct from each other.

Using this definition the importance of making multi-user interaction between either participant peers or instructors endogenous to the Simulation/Serious Game rather than external to it as an alien process is important to the creation of an immersive experience and one which will have knock on effects with motivation and advancement. A picture begins to form of the necessity of trying to integrate each of these features into the Simulation/Serious Game itself as structured characteristics which can be identified but not manipulated by the user. Indeed it is pertinent to go further and ask what advancements in the use of e-learning tools such as these might be made if such an integrated system could be created.

Garris et al (2002), likewise, depict the matter of immersion in the same way, albeit with stronger connections to engagement than the definition given by Jennett et al above. Immersion, when linked to engagement variables, is strongest when interaction between participants and other users and the virtual environment results in immediate and natural results, sensory stimuli from the environment are rich and varied, and lastly, as a point made above, distractions from external stimuli are reduced significantly (Garris et al, 2002: 453). This is fundamentally a matter not only of background mechanics but, principally, presentation (de Freitas, 2006).

Online multi-user environments are certainly one approach towards achieving these means, and given the levels of interactivity they afford as opposed to say group input around a single input station. Interactivity as a means for immersion has been extended by game-based learning in such environments. Aldrich (2009a) provides metrics, in the form of a logarithmic scale (i^n), for the quantification interactivity in these circumstances; interactivity of participants in traditional settings and environments with increasingly games-based tasks.

These levels of interactivity fall between those at level zero (i^0), where the instruction method - be it a book, or a lecture, or whatever - is delivered without engagement by the instructing authority, to level six (i^6), where an open-ended experience is had by the participants and the instructing authority reverts to the facilitator rather than the instructor - a desired state for Simulations and Serious Games as previously discussed in the above sections. Aldrich (2009a) develops this idea in different directions to the instructor as a facilitator line of Kebritchi, Hirumi, and Bai (2010), and Robertson and Howells (2008), is in matters of leadership models sprouting from these different interactivity levels. At early levels, zero to two – level one being the stage Aldrich believes many e-learning courses and textbooks end at, the instructor is assertive and determined to be applying pressure on the participants to comply. Aldrich states that only short-term goals can or should be expected from this approach and participants are more reticent to comply and thus retain information.

At later stages, it follows, the attitude is one of assistance between instructor and participant which have largely opposite effects on cognitive absorption, engagement, and, overall, interactivity. The authority with which an instructor presents themselves, be that from an internal or external source, is of great importance to the probability of success for the participant.

1.6. Facilitating the free swim

The last of Aldrich's slates is really a summation of all the issues and challenges facing the facilitation of the earlier slates. In this the participant, left to experience the environment of the Simulation or Serious Game, must be encouraged to engage with and explore this environment beyond the its initial curiosity (slates one and two), and as a shared experience (slate three). Factors governing motivation, interactivity, and immersion must, therefore go further, and permit more open-ended experiences in order to, ultimately, maintain the interest of the participant. Managing challenge will promote cognitive curiosity; new interactions, likewise, will further experiential learning. Liberty, and good design, will provide the means for this to happen in Simulations and Serious Games. However, it is a matter of course that this can only happen by internal regulation. Experiential learning, as provided for in the Killi's model, can of course be shared and there is value to that sharing process. However, there is a danger that through external regulation the experiential pursuits of individual participants are hampered or held back by the attempts to ensure that all participants experience at the same rate.

From a simple viewpoint, internal regulation is handled by two means: that the participant is aware of what they do and do not know and, thus by their interactions and achievements, secondly, the Simulation or Serious Game is aware of the participants capabilities. This point can be easily demonstrated with the below mathematical problem.

$$\left(\sqrt[3]{e}\cdot\left[\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}\right]^{\frac{2}{3}}\right)^{\ln\sum_{i=1}^{4}8+i} = ?$$

To begin with, the answer to this problem is of little importance as the phrasing of the question itself demonstrates that it is a test of definitions and functions rather than one of calculation. To this end the matter of finding an answer to this problem is simply one of process. How far the person attempting to answer this question will go with this process is dependent entirely on what they are capable of doing. For some, early arithmetical training will be all they have and will recognise the use of exponents, and maybe even the logarithms found in the problem. Others may go further and still not solve it and more may go further still solve the problem altogether.

Though it is a sliding scale the types of people who might attempt to solve such a problem would fall into one of three camps. The first, those with limited subject knowledge and understanding, are likely to be overwhelmed by the unfamiliar symbols or even the many layers of the question. Those in the first camp are unlikely to attempt to answer the question and, even if they should guess at an answer and even further still, if that guess was correct, the person themselves will glean no educative satisfaction from the event (Kirkland et al, 2010). They will be turned off from answering such questions. Those of the second camp will have, albeit possibly to varying degrees, a greater understanding of the problem in front of them and can at least isolate what about the problem they do not understand and can seek to fill that knowledge, else they can deduce an answer with an educational guess and continued experimentation. Members of the last of these three camps know the process by which to answer such questions and, if sufficiently comfortable, will complete them with relative ease.

Cognitive curiosity, as proposed by Malone, will be greatest then in participants of the second camp; as to provide challenges which fall within the empirical powers and knowledge of the participant (Kiili, 2005). For those in the first camp and those in the third such problems can offer little educative satisfaction.

This may seem to be a fairly straightforward deduction - that there is no joy in a problem which is too difficult to solve or too inconsequential - however, the manner in which this factor is handled internally by the Simulation or Serious Game is the difficulty. Managing challenge is principal to fostering cognitive curiosity, which then knocks on affecting all other areas such as engagement and flow, thus immersion and motivation.

The self-regulation, through self-evaluation, of the participant in terms of challenge is possibly the more difficult to guarantee. How are they to know that the challenge they are about to engage with is at an appropriate level for their capabilities, and if it is not, how are they to be made aware of how much of a gap exists between their abilities and the challenge. In short it is a question of whether the participant should be guided to a different set of challenges more suited to their current knowledge and skills or that they might, with some effort, succeed at the current challenge level with a few more attempts and a little aid.

Research in non-invasive adaptivity in Simulations and Serious Games has been undertaken with special emphasis on challenge. The work by Pierce, Conlan, and Wade (2008) in using the ALIGN system – Adaptive Learning In Games through Non-invasion (Fig.1.5) – shows a form of structure in which the interactions of the participant structure the flow of the narrative, challenges, and thus objectives of the Simulation/Serious Game. In premise it is a less constrained method of adapting the experiences of the participant than a Simulation/Serious Game controlled by an Intelligent Tutor System, which provide a more personalised learning experience, tailored by the actions of the participant.

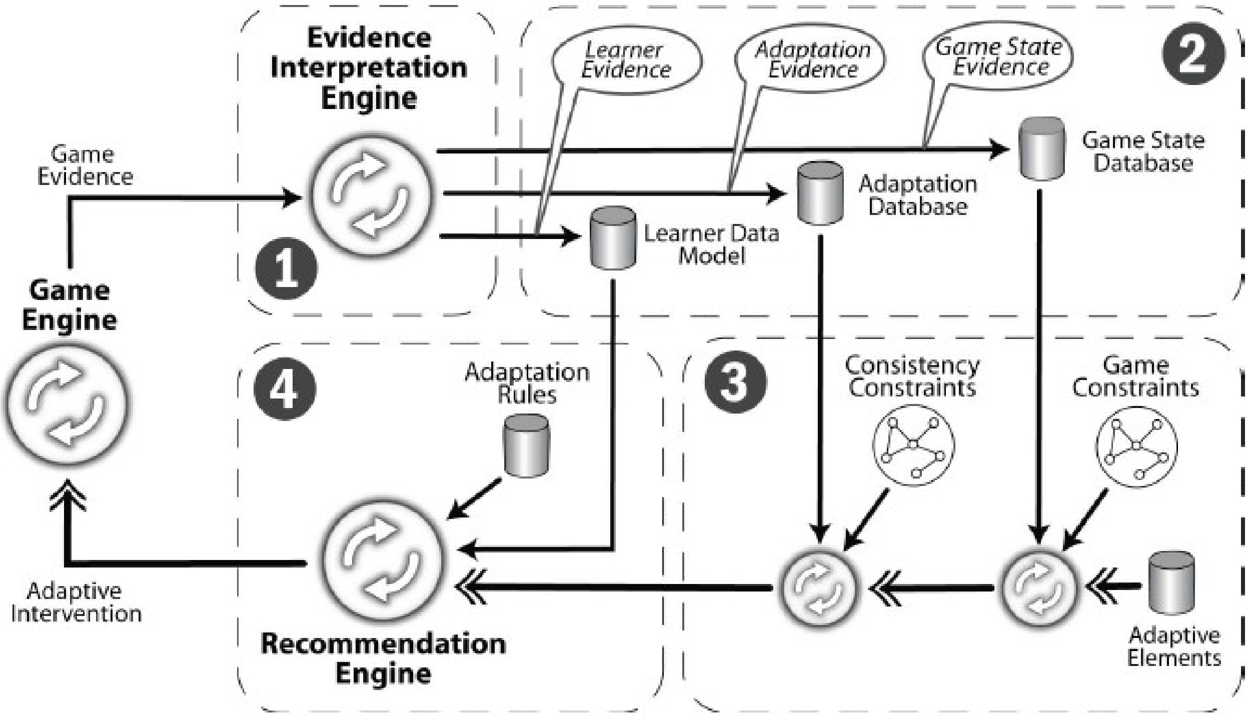


Fig. 1.5., *ALIGN architecture diagram*, (Pierce et al, 2008)

The cyclical nature and constant refinements made through the ALIGN system creates a dynamic participant experience and, with carefully considered and rules-based constraints, this system can provide the means by which a participant can recognise their ability to undertake specific challenges. However, for its dynamic approach the ALIGN architecture is yet still governed by the necessity of its rules-based system and the limitations of the adaptive experiences which can be had from this system.

These limitations can seem rather odd at first. Surely, in a simulated environment of any kind, it is possible just to define the rules of participant-environment engagement to any complexity we would wish? Well suppose, by way of metaphor, that a scenario in a Simulation or Serious Game required that the participant force entry into an elevator. It may be possible that a relatively strong screwdriver might be wedged between the doors and the mechanism could be forced. Should the Simulation or Serious Game see the screwdriver as the only means by which the door can be opened and should the participant fail to find the screwdriver they will be unable to proceed? Why should a screwdriver be acceptable but similarly bladed instrument not be; such as a fire axe?

Well allowing both increases the complexity of the solution from one characteristic - “is a screwdriver” - to two characteristics - “is metal and bladed”. What if this elevator were next to a cafeteria and wooden trays and cutlery were at hand. A table knife is metal and bladed but the strength of the material, presumably stainless steel, of such an instrument may be called into question. Would the non-metal yet bladed dining tray be sufficient? The result of this process would result in a long list of in-game objects suitable of accomplishing the task and either an increasingly precise or increasingly ambiguous set of characteristics which each of those objects must possess. It is conceivable that the designers of the Simulation or Serious Game of this scenario may well just pass a constant boolean CAN_OPEN_MECHANIZED_DOORS variable for each game object. What then if the participant should choose override the circuitry of the magnetic fastenings for the elevator instead of prizing them open? Adaptivity via rules-based systems thus have their limitations for how adaptive they can be.

This brings us to the expectations of the types of experience that a participant can expect. The above scenario of forcing open elevator doors may be one which would provide a thrilling experience for a participant in an action-adventure game. If, rather, the scenario took place in a Serious Game designed to demonstrate the safety features of elevator circuitry and how they might be compromised and, thus, how to identify the means by which they have been compromised – not unlike the training of ethical hackers – then it might be reasonable to expect that the participant's interactions with the game are limited only to electronic interfaces rather than manual force. In this way it is reasonable to suggest that the challenge presented and the adaptivity of the challenge is not necessarily related to the goal of that challenge and what the participant is trying to achieve. As Salen and Zimmerman (2003) point out, hitting a golf ball with a stick many hundreds of yards away from the hole that the player is trying to put it in to is a distinctly odd way of achieving the goal of putting a ball in a hole.

Adaptivity of challenge for the purpose of entertaining is a rather more pleasant idea in terms of participant motivation. As with the description above of ensuring that participants, as much as possible, remain within the second camp of cognitive curiosity, entertainment through adaptivity of challenge, as suggested by Thomas and Young (2010) can be used to maintain engagement as well as dictate flow and the structure of Simulation/Serious Game interaction. This positioning of the participant is described as the optimal game play corridor – difficulty as a function of time – graphically in a narrow frame between boredom and frustration (Fig.1.6).

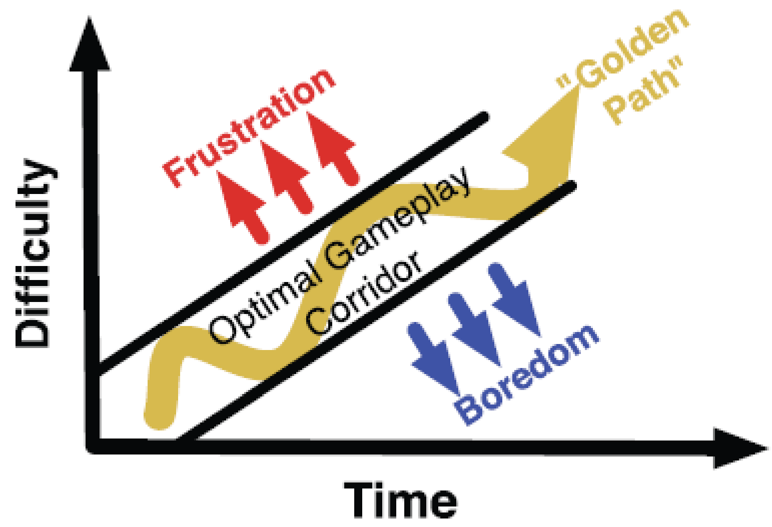


Fig. 1.6., *Optimal game play corridor*, (Thomas and Young, 2010)

Thomas's and Young's Intelligent Tutor System, called Annie, as in Annie Sullivan of Helen Keller fame, has been pitched as a method by which Serious Games can use intelligently tutoring system mechanisms without removing any of the inherent entertainment of game play. Annie's principal is fairly simple. Annie's knowledge of the participant is provided by the participant's own actions. How well the participant performs at a certain task will dictate their progression through the plan state of the game. The plan state is a mapped progression of the participant's abilities as they have been demonstrated followed by a calculation to determine the new challenges they will face. Annie, as a relatively new example of an Intelligent Tutoring System, governs participant interaction in an incremental manner and at a pace which is defined by the participant itself.

Though Thomas and Young (2010) focus on entertainment makes the dubious suggestion that something can only be engaging and thus, motivating, if it is fun, Annie's incremental management of participant achievements and how challenge can be adapted provides a solid case study for participant led flow and engagement in Intelligent Tutoring Systems.

1.7. A case for further refinement

The overview which has been given of the various topics and some of the practices governing the educational experience in terms of Aldrich's slates provides something of a way of determining where Simulations and Serious Games can be refined. Up until this point however, this research has focused only on the integration of pedagogical theories and learning behaviours in Simulations and Serious Games. Whilst these are important topics it would be useful to discuss some of the technological limitations and requirements that might be applicable to provide a broader context for refinement.

As ever increasing interests of the development of technologies and technological approaches grow throughout the world and policy for such changes is laid down (United Nations, 2013), the range of methods by which peoples are likely to interact with Simulations and Serious Games is likely to change. Within the developing world the use of mobile technologies is quickly becoming integrated in all means of development and education (Syngenta, 2011; GMSA Intelligence, 2013a) and amongst Western powers the use of mobile technologies are making considerable gains in education and matters of health (GMSA Intelligence, 2013b). It would thus be remiss to ignore this market for interaction within Serious Games and Simulations particularly as it, currently, provides the means of communication, business, and education for millions of people.

Despite the changes to mobile technology which are likely to occur during the course of this study, it would still be difficult to provide a large-scale Simulation or Serious Game for the developing world using these devices. This is due in part to the unstable broadband connectivity that still remains in some regions, and the relatively high costs to the user for their download data (GMSA Intelligence, 2013a).

Similarly any such advances to alleviate these issues in the developing world will have value in the Western world and Asia, as educational and training opportunities increase in number – particularly in formal training services for business where expansion can be facilitated by rapid and uniform training of all employees. To this end, it would be pertinent to summarize the challenges which any such refinement may face in attempting to incorporate these newer external factors into Serious Games and Simulations and how they might readily be integrated into the current corpus of knowledge and approaches on the subject.

From the external factors provided above and the strengths of projects such as Annie and the ALIGN systems a case can be made for the proposal of a refined framework which can govern the Simulation and Serious Game experience. In doing so it is necessary that this framework be in-line with theory surrounding Aldrich's educational slates and also, given the technical limitations addressed above, one which can have greater autonomous interactions with the participant where no instructor/mediator may be possible. It is possible that as well in creating such a mediator that factors such as immersion and engagement may be increased as, argued before, the chief activities of feedback and assessment between both participant, peers, and mediator will not be external to the Simulation/Serious Game itself. Adopting the epistemic approach of Shaffer (2007b) and the research of tutoring systems undertaken by VanLehn (2006), Johnson et al (2005), and those mentioned above will be key to achieving these goals. To this end the research question of this thesis becomes: can such a framework be developed which incorporates the design practices described in those models and can assessment and feedback mechanisms be developed for the framework which would provide complex and dynamic user interactions and user assessments?

Within the next chapter the framework to achieve these goals will be proposed; the Wrongness Framework. In this chapter the mathematical description of the framework, the proposed integration model, and examples of how this framework might be potentially used will be described. Part of this description will also include the methodological process by which the Wrongness Framework might best be evaluated by using case studies. It is worth noting that, as this is a framework rather than an engine, there will be no specific platform considerations nor source code presented nor even special criteria for the types of suitable media that the framework should exist within.

The Wrongness Framework is to be adaptive, fit for purpose in a variety of Simulation/Serious Game situations. Consequently the variety of the types of case study in which the Wrongness Framework is used will vary in terms of subject content and even in terms of the likely audience, such that their will be no particular focus on creating Simulations/Serious Games specifically for children of a certain age or within a specific field such as medicine. It is thus possible, in evaluation, that the Wrongness Framework may well be better suited to specific Simulations/Serious Games for a specific audience or a specific context or even a particular type of learner but it will be presented in the beginning with as much universality of purpose as is possible.

Whilst the framework itself is to be applicable it should be understood that, as a background entity, its relation to foreground matter such as the graphical and layout quality of on-screen events will be discussed only in potential terms. Rather, the Wrongness Framework has distinct knock-on effects for what can and should be displayed to the user but the means by which this might be achieved is an area of study in and of itself. Consultation will have to occur, where possible or appropriate, for specific media matters to specific Simulation/Serious Game contexts but that will not specifically be addressed by the application of the framework.

* * *

Chapter Two:
The Wrongness Framework: An Overview

2.1. Introducing the Wrongness Framework

The Wrongness Framework is, at heart, a simple idea. There are, relatively, right decisions and wrong decisions, and thus, some decisions which would be more right or wrong than others for any given situation. Determining what is a “right” decision in a Simulation/Serious Game scenario and still maintaining an adaptive and immersive experience would require an expert system of enormous size and complexity. It is, thus, far easier to determine what is a wrong decision for a participant to make. The Wrongness Framework, at its core, simply asserts that what isn't wrong must be right. On face value this may appear to be a fairly self-evident suggestion; an advocacy of the law of the excluded middle. In many ways this is true, however, unlike the rather absolute terms of the law of the excluded middle the Wrongness Framework works on a relativistic scale in which terms, if given enough suitable complexity, any decision can be made as a binary choice.

The means by which the Wrongness Framework determines any decision is with the use of two flags: Wrongness Flags and Cumulative Malpractice Flags. In any scenario there will be actions which can be described as wrong no matter what the specifics of a situation. For example, there is never a case where embezzlement would be correct practice, nor, to give a more practical example, the division of a polynomial by an unknown variable in algebra. Wrongness Flags, therefore, do not deal in moral absolutes then but rather absolutes of practice within a subject.

Cumulative Malpractice Flags are somewhat more refined and complex. These are flags which, whilst not immediately incorrect, act rather as blots against the participant through their actions. For example, a single instance of going over budget in a proposed business plan may not raise alarms the first time – after all why else have risk assessment? - but continually doing so will affect the opportunities that the participant will be able to experience and may even result in disciplinary actions.

At this stage it is important to bring in the epistemic approach discussed in the previous chapter to these interactions. Contextually appropriate reactions to the accumulation or non-accumulation of Wrongness Flags or Cumulative Malpractice Flags form some part of the feedback. This is important as it means that the feedback that the user receives is in part self-initiated. There are clear successes and failures in most tasks and thus the mechanics of a Simulation or Serious Game within an epistemic scenario must accurately portray these successes and failures at appropriate junctions.

This, as a premise, is fairly straightforward. To return to our business like scenario the reward for successfully running a project will be the opportunity to run larger more complex projects in the future. This on behalf of the simulated company is an example of an increased demarcation of responsibility and thus, trust as well as a formal recognition of skills. Should the failure of a project be the responsibility of a single person's actions, in this case the participant's, the result is the opposite, potentially leading to that person's termination. This brings reward mechanisms within a Simulation or Serious Game in-line with the findings of Shaffer (2007a; 2007b) and the recommendations of Habgood and Ainsworth (2011), and Ulicsak and Wright (2010) as discussed previously.

2.1.1. How the Wrongness Framework differs from existing design frameworks

Each framework for the design of Simulations and Serious Games typically focuses on addressing a single issue that has been identified with either participants, other stakeholders, or the Simulation/Serious Game itself. It would, therefore, be most profitable to compare the Wrongness Framework to other frameworks or models, which most closely match its design focus.

Comparable frameworks are frameworks that are focused generally on improving levels of flow and engagement or those focused specifically on feedback or assessment to improve engagement.

One fundamental difference between the Wrongness Framework and other comparable frameworks is that the analysis of interaction is typically only in one direction – that of participant to game. Connolly et al (2012: 673 - 686) identified a number of studies focused on the acquisition of skills using games-based learning, of which a design framework is the outcome of only two of those studies. Both of these frameworks, developed by Connolly et al (2008), and Lucas and Sherry (2004), focus on the interactions of the participant with a game and describe how in-game alterations should be made to accommodate those interactions. This, they argue, can be used to improve participant engagement. The Wrongness Framework, however, reverses the direction of interaction and addresses how a Simulation or Serious Game should interact with a participant based on the input of that participant. This puts the Wrongness Framework in a research focus that is much more closely aligned with the advances of Intelligent Tutoring Systems in the last decade than that of Serious Game or Simulation research.

In the design for Intelligent Tutoring Systems some attention has been given to facilitating what is referred to as "metacognitive feedback". Metacognitive feedback in Intelligent Tutoring Systems is feedback which is given to the participant so as to encourage them to make use of features of the system to reach a more complete answer to a given problem (Roll et al, 2011). The most important aspect of this feedback is that it should in no way relate to the actual problem itself, rather, it should relate to the performance of the participant in their attempt to solve the problem. For example, suppose a participant was given a simple algebra problem to find the value of some variable and then that participant's answer was assessed as being incorrect. An example of metacognitive feedback would be to inform the participant that their answer was incorrect and that they should consider using the answer hint system or that they should take time to review their steps and see if they can spot where they made their mistake.

This is considered by Roll et al (2011), Aleven et al (2006), and Santosh and Koedinger (2005) to be a preferable approach to correcting participant interaction than simply to give only the bare bones assessment of the participant's answer and identify the key moment a mistake was made. Metacognitive feedback, however, is more complex than providing guidance feedback to the participant as it is necessary for the Intelligent Tutoring System to know when to provide this feedback. A number of models have since been proposed and developed to make this assessment of when metacognitive feedback should occur, however as Santosh and Koedinger (2005) describe,this problem has yet to be resolved within an optimal setting.

The two models which are most closely comparable with the design of the Wrongness Framework are the Help-Seeking Model (Aleven et al, 2006: 107) and Intelligent Novice Model (Santosh and Koedinger, 2005: 260). Both of these models provide the means to predict where a participant may be in solving a solution based on a number of system triggers. Depending on which of these triggers are fired the Intelligent Tutoring System can then determine using the default reasoning of the model the position of the participant in their attempt to answer the problem. Once this position has been deduced an appropriate piece of metacognitive feedback can be presented to the participant. This is possible because both the Help-Seeking Model and Intelligent Novice Model abstract the logical progression of problem solving and participant interaction that the Intelligent Tutoring System can make the necessary assessments of participant positioning. More complex applications of these types of models account not only for direct input from the user and their pattern of input but also judgements of that pattern in time.

The Wrongness Framework seeks not only to expand on these principles, but to do so from the perspective of Serious Games and Simulations. To appreciate the challenge of this situation is to appreciate the differences between the approach to problem solving presented in Intelligent Tutoring Systems, and the approach presented in Simulations and Serious Games. In Intelligent Tutoring Systems problem solving often mirrors a typical examination.. A question is given, perhaps with some real-world context, and the candidate is to provide some solution in a way that is appropriate for the domain of the problem for example, a numerical answer for an algebra problem, an algorithm for computer science, etc.

In a Simulation or Serious Game the problem is not, typically, presented in such a straightforward manner so the solution cannot usually be provided with a single direct input. Instead, feedback must come from either the environment, non-player characters, or from some supervisor presence.

If indeed, as the aforementioned researchers believe, that metacognitive feedback provides feedback in a way that is most appropriate to improve engagement and, indeed, flow in Intelligent Tutoring Systems then this is a matter of solving the *what* and *when* of feedback. In bringing the concept over to the realm of Serious Games and Simulations the Wrongness Framework is responsible for not only including those factors of feedback (what and when) but also addressing the question of how that feedback should be given.

This is pertinent as it strikes at the heart of the differences between an Intelligent Tutoring System, and Serious Games and Simulation design. Feedback by pop-up message may be an acceptable form of communication for an Intelligent Tutoring System but when creating a learning environment in a Serious Game or Simulation feedback can come from a multitude of different sources. This means that feedback offered to the participant must comply with the epistemic principles laid down by Shaffer (2007) or the experiential model of interaction of Killi (2005) with feedback being direct or indirect. Ultimately this increases the complexity of managing feedback responses far beyond the methods used in the models proposed for metacognitive feedback and need to be accommodated, which the Wrongness Framework does. The Wrongness Framework, therefore, is a coming together of various artificial intelligence techniques to bridge the design gap between Intelligent Tutoring Systems, and Serious Games and Simulations. In order to bridge this gap, the Wrongness Framework expands the complexity of feedback and assessment, without compromising the positive attributes of the known techniques.

2.2. Example: Contrapuntal Serious Game

To illustrate just how the Wrongness Framework may be used in abstract terms in a Serious Game/Simulation it is easiest to put forth an example from the arts. In theoretical music there is a term known as counterpoint. Counterpoint, in brief, is the process of composing a number of melodies that would be played together at the same time. In an example of good counterpoint these melodies are distinct from each other, such that neither is subservient to the other, yet both are structured in a way which support each other and are considered pleasing to the ear. To learn how to compose contrapuntally (in the style of counterpoint) takes some training and practice.

Handily, there have been some rules developed and codified over the last few centuries to assist composers in how to write in this style. Luckier still is that these rules have been written in a style of what not to do, rather than what to do.

In the simplest style of counterpoint, two melodies both rhythmically identical, identifying whether or not one of these rules have broken is a simple piece of subtraction and evaluation.

The entire basis of counterpoint is the comparison of the measurement of distance (an interval) between the two melodies in one instance and the measurement of distance between the two melodies in the next instance. So if we give each note of a single octave of a musical scale an identifying number then it is possible for the game to identify what type of interval is each interval in the sequence of intervals that is counterpoint. So, assuming the low notes are identified as 'b' notes and high notes are identified as 'a' notes, we can take any a minus b and evaluate the results.

Where δ_x is the interval and $T(\delta_x)$ is the identification of the interval type; either Perfect, Dissonant, or Imperfect.

$$\exists \delta_x : \delta_x = a_x - b_x$$

$$T(\delta_x) = \begin{Bmatrix} \delta_x = 7 \lor \delta_x = 4 \lor \delta_x = 0 \rightarrow Perfect \\ \delta_x = 6 \lor \delta_x = 3 \lor \delta_x = 1 \rightarrow Dissonant \\ \delta_x = 5 \lor \delta_x = 2 \rightarrow Imperfect \end{Bmatrix}$$

Secondly to assess whether or not a rule has been broken it is necessary to compare not only the type of intervals but also the method of motion used to proceed between each type of interval. This too can be determined mathematically. Where μ is motion and $T(\mu)$ is the type of motion, either Direct, Contrary, Oblique, or No motion:

$$\forall \mu = (a_x - a_y) \land (b_x - b_y) \exists T(\mu):$$

$$\left((a_x - a_y) < 0 \land (b_x - b_y) < 0\right) \lor \left((a_x - a_y) > 0 \land (b_x - b_y) > 0\right) \rightarrow T(\mu) = Direct$$

$$\left((a_x - a_y) < 0 \land (b_x - b_y) > 0\right) \lor \left((a_x - a_y) > 0 \land (b_x - b_y) < 0\right) \rightarrow T(\mu) = Contrary$$

$$\begin{Bmatrix} (a_x - a_y) = 0 \land (b_x - b_y) > 0; \\ (a_x - a_y) = 0 \land (b_x - b_y) < 0; \\ (a_x - a_y) > 0 \land (b_x - b_y) = 0; \\ (a_x - a_y) < 0 \land (b_x - b_y) = 0; \end{Bmatrix} \rightarrow T(\mu) = Oblique$$

$$\left((a_x - a_y) = 0 \land (b_x - b_y) = 0\right) \rightarrow T(\mu) = No\, motion$$

There are now enough established definitions for our game that we can create our first Wrongness Flag. The guiding rule of counterpoint, albeit there are others, is that the composer should not write two perfect intervals, one preceding the other, when the motion between them is direct. We are thus, left with the means for the framework to judge if a Wrongness Flag (WF) has been activated in this case.

$$\exists \delta_x \wedge \delta_y \exists \mu(\delta_x, \delta_y) : Perfect(\delta_x) \wedge Perfect(\delta_y) \wedge \mu(\delta_x, \delta_y) \; = \; Direct \rightarrow WF = 1\,;$$

Using a similar method it is possible to specify the logical criteria by which the Wrongness Framework can identify when a Wrongness Flag has been triggered. What, then, of Cumulative Malpractice Flags? In this example music, and the act of composing music specifically, has, like many other disciplines, developed a number of guidelines and theoretical aides to assist in the task of composing. Unlike, for example, the rules of first species counterpoint these are less of a case of “do not do” and more a case of “it's best if you avoided doing this”.

To give an example of such a guideline, few instruments, save the possible exception of keyboard instruments, look preferably upon melodic leaps of an interval greater than an octave. Consequently, including such leaps in any great number is likely to exhaust and annoy the performer of the work. So whilst an outright ban of melodic octave leaps would be unwise, and in fact would relegate famous melodies such as “Somewhere Over the Rainbow” to being incorrect, their use can be monitored using the same mathematical methods as before in the Serious Game, and overuse will result in appropriate disapproving feedback.

In this way triggering Wrongness Flags and Cumulative Malpractice Flags result in providing the input for the feedback and assessment mechanisms in the Wrongness Framework. Albeit it a operating within a different method of user initiated input, this approach is similar to the workings of the Evidence Interpretation Engine of the ALIGN architecture (Pierce et al, 2008) discussed in the previous chapter. To this extent it should be understood that if Wrongness Flags provide the means by which practice may be judged, Cumulative Malpractice Flags provide the means by which the same practice might be refined, and it is a refinement which can be made to a specified finite degree.

2.3. The need for Intelligence in the Wrongness Framework

Yet still our musical example provides, so far, nothing beyond a simple program which acts as a kind of musical “spell-checker”. It knows, after analysis, whether or not a mistake has been made and in some regards to what degree that mistake takes place in terms of seriousness of transgression. The real intelligence of the Wrongness Framework is found in the way that the Framework takes the information of the kinds of mistakes and their degrees, and actively alters the Serious Game state in what is presented to the participant on screen, what actions are denied them, and the means by which the participant may be informed of the kind of mistake they have made.

It is important, as discussed before, that this feedback through assessment takes place in an epistemic manner and in as an immersive way as possible. This method of providing feedback by altering user experience, environmental variables, and even through the agency of in-game characters is pivotal to the workings of the Wrongness Framework. In doing so, the risk that the experiential pursuits of the participant will be compromised by external factors, should be reduced.

There are then only two major ways in which the Wrongness Framework can alter the game state to provide the kind of feedback necessitated by the above stipulations. These are that the intelligence of the Wrongness Framework alters the environment itself as an entity, or, in a more direct way, that some regulating agent present in the game is itself altered – a virtual tutor, or mentor. In reality it is more likely that any Serious Game or Simulation making use of the Wrongness Framework will do so as some mixture of both of these options.

By way of example, if we return to our contrapuntal Serious Game we can now apply some means of guiding and structuring the participant's experience. Allowing that, in first species counterpoint at least, the first interval must be perfect and the last interval of the sequence must also be perfect, albeit specifically an octave, it is perfectly reasonable to suggest that any sequence based on a pre-existing cantus firmus – base melody - might be generated automatically in the same way that in other virtual games a non-player character may have their path-finding generated by some algorithmic process.

These generated sequences would then form the basis of some knowledge bank of acceptable answers in a similar way to the knowledge bank system of Thomas's and Young's (2010) Annie system. From this point, any participant input into the contrapuntal Serious Game is, as it were, overlaid on top of an appropriate existing template. The analysis for assessment will be dependant entirely on how different the participant's input is from the generated template. Even more useful is that, given the incremental nature of the task, it is perfectly possible that a best fit scenario can be employed and that the template will change as necessary to best fit the intentions of the participant. This then affords a much more dynamic assessment mechanism which will facilitate a greater number of permutations and thus allow for much more creative input.

As for the matter of actual feedback this can be achieved in a number of ways. It is pertinent to focus on what might be described as “negative” feedback or criticism although the connotations of that term bely its usefulness. In our contrapuntal Serious Game it might be that the participant is locked out of the rest of the sequence until they have, note by note, entered an appropriate input much like completing a maze or puzzle-sequence. Here the environment is informing the participant that what they have provided by way of their answer so far is incorrect but unless some additional form of feedback is offered on the specifics of where they have gone wrong the participant cannot advance. In part this will be remedied by the complexity of sequences the participant will engage with and the presentation of this complexity; viz. Aldrich's slates of educational experience. Still further this is the matter of through what agency the participant was first taught and maintaining that agency of the instructor. Assuming, as we must for the Wrongness Framework, that the instructor is an internal source then demonstration, and thus remonstration, must come in the same clear manner.

Conversely, rewards, at least from an educational standpoint, must come from the same epistemic root as the identification of suitable rewards are entirely context specific. In a business Simulation of some kind it would be reasonable to found a reward system where success in some project for a participant will mean that they are gifted with greater responsibility and increased opportunities to work on more extensive and complex projects. Similarly, failing to succeed, perhaps repeatedly, will ensure that responsibilities and opportunities are removed.

One might think of job titles, in such scenarios, as badges of achievement but without the ability to alter one's environment through them they become arbitrary serving only as a form of place-holder. In our contrapuntal Serious Game the rewards of undertaking the struggle to learn counterpoint is that the participant, in time, will be able to do more interesting things musically. Here the Wrongness Framework would unlock, for the participant, new exploratory powers and new tools to in order that they should be able to make more interesting music and experience new sound sensations.

What should be becoming clear is that the Wrongness Framework itself must occupy the role of the instructor in Simulations and Serious Games designed with it. In this way, it would seem, that the Wrongness Framework is no different to the other Intelligent Tutor Systems mentioned. This is true in general terms at least however, the Wrongness Framework must go further in the refinements of acting autonomously. It must educate, it must correct, reform, analyse the participant's interaction and govern the guidelines of the participant's experience. What is more it must do so in a way which is contextually appropriate and from within the Simulation/Serious Game environment itself in a way which cannot be seen as external nor alien. This is itself no small feat and it will require the structure of the environment of the Simulation/Serious Game to bend itself to the will of the Wrongness Framework. From this adaptive position the Simulation or Serious Game should thus mould itself dynamically to the participant's own needs.

Of course there are difficulties to be found. Giving the Wrongness Framework the kind of presence it requires in order to achieve this will require no small amount of designed intelligence – which, in and of itself, has its own problems. The relativistic nature of this kind of integration will mean that Simulation scenarios themselves will have be either finely modular in design, which then becomes a matter of managing flow, or less ambitious. Neither of these difficulties are unexpected or, indeed, unmanageable, however, there must be some considerable work on the design of an adaptive expert system to accommodate these needs. The following description provides an abstracted top-level view of the Wrongness Framework schematic and how these goals may be achieved.

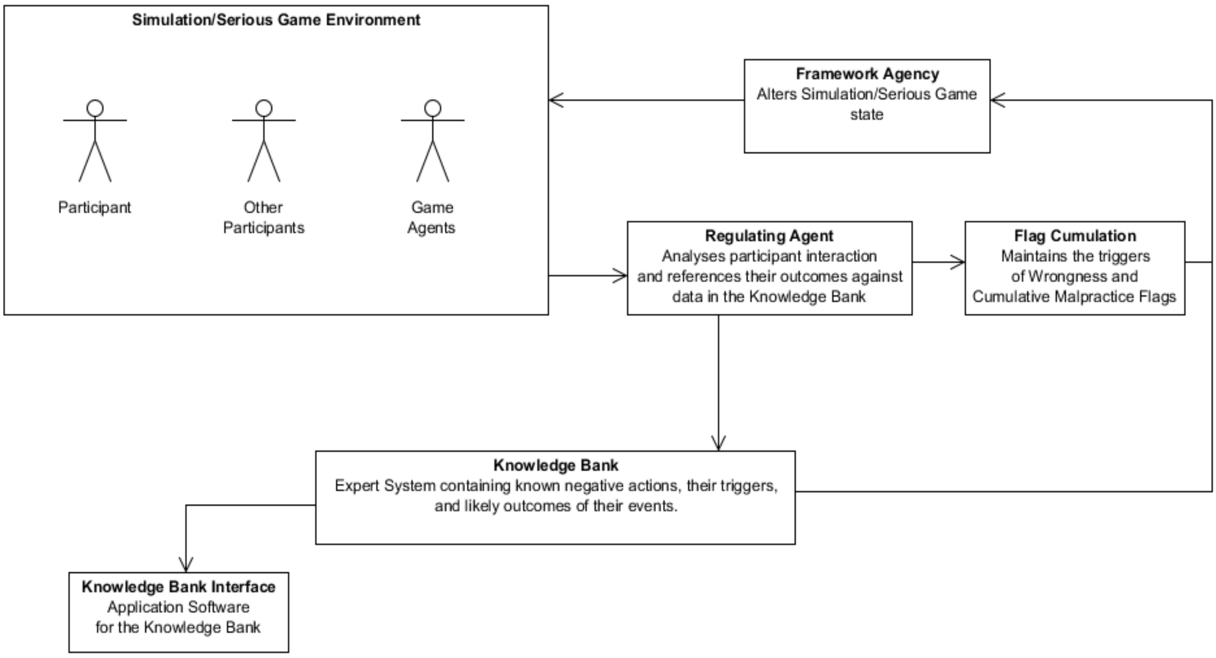


Fig. 2.1., *Wrongness Framework Overview*

In the diagram above (Fig. 2.1.) it is possible to see all of the method of mechanics in the Wrongness Framework. The Simulation/Serious Game Environment is assumed to be a fairly typical collection of the participant, potential multiple other participants, and game agents, such as non-player characters and environmental variables. By far the most important aspect is, as has been described previously, the Knowledge Bank. Within the Knowledge Bank a mass of data regarding the behavioural, subject specific, and challenge based knowledges exist as well as their respective relations and their effects on the game experiences – including any Wrongness or Cumulative Malpractice Flag values. To keep this Knowledge Bank adaptable it is thus, desirable, to accommodate alterations. So that this might be achieved an Interface to the Knowledge Bank is included, however its remit for change should not be permissible during a live Simulation/Serious Game experience.

The Regulating Agent is the analysis module for the Wrongness Framework. Taking input from the participant's actions and referencing it with the Knowledge Bank it is possible for the Regulating Agent to pass triggered values to the Flag Cumulation module which, using a limiting procedure, can determine severity of the participant's negative actions.

The Flag Cumulation and Knowledge Bank, together, form a pool of data from which the Framework Agency can siphon values for its state alteration variables. As one might imagine from the terminology, the Framework Agency – which itself alters the Simulation/Serious Game environment directly, is a complex Finite State Machine. To describe it as a singular Finite State Machine is perhaps a disservice as, by necessity, other game agents may require integrated Finite State Machines themselves. To this extent it is better to imagine it as a mesh rather than as a single entity.

What is clear from this overview is that the visual aspects and even the specific interaction methods of the participant with environment and environment with participant are of little importance to the mechanisms of the Wrongness Framework. The Wrongness Framework requires only that the specific interaction methods and presentation of information for the participant be consistent throughout the entire life of the Simulation/Serious Game experience.

Returning for the last time to our contrapuntal Serious Game it is possible to see the underlying structure of the Wrongness Framework at work. The manner in which the participant interacts with the game is itself unimportant to the mechanism, but with consistent forms of interactions participant behaviours can be assessed via the templates generated by the Knowledge Bank and the Regulating Agent and advances or decrements to experience can be meted out via the Framework Agency. This is not, of course, to suggest that the visual is unimportant to the experience – indeed so much of the matters of engagement and flow as discussed in the previous chapter are reliant on this as a necessity - but rather it is not a contributing factor in the Wrongness Framework. Consequently it is perfectly plausible to apply the Wrongness Framework to Simulations/Serious Games which, through technical or socio-political reasons, might be devoid of specific visual cues.

2.4. Thou Shalt Not: Developing the Knowledge Bank

Developing the Knowledge Bank for the Wrongness Framework is, in many regards, a very similar process to the creation of any generic expert system. As a rules-based system the Knowledge Bank must have a myriad of minor axiomatic absolutes and the default reasoning mechanisms to follow. For computer scientists such as Luger (2005), and Russell and Norvig (2010), the means by which these foundations are laid can be said to be codified into the following guidelines:

- That the database of knowledge be assembled semantically in a standard if...then structure. This is necessary as, in most applications of the Wrongness Framework, a rule which requires anything other than symbolic reasoning will be restricted severely to the technical limitations of sensory stimuli in artificial intelligence and the platform on which the Simulation/Serious Game exists at the time.

- That the database of knowledge be separate from the inference mechanics; hence the Wrongness Framework's separate Knowledge Bank, Regulating Agent, Framework Agency modules.

- That, wherever possible and by means of establishing precedent - particularly in expert rules-based systems - a number of logical cases which may already exist be present in the knowledge database.

- That, wherever possible, rules should be defined unambiguously as possible, even though a specific case may need to call upon a number of rules in succession.

- If this is not possible, for example within a specified frame of reference, that a default reasoning mechanism be in place so that decisions might still be made.

The Wrongness Framework, as a framework for Simulations and Serious Games rather than any other kind of virtual environment, is fortunate insomuch as it is not required to make any positive diagnosis in the same way that, for example, a medical expert rule-based system might need to. This is due, of course, to the correctional nature of the educational process and the manner of structuring learning via what are essentially goal points. As a consequence of this the Knowledge Bank itself can be much smaller – as it need only concern itself with what is wrong - and, indeed, more effort can be made toward making it more complex and precise.

So far, the discussion surrounding the application of the Wrongness Framework has been limited somewhat to the computational aspects of knowledge gathering and interpretation but has not actively investigated the way in which people gather and make use of knowledge. In his guidelines for the design of expert systems, Luger (2005: 284) himself provides a clear example of how a person may think when confronted with a problem.

In the example, in which a person is attempting to balance on a unicycle, Luger suggests that the person at that point does not keep their balance by calculating any number of equations at tremendous speed in an active way but, rather, using sensations of gravitational pull and momentum to gather information they can make informed decisions on how to keep their balance. In an application of a knowledge database this approach might take the form of understanding typical semantic and syntactic habits of a typical user and, if using a search function, suggest results based on a particular phrasing of the search term. In a Serious Game/Simulation using the Wrongness Framework, this kind of feedback to the user must be either audio or visual or some combination of both. Here weak AI mechanics must inform strong AI mechanics and vice-versa in a smooth cyclical loop continually feeding back to the participant in a variety of ways and in ways which they govern.

Because of the cyclical nature of the Knowledge Bank – Game Environment feedback loop, the epistemic approach to Serious Game design is reaffirmed as the most apt choice. As continuous feedback needs clearly defined parameters and epistemic design provides a definitive reward and penalty system which can be factored into the expert system more easily than any other.

In defining the rules of our expert system there are a few categories of content which form the substance of each rule. Firstly, there is the rule itself composed within a symbolic language. In addition to which there should be the means associated with every rule to acquire the input from the participant in order to test if its state is true or false based on that input. Secondly that its associated flags be established. These will include either Wrongness Flags and/or Cumulative Malpractice Flags and their respective weightings and triggers. It might also include any further knowledge map flags which are to be used in the default reasoning of the system. Lastly that the rule, should it be broken, has a direct or associated consequence which can be triggered in the game state. It should be noted therefore, that though each of these respective rule components may be housed in different modules of the system, it is necessary that during the design of the system they are written in complete form.

The consequence connections of these rules are in themselves complex as, unlike examples such as the prior contrapuntal game, not all scenarios are going to lend themselves so easily to mathematical conversion. This is particularly true in terms of human behaviour and though there is continuing research and understanding in this area there is, more often than not, going to be a need for balancing the complexity of consequence and inference from the Knowledge Bank and the detail of a Simulation/Serious Game scenario. Part of the goal of the Wrongness Framework in Simulations and Serious Game is to increase the range of possibilities available and extend the limits of this balance.

2.4.1. Constructing Rules

The means by which rules may be constructed must adhere to the reasoning mechanisms as described above but must also contain a number of necessary components. These are as follows:

- Conditional inputs and identification behaviours for input sources.
- Consequence states for those conditions.
- (Knowledge) Map Identities.
- Map triggers.
- Weightings.
- Default Reasoning Processes
- Weighted Wrongness and Cumulative Malpractice Flags

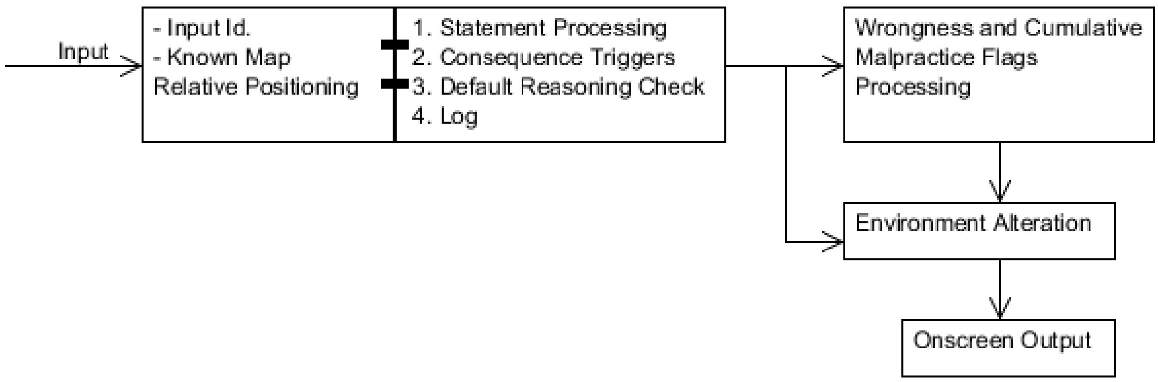


Fig. 2.2., *Rules-Consequence Abstraction*

The above diagram (Fig. 2.2.) shows an abstracted view of the interaction behaviour sequence of rules through these components on a wider scale with other parts of the framework. In this sequence the relationship between the identification of input sources and the triggering of some on-screen response is shown as a process of deriving the correct rules statements from the knowledge map. This rules-based derivation then affects environmental alteration states within the Framework Agency component (as shown in Fig. 2.1.).

In this example the input comes directly from participant engagement with the Simulation.

It is reasonable to question that a premise such as the Wrongness Framework and its concentration on regulating negative interactions within a Simulation/Serious Game might be susceptible to moments in the environment where no reasoning can be offered. This is a consequence of a less complex and more open-ended simulation experience. Whilst that which is not wrong may be given a free pass, by implication the framework must be certain of when to trigger a consequence and when not to.

The Wrongness and Cumulative Malpractice Flags go so far in ensuring that consequences of an appropriate weight are triggered but they, in and of themselves, cannot guarantee such measures. For that a limiting process is introduced which provides some means of safeguarding against unfair applications of negative consequences. Simply it is the means for the framework to recognize formally that it has no means in its knowledge bank to correctly identify a participant's input as an unapproved method of tackling the scenario of the Simulation. Additionally this limiting process can also be used as a beginning basis of increased adaptivity in a Simulation and, with other appropriate measures, a shift towards non-deterministic scenarios might be achieved.

2.5. Challenges for the Wrongness Framework

The Wrongness Framework is a proposal to implement a less complex yet explicit rules-based expert system to alter the states of a virtual learning landscape of a Simulation/Serious Game in an epistemic way such to improve the immersive quality, engagement, and learning qualities of the Simulation/Serious Game. Fundamentally changes brought on by the Wrongness Framework in the background mechanics of the Simulation/Serious Game have direct influence on what is to appear on-screen as a means of communication with the participant. However it is important to make clear that these on-screen changes are not the focus of this study but rather what is the focus is providing the mechanisms of the framework to allow these changes to happen.

For the Wrongness Framework to have relevant application it is necessary for a number of challenges to be fulfilled. These are as following:

A) That possible behaviours of a scenario represented in the epistemic environment of the Simulation/Serious Game can be codified into Wrongness and Cumulative Malpractice Flags.

B) That the general components of behaviours, including their Wrongness and/or Cumulative Malpractice Flags, can be expressed in adequately in terms symbolic logic and with relative weightings.

C) That it can be demonstrated that the Wrongness Framework can increase scenario possibilities of a Simulation/Serious Game rather than restrict them.

D) That it can be demonstrated that the Wrongness Framework provides the means for integrating necessary Simulation/Serious Game processes as detailed within the four slates of educational experience into a single in-world experience.

If these challenges can be met it will demonstrate the validity of the Wrongness Framework as a means of increasing what is capable in Simulations and Serious Games and demonstrate the associated powers of other features of immersion, engagement, and achievement into one means. What is, therefore, required now is the assessment of a number of case studies making use of the Wrongness Framework to test if these challenges can be met, how effectively if at all the challenges are met, and the change in design implications demonstrated by its utilization.

* * *

Chapter Three:
A Methodology for the Wrongness Framework

The last section of the previous chapter identifies the challenges that the Wrongness Framework must consider and overcome. These challenges were based on the findings of previous scholars, their conclusions, and the gap in knowledge that they have identified. The Wrongness Framework as it is described in this section is, therefore, a reaction to these challenges and a consideration of changing implications for Serious Game and Simulation design. It is the goal of this chapter to clarify the means by which the Wrongness Framework has been used to develop Serious Games and Simulations. Additionally it is necessary to identify how the effectiveness of those games in facilitating the key factors of knowledge acquisition, immersion, and motivation has been achieved. Ultimately this is an exercise in attempting to answer the question of whether the Wrongness Framework is capable of providing a framework for in-game assessment and feedback with a reduced need for external validation from a real-life tutor presence.

3.1. What must be assessed?

Before the minutiae of details regarding the methodology for the Wrongness Framework can be discussed it is necessary first to establish what has been assessed. This is due to the range of impact that the Wrongness Framework may have in the field of intelligences in Simulations and Serious Games and this clarification provides a reining in of research focus. This research focus has been addressed in a single question:

Can the Wrongness Framework provide feedback and assessment mechanisms in-game in a way that improves participant immersion, which might otherwise have been interrupted by a real-world tutor? The challenges to the Wrongness Framework, as addressed at the end of the last chapter (p. 43), have provided the benchmarks for the design implementations of the Wrongness Framework in addressing this question.

Firstly it was necessary to establish that the expert system of the Wrongness Framework was fit for purpose and was operating accurately within the context of the Simulation. This means ensuring that the Wrongness Framework was complete for the Simulation context, which is to say that it had accounted for the expert knowledge required by the scenario and was able to make direct or environmental judgements against this expert knowledge.

Challenges addressed: A and B

Secondly that a Point of Enquiry (PoE) has been established within the Simulation/Serious Game which the participants were able to use to assist them in their in-game operations. The PoE will have been one of any kind provided that it conforms to the epistemic requirements of the Simulation scenario insomuch as it, too, does not break immersion or flow (Bouchet et al, 2013).

Challenges addressed: D

Lastly that, through the mechanisms of the Wrongness Framework in place within a Simulation/Serious Game, that participants have been able to demonstrate progression in a clear fashion and that that progression has been used to tailor its complexities to the needs of the participant by way of maintaining the cognitive curiosity of that participant through game flow (Bouchet et al, 2013).

Challenges addressed: C

These design considerations have been made by using the findings of the first chapter. The considerations therefore have been formulated with some regard to the epistemological contexts of the scenario, the four slates of the educational experience, and all other surrounding literature which has been used to dictate the content of any simulation scenario. This has been important as it was the scenario's context itself that has determined the means by which the assessment or feedback was provided. This has also been the case for the interactions of the participants themselves (Charles et al, 2011).

3.2. The application of the Wrongness Framework

As the *what* of the Wrongness Framework has been determined, which is to say a design framework for providing feedback and assessment mechanisms in Simulations and Serious Games, this section details the approaches and procedures that have been used to achieve the *what*. The process for each of the case studies has been developed to be as straightforward as possible. So as to best demonstrate the process that has been used, this section of the chapter has detailed firstly the top-level overall processes of the methodology used in the case studies before going into further detail in specific instances of those processes.

It must first be understood that this research has been undertaken from a Western or Euro-centric point of view. It is important to understand this position as the types of qualitative data and the means of its collection from the case studies, regardless of whether this is by the interviewing of individual experts or the in-game interactions of participants, has been analysed from the perspective of that Western/Euro-centric view. To illustrate what is meant by this, the case studies may have, for example, focused only on Western pedagogical techniques or team practices as a matter of familiarity rather than through active selection. It is therefore to be understood that this review approach may have constrained the kinds of qualitative data that have been collected and this bias should be considered as a limitation of each case study.

No particular consideration has been given, beyond the context of the subject of the Simulations/Serious Games in the case studies, for addressing the needs or best approaches of specific groups who may well make up the contingent of participants. Similarly no particular consideration has been given to addressing a specific presentation style outside of the context of the subject of the case study. Rather the foundations for each of the case studies have been made to mirror the more general understanding of Simulation-based learning as described in the first chapter.

The omission of these special considerations has not been due to a disregard for their importance, but rather as a desire to accomplish two goals: firstly, to keep the assessment of Wrongness Framework as free as possible from the biases of special interest, and secondly, that one accepts that it may not be possible to consider the approaches and requirements of a truly universal application. Consequently, no speculations have been made on results taken from participants which might fall into the remit of special consideration. It is admitted that these factors might have some effect but that their detection and analysis will fall outside of the scope of the initial research of these case studies (Pierce et al, 2008).

Of the two major paradigms of methodologies, qualitative and quantitative, the former interpretive methodology has been adopted as being most suitable for the research undertaken in the application of the Wrongness Framework. It has been considered that a qualitative approach lends itself most to the description of personal experience and variation amongst that experience that can be expected when investigating the issue of immersion, the effectiveness of feedback and assessment, and the learning achievements of participants across multiple case studies (Walonoski and Heffernan, 2006). Though some quantitative analysis has been undertaken within the case studies, these findings will serve to assist further dissection of details from participant review as in line with previous methodologies employed in Intelligent Tutoring Systems rather than to formulate direct conclusions (Mitrovic et al, 2007; Charles et al 2011). This quantitative data, however, has been used to assist in any mathematical modelling for refinement of the Wrongness Framework. These refinements have been provided so as to provide insights into the use of frameworks such as the Wrongness Framework for affecting the algorithmic and mathematical stimuli in the design of Simulations/Serious Games.

3.3. Activities of application

The diagram below describes sequentially the activities that have been used in establishing the foundations of a case study for a Simulation/Serious Game using the Wrongness Framework. These activities take place in two phases, the pre-Simulation and post-Simulation. The activities of quantitative data gathering from the game experience is described between the pre- and post-Simulation phases.

3.3.1. Pre-Simulation

1. Subject Expert Interviews

So as to ensure that the expert systems of the Simulations used for assessment are capable of accurately performing the tasks of altering the virtual learning environment, assessment, and providing consistent rulings from the PoE in game tutor, it has been necessary to undertake expert interviews as an activity recommended by Charles et al, 2011.

These interviews have provided the theoretical basis for specific game-play mechanisms such as scenario progression and the prioritisation of action-based involvement from the participant. They have also been used to define the legislative, simulated, and organizational operations for the feedback and assessment mechanisms of the Wrongness Framework.

The expert interviews have been conducted with the assistance of subject experts in the scenario context for each of the case studies (Steenbergen-Hu and Cooper, 2014). A number of point of contact experts were consulted, around 8 – 12 for each Simulation scenario that used this process. These experts were interviewed regarding the typical interactions that should be present within a Simulation, the means for progression for the participant, and, most importantly of all, for the information required for determining the Wrongness and Cumulative Malpractice flags for the Simulation. This information was then used to design various components of the Wrongness Framework so as to provide for its participant assessment and the activation triggers for environmental changes within the Simulation. The evaluative process seen in Fig. 3.1. (shown below) which made use of the results of the expert interviews should properly be read from left-down to right-up.

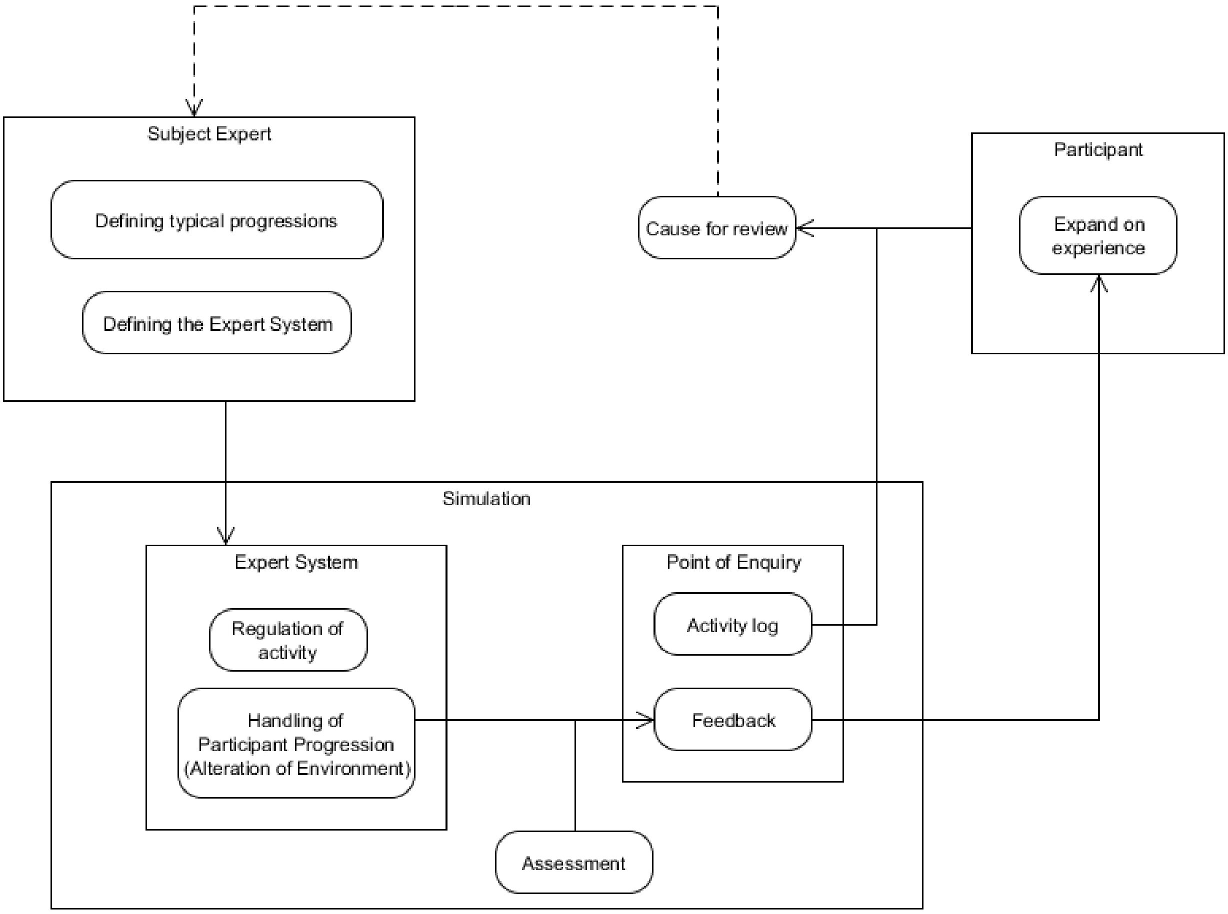


Fig. 3.1., *Evaluative Process*

3.3.2. Post-Simulation

2. Surveying the participants

The Simulations of each case study that have had active participants have been undertaken with a large number of participants but were typically handled in 10 – 12 per group. Participants were within the age range of 18 years old and greater and should, according to the definition given in section 1.1 be considered *older*.

The participants of the Simulation experiences were surveyed for their perspectives on their achievements, their perceived levels of immersion, and their progress. Additionally participants were also asked to comment on their actions in light of the in-game Point of Enquiry assistance and feedback.

The survey of the participants has also taken into account the comfort levels of the participant regarding the technologies which they were using, noting issues they may have had during their interactions, as well as their general confidence regarding the subject knowledge of Simulation scenario (Bouchet et al, 2013). Each scenario took a relatively short length of time to complete, typically around 20 – 30 minutes. Some prior subject knowledge has been assumed for the participants in certain case studies. Where this has occurred it has been noted in the relevant chapters.

The survey of the participants was in the form of a formal questionnaire as well as seminar based feedback sessions directly after the Simulation. These sessions were used to provide the opportunity for participants to offer more detailed feedback so as to make use a variety of surveying techniques (Cen et al, 2006). Participants were not asked to submit some personal data regarding their specific age ranges, genders, and similar kinds of identifying information. Participants were also given the right to refuse to submit personal data if preferred.

After the Simulation event, the event logs of each participant were analysed for how often the participant made interactions with the PoE, their overall progression, the advancement of the game environment, and the typical interaction time for each event (San Pedro et al, 2011). This data was then used to provide an overall trend in addition to the qualitative assessments of the participant so as to ensure for validity that actions have matched what has been described in survey.

* * *

Chapter Four:
Applications using the Wrongness Framework: AdQuest

(Ethics Reference: 16758)

Within this chapter and the following chapter are presented two major case studies of the application of the Wrongness Framework. Each of these case studies is different in its scope, context, and purpose but each demonstrates how the simulation of procedural professional tasks might be designed using the Wrongness Framework to incorporate more complex factors within any Simulation scenario. This might include the use of environmental variables, interaction design, and other such factors. More important than the mere inclusion of these factors is the demonstration of how those factors can be managed from within the Wrongness Framework itself. Each of the case studies found in these chapters demonstrate the application of the Wrongness Framework as it is described Chapter 2 and in light of the challenges addressed in Chapter 3. However, each case study represents only a selection of components from within the model of the Wrongness Framework itself and the designs presented for those case studies should be understood within the particular remit of what is attempted to be demonstrated.

The first of these case studies is the small scenario of the Serious Game "AdQuest", designed and developed in partnership with the UIX designer Vanissa Wanick (University of Southampton). The following few paragraphs describe the development of AdQuest with respect to this partnership. This is in order to ensure that work can be correctly attributed to either Vanissa Wanick or to the author. For purposes of clarity the author of the thesis has referred to himself in the first person for the following passage for the description of the development of the AdQuest scenario.

The development of AdQuest took place over three stages. The first, the conceptual design phase, involved researching the 34 rules used in the AdQuest scenario based on the needs of the AdQuest case study. The division of this research was simple: rules concerned with direct assessment of advertisement design were researched by myself, and those specifically pertaining to Chinese advertising culture were researched by Vanissa.

Once these 34 rules were collected and reviewed from a number of sources I decided upon the quantitative metrics used in the AdQuest scenario and the means by which they might be collected.

The second stage of this development process began with designing the hypotheses to be tested and storyboarding the scenario. Storyboarding in this sense should be taken to mean a scene by scene design of each thing that a participant might see. It was then mutually decided, based on the rules we had each researched, that the AdQuest scenario should be for the design of luxury brand goods. After ethics clearance was obtained, Vanissa wrote the dialogue for the game describing the context for the scenario and obtained a Mandarin Chinese translation of that dialogue with the assistance of two other colleagues. During this second development period I was responsible for designing the assessment mechanisms for the scenario, deriving the logical parameters for those mechanisms from the rules researched in the first stage.

The third development stage was devoted to the final creation of the AdQuest scenario. Vanissa's responsibility for development was layout design and the creation of all the graphical assets, which included character design, product image sourcing, user interface components, etc. My role was to take these graphical assets and to program the layered AdQuest system beginning with simple interaction design with the UI elements to the creation of the rules-based knowledge bank of the AdQuest scenario; the default reasoning, the assessment weightings of each component, etc. This required the programming of the means of assessment into a single, coherent assessment module capable of assessing the submissions of the participant against results expected by the default reasoning. The intention of the game was to test the Wrongness Framework. The intention was not to be completely accurate with the business logic as this would have taken far more time than the scope of this PhD. As an illustration, it was perfectly useful.

4.1. An Overview of Game Play

AdQuest is a Serious Game in which participants play the part of a layout designer for a design agency that accepts project from perspective luxury goods clients. In the scenario taken for the first case study, the participant was asked to provide a layout design for an advertisement of Chanel No. 5 perfume specifically targeting the Chinese market and within a population demographic of 18 to 35 year olds. Once the participant's design was submitted it was then assessed and feedback was given in a brief performance review with the user's in-game project manager.

The aim for the participants was to design what they believe to be the best layout design for an advertisement of that specific brand within that market/demographic and that the outcome of such a campaign would be relayed to them in terms of a ratio of expected market penetration. The style of this feedback was chosen only for the purpose of maintaining an epistemic context and in no way represented a true indication of what appropriate marketing theories might have suggested as this was not the focus of the AdQuest scenario.

Fig. 4.1., *Title Screen, AdQuest English Version*

Great care was taken to ensure that the delivery of the design brief and the contexts within which it was framed fit the epistemic context that the users might expect and to address the development of AdQuest within the remit of Challenge A (as described in the previous chapter). For example, the PoE for the game was managed through the manager character (as illustrated above). It is through direct communication with this in-game character as PoE that the participant was informed of their design brief and what is known of the target market/demographic. Similarly, it was through this manager character that the participant is given feedback on their efforts and of their marketing campaign.

Other than standard navigation interaction within the games, which was handled in a linear way via simple button presses, the majority of the assessed participant interaction took place during the design stage of the game. In this stage participants were asked to create their layout design using a selection of pre-generated components. From within the context of the game it has been assumed that these components represent the best possible choice selections for each of the component ranges.

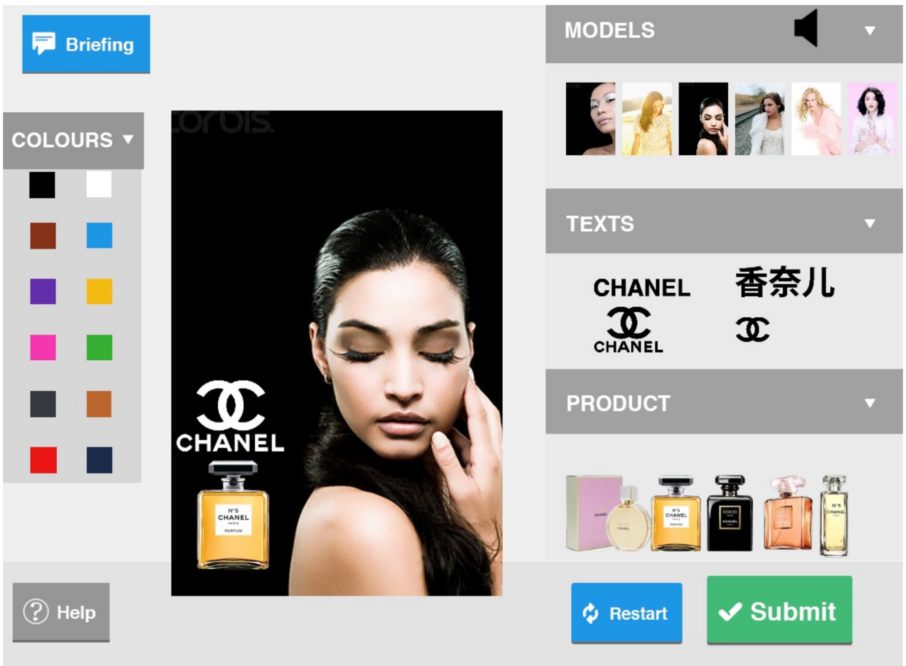


Fig. 4.2., *Design Screen, AdQuest English Version*

The layout designs the participants were asked to create contained only three major layout components – background image, product, and logo - which can be chosen, altered, and positioned as the participant sees fit. It is worthwhile noting at this point before getting into the mechanics of assessment that a number of unsuitable products were included as options purposefully to provide harsher cumulative malpractice flag weightings for assessment. In this instance their unsuitability is attributable to the fact that the advertisement brief is specifically for the Chanel No. 5 product and other products were carefully chosen from the Chanel range which might have attractive attributes for any layout design – such as colour, shape, etc. - yet nevertheless would be completely unsuitable from the perspective of the brief.

By clicking the help button the interaction commands used in AdQuest were displayed to participant. This was a somewhat different from the typical Intelligent Tutoring System activity of offering specific advice on what should be selected as this scenario was not designed to act as a guided assessment. However, the participant was able to refer at any time to the project brief by clicking the appropriate button which displayed, once again, what was expected and what was known about the market for which their design was destined.

The participant was free to choose to include or not include each of the components in any combination of logo, background, and product. However, once they were satisfied with their design the participant was to submit it for circulation. It was then presumed that the participant's design was then used in an in-game advertisement campaign. After the results of this campaign were analysed, a cover for the Wrongness Framework's assessment of the design, the participant was debriefed by the manager character on the success of their advertisement in terms of market share increase and their actions.

Fig. 4.3., *Reaction stills from campaign results, AdQuest English Version*

During this debrief feedback on a suggested course of action for the participant was given which may be used such that they might correct any errors they have made and the process for doing so. This feedback also detailed the satisfaction of the agency with the participant's work and the participant's position within that company; such as to whether they might vie for promotion or more challenging tasks in the event of a positive outcome. As with other elements, care was taken to ensure that the feedback was itself epistemic and that the inherent reward structures within the game matched what should be expected in industry.

However, it should also be noted that the feedback mechanisms for this game were much less well defined than the assessment mechanisms as there is no longer term track within the game which the participants may engage within. Consequently the feedback offered was specific only to this particular scenario and presumes the participant, in-game, to have some imagined pre-existing history of expectation in their role with the in-game company. This means that participants of the game were treated as having some prior knowledge of the design industry and that expectations of that presumed knowledge can be found within the assessment mechanisms.

4.2. Developing the Rules-based System for AdQuest

To begin, a number of rules – around thirty - were sourced from various professional texts and academic sources. The nature of these rules were split into three general camps: the general layout rules which would be applicable in most circumstances (Felten 1947; Landa 2010) and those which are specific to luxury brand advertisements (Checchinato et al, 2013), and those still further which were specific to the Chinese market (Hung et al, 2011).

It is worth attempting to frame the context for these rules properly as they were collected. Firstly, as with all topics of a more subjective nature it is difficult to treat these rules in the same way that one might for a subject with more locked-down definitions such as, for example, a legal framework for an industry. Rather, they are seen, by their own authors, more as guidelines which ought to be followed as a product of academic research or personal professional experience. Secondly, there is no escaping the fact that the during the development of AdQuest these texts were treated as authorities in their own right and, beyond a wider investigation of the topic of layout design for luxury goods in the Chinese market, the rules found within these texts are treated as being sound. As with all such systems though, the Rules-based System of the Wrongness Framework must presume that the information it is given is sound. Hence it is proper to refer to the Rules-based System for AdQuest as an Expert Rules-based System rather than as a standard Rules-based System.

Designing the mechanics by which the Rules-based System interacts within the game was slightly more complex. It was necessary at first to determine, via the selected three layout components, their relative dependencies and independence within the rules themselves. Each of the components have a number of relativistic attributes, such as size, colour, shape, position, etc., which, acting as weights, will alter in accordance to selections made both on themselves and on each of the other components.

To accomplish the management of these weightings a path of least change from a super component was developed between each of the components to determine their relativistic dependencies. Simply this is a matter of determining which of these components changes least and with the least dependence on information from other components. In this AdQuest scenario the role of super component was fulfilled by the background component as it has the smallest range of alterations possible in its optimization.

The use of a super component in a path of least change should, in this context, make sense. If each of the components is going to have some selection of attributes and those attributes are themselves weighted on their own internal optimization then there must be something by which that optimization is judged. For example, suppose a rule exists hypothetically within the system which says “logos for Chinese markets should always be fuchsia pink because that colour is culturally linked with luxury brands” then it would make sense that the colour of the logo should, in all cases, be fuchsia pink. In the event that no such rule exists then the suitability of a colour will relate to other colours found within that design. A white logo scores badly on a white background because of reasons of legibility but may score much higher against black backgrounds when used in conjunction with other appropriate schemes.

From the perspective of design the choice for managing this path of least change was clear: a weighted node system would be best. Primarily such a system would give the assessment enough flexibility so that multiple “correct” answers could be given for the participant's layout design. However, the real strength of such a system is that each individual choice of starting background, logo, and product image could still have a calculated optimal arrangement of attributes and consequently assessment against the rules could be performed rationally. In effect each attribute could be valued with some percentage against its optimized form. Assessment, therefore, would simply be a summation of these percentages as a differential.

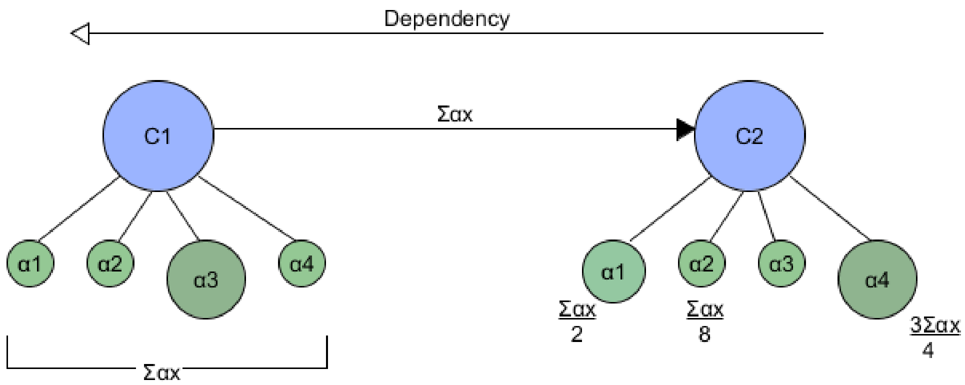


Fig. 4.4., *Abstraction of node dependencies with hypothetical weightings*

This approach, whilst reasonably satisfactory in management of Cumulative Malpractice flags, cannot account for Wrongness flags in a way which is fully representative of their importance against a perceived limit. To solve this issue it is necessary to provide the summation of Wrongness flags as a coefficient to the Cumulative Malpractice Flag summation. Thus the form of the weighted assessment mechanism is:

Let *W* be the set of weighted variables for a given assessment, *x* be each weighted variable in the assessment, *c* be the value of any associated Wrongness Flag with this assessment, *r* be the cardinality of the set *W*, and *d* be a importance weighting value for the entire assessment:

$W: W = \{x_1, x_2, x_3, ...\}$

$x: x \in W; 0 \leq x \leq 1$

$c: c \in \mathbb{R}; 0 \leq c \leq 1$

$r: r = |W|; 0 \leq r$

$d: d \in \mathbb{R}; 0 \leq d \leq x$

$$\frac{c}{r}\sum_{n=1}^{r-1}(x_n - d) \rightarrow y: \; 0 \leqslant y \leqslant 1$$

As the result of this calculation – the value of c being calculated in a similar fashion – provides a decimal between zero and one optimizations for components can be converted to a percentage in which assessment can take place in terms closer to what the average participant can readily understand. This also means that in areas, such as academia for example, where percentages are used and ranges of percentages in assessment have defined meanings then those definitions can still be used when otherwise non-simulated tasks are simulated in this way.

For the sake of optimization within the system and in light of the rules collected on layout then it is possible, with some prior knowledge, to make value judgements on the success of hypothetical changes to a component. For example, with a given background and application of the appropriate rules it is possible to map the most optimal locations of a product; assuming all other attributes of that product are correct.

What this means for long term development is that with a demonstration of previously acquired knowledge a participant will be automatically rewarded in such a way that a knowledge progression map might be predicted for that participant and refinements to weightings could be made to increase the challenge of a task. This is the epistemic principle of gameplay reward-penalty interaction as it was discussed previously in chapter one. In principle this is possible although, as was stated earlier, long term development plans were not incorporated into this scenario and natively AdQuest has no way of measuring and maintaining this track.

4.2.1. Translating findings with regards to identified layout rules into parameters for the AdQuest Knowledge Bank

The development of the rules-based system for the AdQuest scenario had a single goal. This goal was to translate the rules or guidelines of layout design as they had been researched in such a way that fit with the design principles of Wrongness Flags and Cumulative Malpractice Flags. In this way it was possible to offer some form of objective assessment mechanism for a task for which might otherwise have been previously subjectively assessed.

It is important to reiterate once again that the purpose of the AdQuest scenario was to develop and test assessment mechanisms in the Wrongness Framework and as such little attention was given to providing what might be described as accurate epistemic feedback. By epistemic feedback it should be understood that that would be feedback, which relates directly to the context of the scenario. Any such feedback was given to the participant only with the understanding that this was to be understood as giving context to the participant's interactions not as an accurate representation of their impact. This is most notable in the inclusion of a marketshare percentage given to the participant as a framing device for the scenario and, therefore, this marketshare percentage score should not be mistakenly believed to have been derived from some known business calculation rather, only as an epistemic reflection of the participant's submission against an expected result derived from the rules-based system.

The AdQuest scenario made use of three in-game layout components for assessment; the background, the logo, and the product image. Whilst the dependencies of these components has been expanded upon before it is necessary to also explain the data design that was created for those components. To begin each of the rules of the 34 were collected and studied for common variables that might be shared amongst them. For example, many of the layout rules proposed by Felten (1947) refer to the placement of components, their size, shape, and, in some circumstances, their colour qualities. From this basis it was possible to derive a number of simple variables that might be used when making rules-based assessments for each of the moveable components (logo and product); such as their position on the canvas in Cartesian coordinates, relative sizes, colour, specific image, etc. Once these simple variables had been assigned it was necessary to decide upon some default maps for sample layouts that could be used to formulate the programmed rules.

The means of development used was quite similar to constraint-based rules programming that has been used in Intelligent Tutoring Systems, as in the research of Mitrovic et al (2007) and Mitrovic and Suraweera (2000), and other rules-based systems. In this approach constraints are formulated which describe general commands and procedures in the event that a constraint is broken or adhered to and these commands are fulfilled in conjunction with other constraint commands. It should also be noted that this is not the same domain as that of Constraint Satisfaction Problems (Russell and Norvig, 2010) and shouldn't be confused as such.

The benefit of this approach is that allows the rules-based system be more dynamic in its ability to make judgements as there is no need to attempt to model a typical or expected behaviour for the participant. Despite its simplicity to do so requires the careful mapping of in-game variables and how their information might be used to formulate constraints.

In order to achieve this goal a map of an optimized layout for each of the possible backgrounds was created; a map in this instance simply refers to a defined set of values for each of the component variables. This map was created after a number of layouts were created and cross-referenced with the collected rules in consultation with Vanissa Wanick as a layout expert. Each of these layouts represent what was believed to be the ideal layout of components for each background. Once these ideal layouts were mapped the next task was to assign the constraints to each of the variables in accordance with the principles of the Wrongness and Cumulative Malpractice Flags.

For example, each background image has a background colour and for each background colour there might exist any number of suitable foreground colours for the logo component. Rather than map them all by what is most suitable it is a much simpler task to assign to values, which are less suitable than to those that are. So, if the background colour of a background image was black then it would be unsuitable if the logo colour variable was also black as it would mean that the logo wouldn't be visible. Similarly very dark shades colours would also be unsuitable for the same reason. These colours would therefore be given a Wrongness Flag value to signify their complete unsuitability.

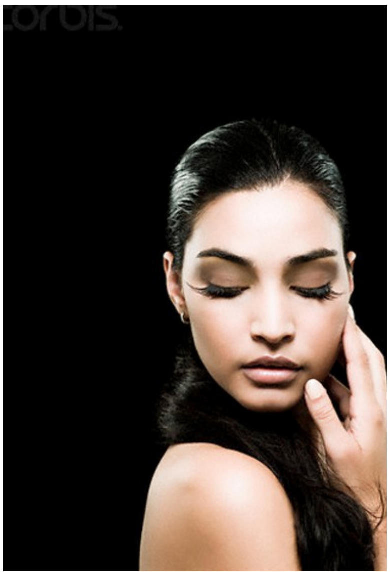
It may also be the case that within that same background image a palette of colours, or more properly hues, can be created and, therefore, some colours are less suitable than others but not completely unsuitable. This is where the principle of using Cumulative Malpractice Flag values becomes useful. If one were to simply use RGB (Red, Green, and Blue) values to describe different colours in an unrestricted palette then to map a rules-based system for only positive values of those colours would require some way of handling approximately 1.6777×10^7 different possible combinations; the rule of products providing this value. Even picking some X number of values of suitable colours from this range would be an onerous task.

However, if one were to instead either reject or look unfavourably on colours chosen from given ranges and to allow colours from other ranges to pass without note then the possible number of different combinations reduces to Y multiplied by 6.5536×10^4, where Y is the total number of possible colours minus the suitable hues. This is a noticeable difference in the number of options one would have to account for. From this point one could simply pick a range of hues and saturations that would be unacceptable and to what degree they would be unacceptable and from this point assign relative Cumulative Malpractice Flags values for them.

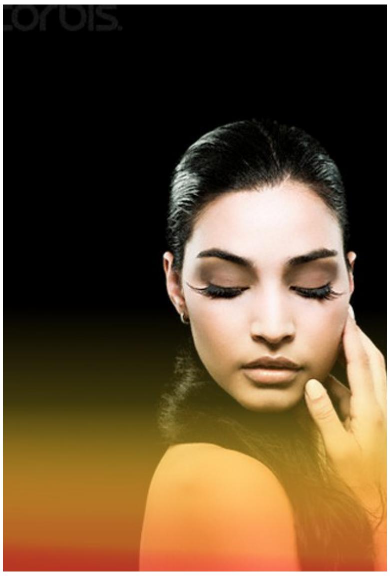
In the case of AdQuest participants were given only a limited palette of colours to choose from to reduce the possible complexity of colour based assessment however, the assignment of Cumulative Malpractice Flag values followed the same principle as in the hypothetical situation detailed above. In the case where the background image was black the colour variable Wrongness Flag value for the logo component was set equal to 1 if the participant chose the colour black. Each other colour in the palette was assigned a value between 1 and 0 where the value represented by percentage its distance from the identified least desirable range.

These values would alter according to the identified ranges of colour established in the layout map for any given background image. Therefore, whilst a black logo component colour would be given a Wrongness Flag value of 1 when the background image colour was also black, it was given a much lower score in the event that a background image colour was white. These alterations were also true for each other colour in the selected palette. With this value a logo component colour coefficient was created. Simply this was the difference between 1 and the given Wrongness/Cumulative Malpractice Flag value of the selected colour. In this way a *wrong* selection yielded a coefficient of 0, a *less favourable* result yielded a coefficient of somewhere between 0 and 1, and an *unflagged* result would yield a result of 1.

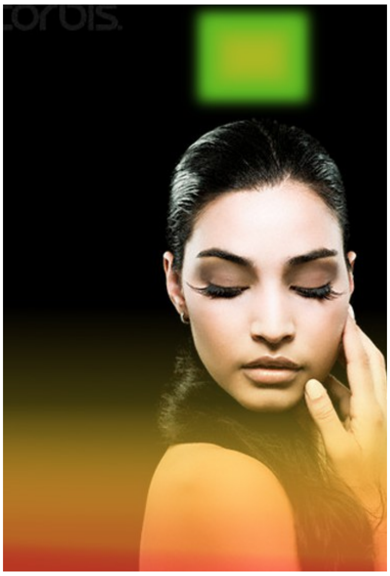
By way of further example the general positioning of components on the background was given a value using a similar process. Take, for example, rule 9 of the 34 given (found in Appendix A) which states “Vertical lines draw the eye down”. In the example image given below a vertical line can be determined from the model's eyes and their pose.



This being the case that leaves a vast area, shown in the image below as a colour gradient, of potentially ideal locations for the logo and product placement.

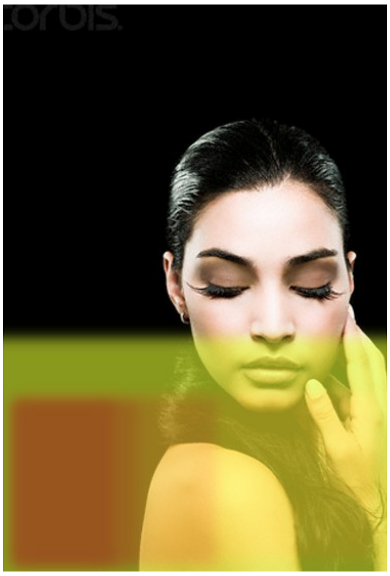


This is not the only possible location for the image however, as it would be possible to follow the vertical line of the model and position the logo and product components directly above the model. This would still maintain the vertical line and be in-keeping with the conditions of rule 9.

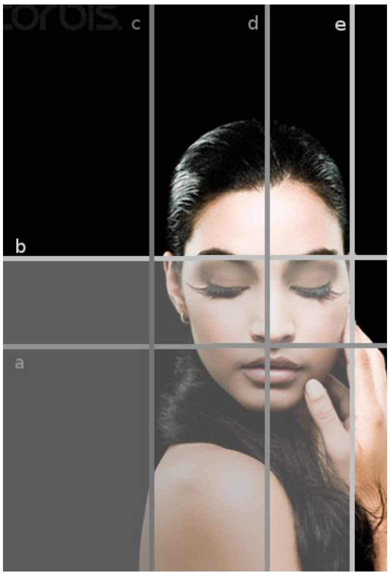


However, this would be less suitable as the eye is to be drawn down via vertical lines and, therefore, the viewers attention would be drawn away from both the product image and the company's logo. Being unsuitable it can be dismissed as a means of determining a range of locations.

This then leaves us with the original coloured area, however, it is not the case that this area is all equally as suitable. Typically positioning the logo and product components in the centre of this area would be less desirable than to one side for a number of reasons. Firstly, components that overlapped with the model would reduce in weighted contrast with those that do not – rule 18 "[c]oloured objects should be weighted heavier than black coloured objects". Secondly, the shape of the model's pose suggests that positioning should take place in the empty space – rules 6 and 11, "[b]alance should be in relation to the optical centre rather than the mathematical centre" and "[i]rregular shapes point toward the area it best emphasises". Thirdly, as the focus for the AdQuest scenario was to design an advertisement for a Chinese luxury brand market the components themselves should be stacked to match the constraints of rule 12 - "[a] design is at its most logical, rhythmically, when each of its components conforms to the direction of reading for that culture". These rules can be seen reflected in the suitable coloured areas shown below.



In this instance for the AdQuest scenario it was possible to determine the limits of the ranges of possible logo and product locations. The coloured grey rectangles are determined by the absolute values of the comfortable, or suitable limits, represented by lines *a* and *c*, and those which represent the extreme limits for the layout, represented by lines *b* and *e*.



The areas shown in the layout above, indicated by the intercepts at *a{c...e}* and *b{c...e}*, provide the ranges for layout positions where area under the *ac* intersect is most favourable and the area under the *be* intersect is least favourable in this suitable range.

In the AdQuest scenario such maps were developed for each background and, as with the colour coefficient, the relative distance of components' positions from their ideal positions were given values ranging between 1 and 0. This demonstrates the need to have developed a dependency of advertising components as described previously as these coefficients were only measured relative to some other given component. As before this value was subtracted from 1 to give a position coefficient to be used. These coefficients, and others similarly derived from component information, were then used to form the major values for the constraints in the AdQuest knowledge bank. What made these coefficients different from other information passed to the knowledge bank was that as their values could be altered so to might the assessment in more refined ways.

However, certain participant decisions could be flagged in more definite ways. The most noticeable example of this used was the presence of different products other than the desired product. In the AdQuest scenario participants were asked to create an advertising design specifically for Chanel No. 5 perfume. In providing other products it was possible to determine how closely the participant had paid attention to the design brief. As far as assessment was concerned if the participant was to choose a product other than one of the two samples of Chanel No. 5 that would trigger as a Wrongness Flag in the assessment module. From an epistemic point of view this was a necessary inclusion as advertising designers must choose their source materials carefully. In a real-world environment designing an advertisement for a given product and then choosing the wrong product would have serious consequences for the designer and, so, to give credit to the participants they were treated with the same level of expectations.

The development of the assessment module of the knowledge bank was developed on a branching series of internal calculations. These calculations were defined in terms of what assessment criteria was asked and would branch into various levels of complexity. For example a simple single-layered assessment criteria to be asked was "What background image was selected?", as the rules found in Appendix A provide for specific background images to be favoured over others. A more complex branch, as example, was found for the logo component which would include assessment criteria such as "Which logo image was chosen?", "Was its colour suitable?", "Was it positioned suitably?", "Relative Size – Comparison Product", "Relative Size – Comparison Background", etc.

In this way the lower order assessment criteria as determined by each constraint would affect the calculation of each higher order assessment criteria. For example, if the position of the logo was mismanaged it might also affect the assessment of how large or small the logo ought to be given its available space in its area of the background canvas.

Once all of these calculations were completed their local scores were calculated for the purposes of generating feedback, and these scores were altered by predetermined assessment weights so as to provide the participant with a mean score for their layout design. Whilst feedback was not the focus of the AdQuest case study, rather its purpose was to demonstrate the effectiveness of the Wrongness and Cumulative Malpractice Flags for assessment, feedback was offered to the participant in a basic form. This was accomplished by using the aforementioned local scores for each constraint so that, if they had made any specific errors in their design the AdQuest knowledge bank could direct them to the source of the error. This feedback could include an explicit correction of a particular aspect of their design and might also have included feedback of a more abstract and epistemic type relating to the context of the scenario.

4.3. User Tests

The user tests for the AdQuest scenario took place over three days during an extended simulated assessment with over one hundred design students at the University of Southampton. These students were chosen because of their previous experience with simulation-based tasks, their expertise in design, and lastly because each was an overseas student from the Chinese mainland. Of the 102 postgraduate students who played the AdQuest scenario and were surveyed, 50 were Design Management students, the other 52 were Advertising Design Management students.

Only one student at a time was called up to each terminal to play the AdQuest scenario and then complete a survey of their experience in approximately 15 minute slots. This was done to ensure that, at least during their play time of the scenario, the student would not receive any additional support from either their class mates or academic staff which would compromise one of the fundamental tests of the Wrongness Framework - that external interference effects negatively on user experience.

As part of a separate experiment, unrelated to the assessment effectiveness of the Wrongness Framework, fifty of the students completed the AdQuest scenario in English, the other fifty-two students completed the AdQuest scenario in Mandarin Chinese.

The changes to the Chinese language version of AdQuest were at a purely surface level, insomuch as the interface design was changed to a localised Chinese setting in terms of colour, text, and character design. No part of the Wrongness Framework assessment mechanisms were altered as part of this change and, as such, from a perspective of background mechanics the Chinese version is identical to the English version of the game.

Fig. 4.5., *Title Screen, AdQuest Chinese Version*

As well as surveying the participant's experience their achievement scores were also noted for comparison of how well they performed during the scenario against the experience they purported to have had. For their experience the participants were surveyed regarding various aspects of the scenario, described in more detail below, in the form of a statement with which they could agree or disagree on a graded scale. Their assessment scores (Interaction Index, Market Share Increase, Optimization Grade) were given as part of the feedback section of the scenario itself as a simple decimal.

The following is the list of User Experience Statements (UES) found in the survey for the scenario. The participant then graded how much they agreed or disagreed with each UES on a scale of 1 to 5 – 1 being “completely disagree” and 5 “completely agree”:

1. I've felt a connection with the character design.
2. The colour scheme of the interface design was appealing to me.
3. The game interface design was easy to use.
4. I liked the look and feel of the interface design.
5. I've enjoyed my experience with the game.
6. The interface design made me feel immersed into the game.
7. I feel confident with designing advertising for luxury brands.
8. The feedback was fair.
9. The task was challenging to me.

The quantitative values of the assessment scores taken directly from the scenario itself provide different information. The Market Share Increase is the easiest value to explain as it provides the epistemic “score” for the participant's efforts. This value is to be understood by the participant as how close their design got to achieving the 30% market share increase that was established as the goal in the scenario brief and is a way for them to judge their own performance. This figure was tied to the Optimization Grade. The Optimization Grade, itself, is simply how close the participant's design was to the possible default optimal designs. This score was achieved by simple comparison with the participant's node weightings in the assessment mechanisms and those of the default reasoning. The Interaction Index is a summation of values from the number of clicks, drag operations, resizes, and other alterations made during the game. A higher score indicates greater interaction with the components.

4.4. Expectations of the AdQuest User Tests

In order to make the most of the AdQuest user tests it was necessary to formalise what was expected in terms of results and participant experiences so as to assess both how successful the AdQuest user tests were and also to uncover any areas where improvement might be made. As with any testing protocol of this nature it has been necessary to assume that the AdQuest scenario has been developed effectively. To have been developed effectively is to say that the assessment mechanisms have been properly designed and implemented using the Wrongness Framework, and that the rules used to develop the rules-based system of the scenario reflect the prior knowledge of the user base. Thus, in the event that all of these assumptions hold it is then reasonable to predict that the results of the survey and scores taken directly from the scenario will reflect the following expectations:

- If the layout optimization procedures of the assessment mechanisms match a rules base for advertising design familiar to the typical assumed user's knowledge, then the majority of those users of that type should be placed in the upper bracket of the optimization score assessment.

- If the majority of users are placed in this upper bracket and those users report favourably to the statements of satisfaction, comfort, and challenge in the user surveys then there is a good case for assuming that the challenge of the task is achievable to that standard by most users of that assumed knowledge.

- Still further, if a significant minority are placed in the middle and lower brackets of optimization score, and also report favourably in the majority for feedback, and, additionally, respond to the other survey statements in the manner described above, then the challenge of the task is therefore not so easy as to be negligible to a user of that assumed knowledge.

- If the challenge of the task reflects the situation described above and a majority completes the task successfully within the higher brackets and reporting favourably on the experience, then the assessment mechanisms must be functioning and be well suited.

4.5. Results Analysis from the AdQuest User Tests

Firstly, it should be understood that, although they will occur in this section, the comparison of results from a perspective of non-localised and localised versions of the scenario is not the main focus that will be taken from results of these user tests. Indeed, whilst they may provide interesting surface level design ramifications for future work and may explain user attitudes during the tests, they have little bearing on assessing the effectiveness of the Wrongness Framework to manage the assessment of non-procedural and dynamic tasks. As such, results from survey questions which relate directly to user interface experiences will not be the primary focus of this analysis. It should be noted that the figures given within the results analysis in this section have been normalised so as to ensure that whatever comparisons occur that those comparisons are made relatively rather than absolutely.

Lastly to this preface, after testing some participants provided their own verbal feedback regarding their experience in addition to their survey results. As with the surveys this feedback was given anonymously but done so in an informal manner. Some notes were made, with participant permission, of this verbal feedback of common experiences and, whilst it will not be recorded in the results section as part of a formal analysis, it will be included in the proceeding conclusions section (4.6) to help to provide greater understanding to the attitudes and moods of the users themselves during the AdQuest experience. No attempt will be made, as a means to protect the participant's anonymity, to connect any comment with their qualitative or quantitative data so as not to compromise the validity of their survey results or comments by drawing unwarranted connections between their experience, comments, and abilities.

The table below shows the results of the User Experience Surveys for both the English version test group and Chinese version test group. The headings of "Agree %" indicates the percentage of each test group who gave an answer of 4 or greater to the survey question. Similarly "Disagree %" indicates the percentage of each test group who gave an answer of 2 or less to the survey question. The heading "Neither %" indicates the percentage of each test group who gave an answer of 3. Additionally the mean value for each survey question is given to provide further trend information.

	English Version				Chinese Version			
Survey Statements	**Agree %**	**Disagree %**	**Neither %**	**Mean**	**Agree %**	**Disagree %**	**Neither %**	**Mean**
1	62.00	6.00	32.00	3.68	57.70	21.15	21.15	3.25
2	54.00	18.00	28.00	3.38	50.00	13.46	36.54	3.39
3	92.00	2.00	6.00	4.42	82.69	9.62	7.69	3.99
4	66.00	14.00	20.00	3.70	46.15	13.46	40.39	3.35
5	70.00	4.00	26.00	4.00	73.08	7.69	19.23	3.79
6	58.00	6.00	36.00	3.62	44.23	13.46	42.31	3.31
7	52.00	12.00	36.00	3.62	38.46	30.77	30.77	2.69
8	64.58	6.25	29.17	3.80	57.69	9.62	32.69	3.48
9	18.00	48.00	34.00	2.62	36.54	34.62	28.85	2.78

Table 4.1., *Survey response results*

The survey statements found in Table 4.1 relate directly to the ordered survey statements found on page 61. Of the results listed in Table 4.1. of particular interest are the responses to statements 3, 5, 7, 8, and 9. The statement 3 ("The game interface design was easy to use.") responses suggests that for the vast majority of the participants of both the English and Chinese versions of the game few issues with the scenario should be directly attributed to the way in which the participant interacts within the scenario. Similarly, the statement 5 ("I've enjoyed my experience with the game") responses indicate that, for the majority of participants again in both versions – albeit a smaller majority, that the task was not so challenging nor so dull for the given time period it takes to play the scenario that the user's interest in performing the task waned.

This appears to validate the idea that the scenario task for AdQuest was set a level suitable for the typical participant from this group; thus corresponding with Thomas's and Young's (2010) aforementioned optimal game play corridor (Fig. 1.6.).

It is important to understand the results of statement 7 (“I feel confident in designing advertising for luxury brands.”) before moving on to the quantified results of the assessment scores as there appears to be some notable disparity between the participants of the English version test group and the Chinese version test group. Those in the English version appeared to be much more confident with the idea of designing advertisement to the scenario brief as only 12% of that group disagreed with the statement. Those in the Chinese version test group were far less certain of their capabilities as those that disagreed with the statement formed 30.77% of that test group. This is quite a peculiar disparity which calls into question the results given for the English version.

Whilst it has been noted that from the perspective of the AdQuest case study there was no change in the assessment mechanisms between the English and Chinese versions of the game, the differences in results in the two versions for each test group cannot be determined by quantitative correlative analysis of participant completion score and participant interaction score; English Version $\rho = 0.0029$, Chinese Version $\rho = -0.0527$. Yet, in spite of this the feedback from the participants of the AdQuest scenario rated their assessment to be accurate on the whole. It is thus the belief of the author that it would be of value to explore different explanations as to the disparity of the results received as the focus of the AdQuest scenario was to provide and demonstrate useful assessment mechanisms using the Wrongness Framework.

The confidence of one group as opposed to another cannot be attributed to the participants being drawn from two different degrees subjects as the participant groups of both the Chinese and English versions of the game were more or less equal in terms of their mix. It is possible that perhaps, as native Chinese speakers, the participants of the Chinese language version of the game, assessing their own understanding of the scenario brief as they played it and how they performed, believed the brief to be more complex than their English language version counterparts who might have satisfied themselves to the simplicity of the task due to performing this task in a second language.

This seems unlikely however as the quantitative results of each group are not particularly distinct and each participant group performed well at the task on the whole. Perhaps then it is a case of one group adopting a more reflective perspective of their understanding in light of their performance of a task or it may simply be a coincidence that the Chinese language version group hold more members which are unsure of their abilities.

Taking the position of the explanation above in the analysis of statement 7, it is possible to look at the results of statement 8 ("The feedback was fair.") in light of those results. The feedback to the participant, limited though it may be in this case study, does provide the participant with some vital information. For example, where the participant may have made mistakes those mistakes were noted and the participant was made aware of them. If the participant was to be rewarded or penalised, a notification of that reward or penalty was also given to the participant. Because of the inclusion of this rudimentary feedback the participant's confidence in completing their task, influenced by this feedback, will have determined how fairly they believe their work to be judged. What should be remembered is that this scenario was designed with prior knowledge of similar tasks assumed on the part of the participant because of the participant test population.

Despite the lack of complexity of the feedback offered through the Point of Enquiry character it appears that both the English version users and Chinese version participants were generally satisfied with the feedback they received; 64.58% and 57.69% respectively. Whilst it was the point of the AdQuest scenario to demonstrate how assessment mechanisms could be realized through the Wrongness Framework, the feedback mechanisms of AdQuest are linked to assessment mechanisms in a way which is worth expanding upon for the purposes of this section.

In principle the two mechanisms should be separate design issues as they are in real-world scenarios. For example, if a task is performed poorly and is assessed as being of a low standard then, objectively, it does not follow that the subsequent feedback on the performance of that task should be considered unfair if it is accurate despite being negative.

However, in an automated system where feedback mechanisms must be triggered from participant assessment then the link between assessment mechanisms and feedback mechanisms becomes more entwined as the triggers for the specifics of the feedback are activated or not based on the calculation of the assessment mechanisms. That so small a percentage of participants registered dissatisfaction with the feedback they received suggests that, as they are all participants with prior subject knowledge, and that feedback in the scenario is intrinsically linked to assessment mechanisms as described above, that the assessment mechanisms themselves must have been working to at least a satisfactory level for this participant base.

The responses for statement 9 (“The task was challenging to me.”) show, much like the results of statement 7, a noticeable disparity once again within the groups and once again with the Chinese language version group being less certain of their abilities. This, perhaps, may lend more credence to the idea that the Chinese language group take a more reflective approach or indeed more unsure of themselves within the scenario task – as expanded upon above. However, based on the responses given to statement 8, the participants of the Chinese language group do not seem to have had their opinions of the feedback they received noticeably coloured by any such perceived uncertainty. This suggests that the assessment of this participant group's attitudes may be valid.

The quantitative data taken from each participant taken from the scenario also provided some interesting insights to the idea discussed above; that of how the participants of both groups perceived their abilities in completing such tasks.

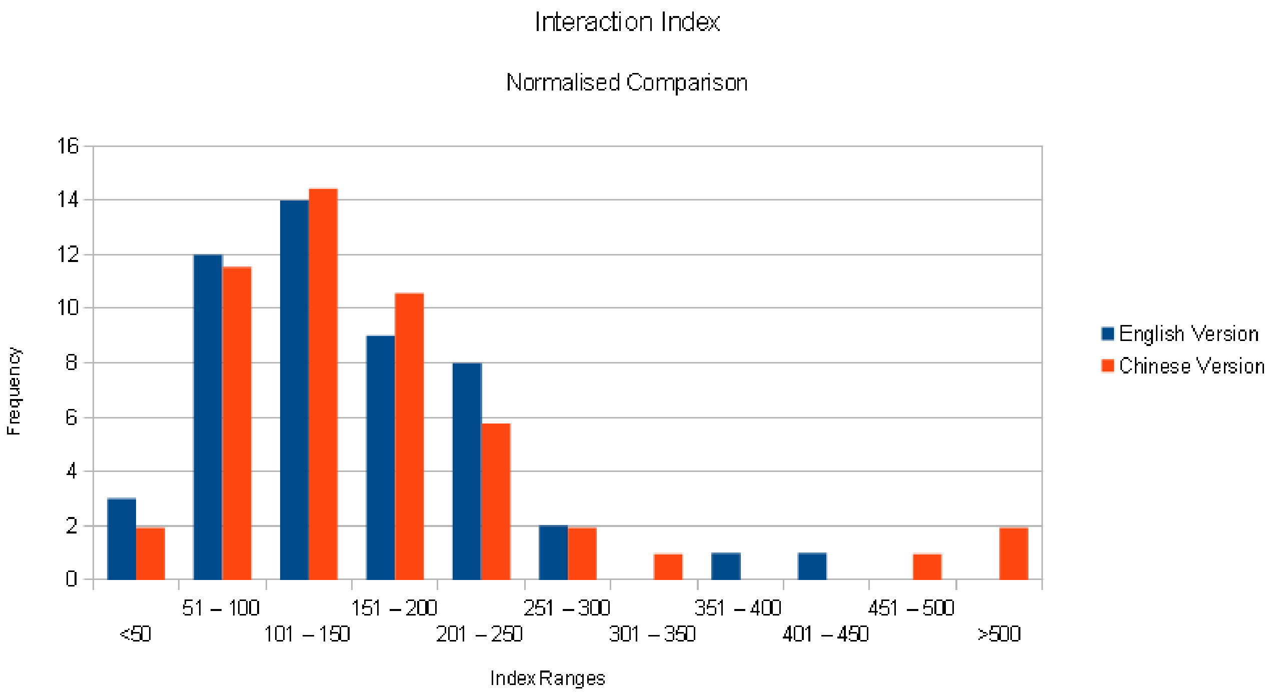


Fig. 4.6., *Interaction Index User Comparison Graph*

The results shown in Fig. 4.6. demonstrate that, for the most part, the interaction index scores for both participant groups trend in largely similar ways. The mean interaction index scores of the English and Chinese participant groups, 149.2200 and 158.3395 respectively, show that whilst the Chinese participant group had a greater interaction with the task itself – the scores showing how much alteration and deliberation took place during the design phase – that the difference between the two groups is marginal. Indeed this difference can be attributed mostly to the few outliers within or just beyond $+3\sigma$ which exist in greater frequency in the Chinese language participant group. Even allowing for the 5.76% difference between the means of the two participant groups a strong case cannot be made that one group or another had significantly more interaction with the scenario during the design phase than the other.

If this is the case then it appears to be less likely that the difference of the responses to the statement results are the case of a greater reflectivity on the part of the Chinese language participant group.

Instead to invoke a simpler explanation would be that the results given in the survey should, by and large, be taken at face value and that the question of disparity is simple - that a disparity exists between members of the participant population based on ability which may or may not be attributable to a localised user interface.

Comparison with the survey results, interaction index results, and finally the optimization score of the participant groups may provide the final piece in explaining any such disparity between the participant groups.

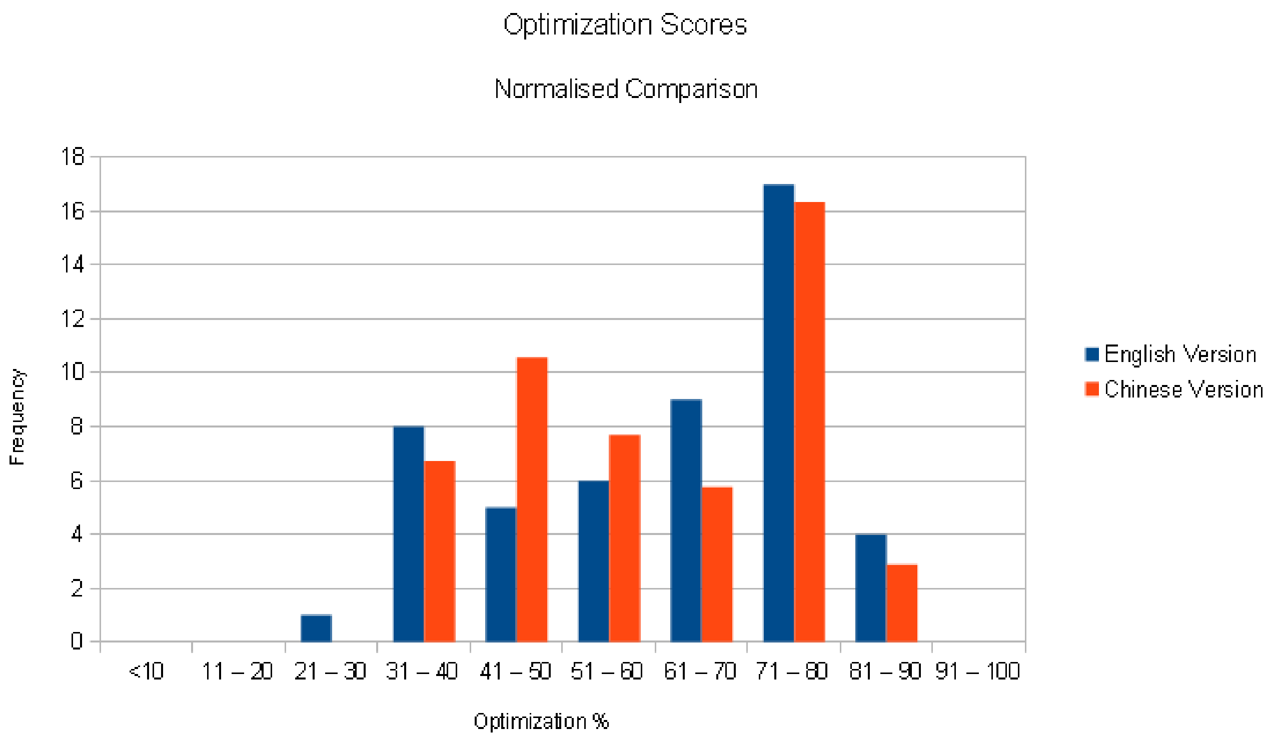


Fig. 4.7., *Optimization Scores User Comparison Graph*

The frequency of optimization scores in Fig. 4.7. clearly demonstrates a disparity between the abilities of the English version participant group and the Chinese version participant group. The majority of users in the English language version group, 60%, were placed within the higher optimization range of 61 – 100%. However, only 50% of the Chinese language version group were also placed in that range.

Still further as 36.54% of the Chinese version group were placed in the middle, 41 – 60%, optimization range as opposed to 30% of the English version group, the respective abilities of the two groups becomes more apparent. The remaining percentages of the English and Chinese version user groups in the lower optimization range of 0 – 40%, being 10% and 13.46% respectively, are somewhat outlying scores on the face of the entire participant population for the tests.

To put the results of the optimization scores in their proper place it would seem that a case could indeed be made to support the argument that, from the perspective of the Chinese version group results, one group appeared to be more able than the other within the scenario. In light of this deduction then it would be that ability disparity is the assumed position rather than the other earlier stated argument of reflectivity.

Despite this argument gaining validity it is worth exploring alternative explanations. It is of course possible that the translation for the scenario was faulty and misrepresented what was expected of the users of the Chinese version. This, however, seems unlikely as an explanation as the scenario dialogue was translated and checked by three native Chinese speakers before the user tests took place and, additionally, during the scenario a helper was on hand in case the participant had any questions which could not be answered by the PoE in game. Therefore, if this disparity of ability in scenario as confirmed by the optimization scores and interaction index scores is assumed, then it answers for the disparity of participant experience in the survey results meaning that those results can be, and in this analysis have been, taken at their value.

In light of this analysis it seems reasonable to make the following claims about the relative success of the assessment mechanisms within the AdQuest scenario:

- Given the high success rate of most participants, as demonstrated by the frequency of participants within the higher range of the optimization scores, it is the case that the challenge of the scenario appears to be pitched at a level which would make it possible for the majority of users with similar backgrounds to achieve similar results.

- Additionally, as there exists a notable percentage of participants from both groups in ranges other than the higher range of the optimization score and that the results of the interaction score take are very similar across both participant groups, then there must still exist some challenge for success even for users of a similarly experienced background.

- In light of the survey results regarding user feedback, task challenge, participant confidence, and ease of use, and assuming the difference of participant ability to be true, it would appear that the assessment mechanisms function satisfactorily in assessing performance and triggering feedback for such a topic given a specific rules-set. This can be said insomuch as participant experience appears to be in concordance with participant performance against test expectations as described in the earlier dedicated section.

In this regard it would appear that the results of the AdQuest user tests do match with the noted expectations of those user tests broadly. There are minor points which need be discussed regarding further optimization of the AdQuest scenario and the assessment mechanisms employed, however, it would appear that, for the most part, the user tests were completed with relative success. This means then that it can be said that the remit for the case study, providing assessment mechanism designs using the Wrongness Framework in subjective/dynamic topics, has likewise been demonstrated.

4.6. User verbal feedback

As mentioned earlier, it is important to consider the informal verbal feedback of the participants from the test groups where it was given. Much of this feedback is, understandably, personal to the particular user who gave it. However, as personal feedback it provides an opportunity for addressing the particular needs or concerns highlighted by certain members of the user population. It is worth reiterating though, that as this feedback was given informally and anonymously it will not be used to colour opinion of the results analysis but may be used at face value for considerations when drawing conclusions from the AdQuest experience.

- “As a dyslexic it was difficult to get all of the details for the task from the brief text.”

No attempt was made in user interface designs either the English or Chinese versions of AdQuest to accommodate for those people with any of the conditions which may make reading on-screen text difficult. This is not to say that the presentation of the brief during the appropriate phase of the scenario was particularly wordy but rather it was not separated in such a way which might have made the information of the brief clearer. All the information regarding the specifics of the desired demographic of the advertising campaign – their known preferences, characteristics, etc. - were included in a single section of on-screen text.

Whilst this is not an issue which is directly related to the matter of designing the assessment mechanisms for the scenario, this is not to say that this is an unimportant or trivial issue. Indeed, despite the fact that fixing this particular issue falls into the remit of user interface design, the effects of not fixing this issue will have impact on the performance of the user within the scenario. In this case the importance of resolving such issues cannot, and must not, be dismissed as without clarity of purpose no user can expect to perform within any Simulation or Serious Game at their best and such issues will ultimately affect the input and output of the assessment mechanisms themselves.

- “I know it was the wrong product but the colour was better.”

This is a particularly interesting piece of feedback in terms of how the Wrongness flags and Cumulative Malpractice flags are used in the design of a Simulation or Serious Game scenario. As mentioned previously in the section outlining the game play of AdQuest, a number of incorrect Chanel products were included as part of the design phase of the scenario for the purposes of demonstrating how Wrongness flags might be used within the assessment mechanisms. In practice this meant that to design an advertisement with the wrong product, regardless of any other desirable qualities that an incorrect product choice may have – size, shape, colour, etc., all of which accounted for within the attribute weightings of the nodes described earlier, was, and should, still be seen as unacceptable and unprofessional behaviour in advertising design.

Despite this position it brings into question the idea of including deliberate “red herrings” in a scenario in such a way that there exists traps for the user to fall into and reduce their chance of success. In the context of the AdQuest scenario their inclusion is, arguably, non-controversial. After all, layout designers for advertisements will have a wide range of stock product images for a company as well as their models and logos; as was presented in AdQuest in an effort to remain epistemic to the professional context of the scenario. It then becomes a matter of professional diligence for any such designer, when consulting a catalogue of images and no matter the size of that catalogue, to ensure that their choices are made in accordance to the confines of their design brief.

The use of Wrongness flags in the knowledge management of the Wrongness Framework is at the core of its design principles and so they must be handled effectively. Ultimately their inclusion in the assessment mechanisms of a scenario of greater scope than the one shown in AdQuest accounts for more than just the feedback and grading of a participant's performance but also so that they manage, as with the Cumulative Malpractice flags, the environmental changes to the scenario and future scenarios within a more complex Simulation itself. This being the case Wrongness flags provided the participant with a chance to fail if they commit the triggers of those flags no matter how many times the participant could be said to have adequately navigated the minutiae of the scenario's Cumulative Malpractice flags. As such their importance cannot be overstated.

However, the nature of this feedback suggests that there may exist some default working principles which exist in participant of a given discipline that might override some default reasoning on the part of the Wrongness Framework. For a layout designer perhaps the identification of the product itself is a secondary consideration when presented with a catalogue of choices all of which have desirable qualities for their designs.

That being the case, perhaps it is better to also account for an understanding of the types of typical participant for each scenario if that scenario presumes experience as AdQuest does. In doing so a greater level of complexity could be allowed for in the Simulation/Serious Game designed with the Wrongness Framework and, in essence, provide a greater range of possibilities for the user to engage with.

The default reasoning could be improved to account for these assumptions in such a way that might allow for more refined feedback triggers. This being possible, and duly tested, then it becomes a much simpler matter to map all of the intricacies of a syllabus of learning in a way that would allow the achievement of one of the goals of the Wrongness Framework: that it might monitor, track, and lead the progression of long-term users in larger Simulations/Serious Games with little need for direct interaction with a real-world tutor.

4.7. Conclusions on the AdQuest Experience

As per the map of the evaluative process (shown in Fig. 3.1.) it has been the task now to review the AdQuest scenario in light of the data analysis of the previous section and draw conclusions on how further designs using the Wrongness Framework's featured components may be improved upon based on this experience. As discussed in the results analysis section, each of the user test expectations listed previously were met broadly confirming that which was hypothesised. With that being the case it is reasonable to take the position that the designs for the assessment mechanisms were validated and demonstrated successfully. If this position is held then it is justifiable to say that a pivotal part of design using the Wrongness Framework has been shown to be effective.

The Wrongness Framework's chief goal is to provide the back engine design to facilitate knowledge management in Simulations and Serious Games. This is the case insomuch as assessment and feedback mechanisms directly relate to the participant's experience and their own path of learning through knowledge management. In this regard the AdQuest experience is successful. However, a number of small changes or additions, as identified in analysis, might make for more refined and improved participant experience.

Firstly, some account should be taken of typical participant experience priorities if working within a Simulation or Serious Game where prior knowledge is presumed of a given procedure or similar codified behaviour.

This would not be easy, and ideally it would result in the formulation of a logarithm based scale which could be used to identify the various intra-nodal weighting differentials in such a way that the knowledge management of the Wrongness Framework through the Wrongness and Cumulative Malpractice flags might be more readily codified. This would result in a far simpler grading scale of subject knowledge and may influence the manner of its delivery and assessment. By this definition it would be difficult to codify these priorities in such a way that could be standardised and could have application in all instances.

It might be the case, however, that rather than attempting to create such a code and implement it in multiple design contexts to propose it merely as a hypothetical or philosophical position on knowledge management design within the Wrongness Framework and attempt to make applications of it where possible, practical, or sensible to do so. Despite the difficulties associated with attempting to do such a thing, as the results and feedback for the AdQuest user tests show, implementing any such measures on the scenario design might increase user-scenario interaction, understanding, and facilitate the expansion of more complex and interesting assessment criteria. However, it is not within the remit of this thesis to propose any such addition to the Wrongness Framework particularly as it is an area which would require dedicated research and empirical data to support any such codification of the principle. It does, however, provide an interesting position to take in the expansion and refinement of the Wrongness Framework for future research projects.

It is, of course, worthwhile remembering that the AdQuest scenario takes place in some non-existent longer term Serious Game or Simulation. It is difficult, necessarily, to decide where in that longer term experience the scenario takes place as it presumes a certain amount of prior knowledge. Is the longer term Serious Game/Simulation an advertising design training exercise wherein participants will be taken from no knowledge whatsoever to some competency in the subject or, rather, a Serious Game/Simulation attempting to provide a professional experience for those already trained? This distinction is important when considering the concept of participant experience priorities as, depending on the nature of the Simulation/Serious Game itself, these experience priorities may be encouraged or discouraged as seen fit in preceding scenarios.

In the case of a Serious Game/Simulation which presumes no prior knowledge this takes the problem back into the hands of the experts interviewed to develop the user progression map and rules-based assessment rather than to leave it to adaptation; in essence a tailored experience.

The final point of conclusion for the AdQuest scenario is to consider the role of feedback in promoting learning during the task experience rather than simply to make an assessment of the degree to which the participant passed or failed the assignment. Such a change would be simple to implement. One such way would be to offer the participant an opportunity to submit a draft proposal for their design as an alternative to a final submission. Using this option, and exactly the same assessment mechanisms as the final submission, it would be possible for the Point of Enquiry character to offer some preliminary feedback, or indeed other users to more easily facilitate Aldrich's third slate, and point out where the user may have gone wrong, if at all, and provide the participant with another opportunity for submission.

To facilitate this option it would be necessary, however, to make an account of time – a factor not present in the current AdQuest scenario – and also the relationships between participant as designer and project leader. Time becomes a necessary factor from two perspectives: design drafts rarely have the luxury of endless submissions, and it would provide the scenario with additional pressures which would exist in the real-world and provide further data to analyse. The relationships of the user to their in-game colleagues also presents the opportunity to explore facets of emotional intelligence in decision making and task delegation – a fascinating prospect for long term Simulations/Serious Games.

The case study found in the following chapter provides an opportunity to explore the addition of these complexities with a special focus on application within a longer term Simulation. In doing so it is hoped that the work of AdQuest can actively influence future designs in allowing for greater freedom to be demonstrated in exploring feedback mechanisms now that the assessment mechanisms of the Wrongness Framework have been validated.

* * *

Chapter Five:
Applications using the Wrongness Framework: Promasim

(Ethics Reference: 15867)

Much like the preceding chapter, this chapter is again another presentation of the Wrongness Framework within the context of some Serious Game/Simulation design. This case study, Promasim, is a little different from the AdQuest case study insomuch as Promasim has not been developed nor user tested as AdQuest has. The purpose of this case study is to examine the wider design principles of feedback mechanisms in light of the findings and validation of the assessment mechanisms in the AdQuest case study and with the direction input from a series of expert interviews. As one might expect, there is no empirical data to support this case study such as the previous AdQuest study and, as such, Promasim should be seen as the extrapolation and exploration of the design of a larger Simulation using the Wrongness Framework. As Promasim is not to be actually developed then it should be understood that references to specific scenarios or Simulation interactions are used only as extended examples of core feedback mechanisms and not simply as specific solutions to specific problems within Promasim. Any discussion regarding the nature of those interactions is only to serve as providing a frame of reference for the discussion of a specific element of the feedback mechanics and must be simple enough to serve as a context.

5.1. What is Promasim?

Promasim, a portmanteau of project management and simulation, is a larger Simulation in which users assume the position of a project manager for a short term software development project. There are a few points which need to be clarified before it is possible to continue with the description. Firstly, by larger Simulation it is meant that the Simulation makes use of multiple multi-levelled scenarios rather than only one or two. The design ramifications of a larger Simulation means that there is scope for user's in-game actions to have near or far reaching consequences and knock-on effects from one scenario to another.

The impact of these consequences is discussed in greater detail in later passages describing the design of in-game mechanisms that are affected by those consequences. Secondly it is important to discuss the context of the Simulation. As far as designing a case study to demonstrate these aspects of design with the Wrongness Framework there is no particular or intrinsic reason why a software development project should have been chosen over any other scenario context. The reason for choosing a software development theme for a project management Simulation is simply a matter of two reasons: familiarity of the topic on behalf of the author, and that software project management is a collection of procedural tasks of varying complexity which lends itself well to Simulation adaptation. Whilst these choices may seem to suggest similarities between the Simulation of this case study and the Serious Game of the previous case study there are important differences between them as well. Promasim, unlike AdQuest, is driven not by the procedural tasks of its context but on the emotional interactions of users with in-game characters and with one another. Additionally, and unlike AdQuest which involved user tests with a user group with experience of the subject matter, Promasim has no user tests.

5.2. Some initial philosophical implications for design

This puts the knowledge management designs for all of the Wrongness Framework components for the Simulation in a novel position; that there is no compelling reason for the rules-based system of Promasim to be sound. All that is necessary is that the default reasoning of the Simulation be internally consistent and valid in its conclusions. What this means for the expert interviews is that the individual answers to certain questions can be assumed to be true in the event of concordance with other answers, and can be acted upon within the design of Promasim as if they were true with no need to test the veracity of those positions. In this way Promasim adopts a much more experiential design context wherein hypothetical users can react to the reality of a Simulation world as it is described by experts.

The whole of Promasim is to be framed on the basis of seven scenes, where those scenes represent the broad seven phases of project management; pre-planning, budgeting, feasibility, design, development, testing, and installation.

The activities of those scenes will therefore be related to the typical activities of project management associated with those phases and will include typical interactions associated with those phases.

As such this allows that the participant will still be able to partake in the procedures of project management used for assessment mechanisms but that they will also be expected to deal with unique issues regarding stakeholder management which add a greater complexity to the task. As an participant might be expected to develop a simple UML class diagram for a package in the software, so too will they have to ensure that it is presented in a style which their non-technical customers understand.

In addition to these tasks the user will be have the ability to interact with their team members, in-game non-player characters, through a variety of communication means as well as other project stakeholders. These characters are to form the greater part of the feedback mechanisms to the user and provide many of the unique scenario interactions. In this regard the use of these characters is mirrored exactly as described in the diagram for the Wrongness Framework's process in Fig. 2.1. So it is important to note that using these characters as environmental variables establishes them, and not software project management procedure, as the context for the entire Simulation. In this regard the format for the entire larger Simulation of Promasim matches that of the classic ITS model as described by Charles et al (2011) and Sleeman and Brown (1982). It is necessary to describe the in-game characters of Promasim in slightly more philosophical terms to fully appreciate how they might be designed to facilitate this kind of learning experience.

As such the question of presentation should be addressed. It is likely, in light of the findings of the expert interviews, that any representation of a character might take place in a variety of ways. For example in a face-to-face meeting with a team member, the user may see, be able to react, and confirm emotional biases in the in-game character's body language which may colour the user's feelings towards that character and how they interact with them throughout the lifetime of the Promasim experience. This is a perfectly natural and proper response when discussing the addition of emotional intelligence in working experience.

Likewise a character may interact with the user via a medium such as email in which case the representation takes on the form of the written word where emotion is inferred through phrasing and in a way which may be left open to interpretation. What is of fundamental importance is that the nature of those characters are going to have to be consistent, or rather that they will change in a manner which can be reasonably predicted by the user.

5.3. The expert interviews

The initial philosophical design implications discussed in the previous section do at least provide a focus for design. From a mechanics point of view there is little of interest in just translating the ordinary, or rather “textbook”, project management activities. As one example, cost triangulation methods used for project resource estimates are a typical activity that a project manager may engage with, and in software development there are a number of such methods that a project manager may use or be tested with. However, the ability to calculate such costs with any of the potential methods cannot reasonably be argued to require the benefits of Simulation. Rather, this becomes something of a pen-and-paper exercise and even if it is a task which is digitized then any such representation becomes like an independent application within a more interactive Simulated experience and its educational impact becomes suspect. As such the question of management activities in a simulation focusing on project management is more aptly applied to the interactions of people or characters within Promasim.

To understand more fully the management of skilled people, professional expectations, and good stakeholder management, a number of expert interviews were undertaken with people who have a great deal of first hand project management experience. The interviews themselves were framed to include the context of the proposed scenario for Promasim so that any questions answered by the expert could be used as directly applicable and without extrapolation from a general idea. Similarly the questions which were asked as part of these expert interviews were focused on stakeholder interactions within a project and activities which would affect those stakeholders. This is to say that the questions of stakeholder interaction were based around direct communication principles with that stakeholder rather than, for example, theorized strategies for project management.

It should be understood that these interviews were undertaken with some degree of anonymity and, consequently, the interviewees will not be named here so as to protect that anonymity, however excerpts from the interviews which were recorded will be transcribed in passages where it is relevant to do so.

Where it is possible to do so initial predicates for design decisions will be noted as they come up for simplicity's sake and then summarised in a concluding section. As the complexity of expert experiences and expert answers to certain questions will differ, even if that difference is only slight, it is necessary to explore those differences and to make decisions on which answer will be adopted within the design of the Simulation. Unlike AdQuest where a hypothesis was presented, user tests were undertaken, and results were analysed for conclusion, the Promasim experience will remain within the realm of reactive design to identified challenges rather than attempting to present a optimal version of itself for testing.

The first, and possibly most interesting, design implication to come from the expert interviews was that of time. Almost all of the experts when asked regarding the question of scheduling and time management with project activities answered that they would, ideally, hold meetings or communicate with project stakeholders in the morning and at the beginning of the week. This, in itself, isn't a particularly remarkable discovery; there is a great deal of common sense to this approach as it allows weekly scheduling to be reactive rather than reflective in focus and plans for the week are decided during that week rather than given as a check-list of documented actions from the week previous. What is of interest are the insights regarding the experts' experiences on team performance based on these time factors.

Below is an excerpt from one interview which highlights the kinds of time-performance factors discussed above. Again, for the sake of interviewee anonymity the expert is not named but can be identified in the transcription by the letter “I”. Similarly the interviewer is identified by the letter “Q”:

Q: How often would you prefer to meet/communicate with members of your team?

I: Can I personally disagree with that? [...] I think with that one there's two different things. You've got meetings and you've got communications?

Q: Sure.

I: My view is in a 3-month exercise, you'd get the team together once a week. But the chances are you'll communicate with members of the team on an hourly/daily sort of basis.

Q: Sure.

I: Communications will be a subset within the meetings cycle.

Q: If you would meet/communicate more than once within a given period of time, would you arrange those meetings to discuss specific topics uniformly? So what that basically means is do you deal with certain business on specific days of the week or do you tend to take that, or do you tend to roll with the punches?

I: My way to do it would be that you would bring the whole group together one day once a week say on a Friday. The trouble is it could be either Monday or Friday; Monday you can plan forward, Friday tends to be looking back. My personal thing would be to perhaps meet some time on the Monday, look at the following week and say "Right, we did that last week, this week we're going to do X, Y, Z, and so by next Monday will have either delivered or not delivered and, because it's a very short program we should know why it's not delivered." [Scheduling specific days] would mean you get bogged down in paper work and the meeting should perhaps be no more than forty-five minutes to an hour or you're stretching people's patience.

The information from this excerpt, and similar expert conclusions, provides the opportunity to explore how time-performance is managed within Promasim. Take as a predicate, as derived from the expert interviews, that as individuals become irate, dissatisfied, bored, tired, or any combinations of these states that their work performance will suffer. Assuming this to be true in all cases, for the purpose of Simulation, then a number of different ways to represent the effect of time on non-player characters becomes possible. For example, perhaps all non-player characters will be given some fatigue limit and performance coefficient such that when that limit is reached by the assignment of daily project activities the value of the performance coefficient is degraded to mimic the drop in work quality for that non-player character.

This, however, is not a particularly satisfactory way of achieving that kind of in-Simulation effect and would be a rather crude representation of Cumulative Malpractice Flag and Wrongness Flag triggers using broad and unrefined environmental triggers. As the excerpt above suggests this matter of time is not simply a matter of days within weeks but rather hours within days. If, by necessity, each in-Simulation day must be broken down into hours then it must be possible to complete specific daily project activities whilst within the time references of the Simulation otherwise it will be very difficult to incorporate any kind of time-performance factors within the Simulation in a subtle and necessarily complex way.

The level of detail in such designs is a matter for some discussion as well as people within teams have different thresholds and limits themselves. Those limits though are not all encompassing. Tolerance of irksome behaviour will, no doubt have some affect on project team members even if they themselves choose not to be affected by it; making such a deliberate stand necessitates at least some portion of personal attention which cannot be given elsewhere – a distraction.

Irksome behaviour is not necessarily just behaviour which is intended to annoy purposefully. All the experts interviewed were given two simple hypothetical situations which may arise within the lifetime of a project. The first hypothetical situation is concerned with how the experts would handle a junior team member who may lack a specific skill or skills and what they might do about training them to bring them up to speed in terms of their abilities so that they may succeed.

This shortfall is irksome to the project manager as it provides an additional complexity to risk management within the project, and it is also irksome to other team members on whom this shortfall may impact either directly or indirectly. The interview excerpt below demonstrates a commonly held view amongst the experts.

> **I:** In a few months there is no feasible prospect of training anybody to a level where they may now make a difference. [...]
>
> As a the project manager I'd just make sure that they weren't given anything that took them out of their envelope. And as a completely separate issue that when this project is over you really need to do some training. So separate arrangements would be made. So I would avoid the situation in the first place. Depending again on the organisation and the impact I may excuse such a member from the project team. I'd say "Go away, don't want you. Explain anything you like to your boss but you're not on my team." [...]
>
> Some clients are perfectly comfortable with the odd mistake and feel quite positive about their contributing to the training of junior people. In which case that's fine. Most clients however, forking out several thousand, tens of thousands, hundreds of thousands of pounds aren't quite so benign. This junior member wouldn't be on my team, he wouldn't be seen, he'd have nothing to do except maybe make the coffee or something like that. I'm very hard about that. I've been in too many situations where you do what you can for a junior member and everything just goes pear-shaped.

A majority of the experts who shared this position also favoured the use of mentoring systems whereby inexperienced team members may be paired with more experienced team members so that they might make up their skills shortage by learning in the field. This type of approach would likewise impact on in-Simulation time management in two ways. Firstly the mentor, junior member, and project manager will now need to take up additional duties which will consume time in their schedules and affect their performance in the ways described above.

Secondly, it may be the case that there is some issue regarding the effectiveness of that mentoring experience and that for those people there is increased dissatisfaction for the project experience as whole because of this one part of that experience.

The second hypothetical situation is concerned with situations where there is a clash of character or misunderstanding between project team members and even clients themselves. In this hypothetical situation the irksome behaviour becomes an issue of personality rather than performance.

To this end it is worth asking whether such behaviour should be included in Promasim or not and what its value its inclusion would add for the user. The next excerpt provides some insight with regards to answering that question:

Q: Suppose there are some personal difficulties between members of the team. Has your experience innovated any way of dealing with and resolving personal difficulties quickly and effectively?

I: Quick? Never. In my experience personal difficulties manifest usually over a period of time. It's unusual to have them instantly. Once manifested they're pretty much set. I've never been able to get chalk and cheese to mix. I have been able to get chalk and cheese to work together - which is a completely separate issue.

Step one is to make it clear to everybody, team included, you can park your ego at the door when you come into work. Step two is, and the method I follow is documented in a number of places, it simply says to the people who are disagreeing - both of you have a valid point of view. There's no such thing as somebody's right and somebody's wrong. You now need to go away, lock yourselves in this room get the fire going in the chimney and give me the white smoke when you've agreed three things:

1) Articulate until you're blue in the face what it is you need and what it is you want, what it is that is important to you.[...] You've got to be honest. You're not honest this isn't going to work, one of you is going to lose your job.

2) […] You have to accept the legitimacy of the other person's view. If you harbour any doubts as to that legitimacy you cannot proceed.

3) […] What's your solution? Agree a solution. Both of you have to agree a solution. So that's my method.

As one might imagine, this kind of approach to resolving such difficulties is common place amongst the interviewees. By and large, they all agree that the best one could hope for is to achieve a professional working relationship between two or more team members and each propose methods similar in style to the one described in the above excerpt. The matter of whether it is of value to add such dynamics to the Promasim experience is open to question. Unlike the scenario of field-training for junior team members which at least provides and opportunity for project managers to elevate the performance of their team, this kind of scenario only increases the opportunity of failure in a way which cannot be justified as adding value to the educational experience. This approach is based in the understanding of game-play as expanded upon by de Freitas and Jarvis (2009) in Fig. 1.3. wherein reward structures and motivation in Serious Games/Simulations should not feel Sisyphean; as the constant monitoring of professional conduct amongst flagged non-player characters would – even if it is a potential occurrence in real-world project management.

What is important to take from this then is that time-performance is much more nuanced than the idea of activity impact versus activity scheduling, and as such time-performance variables have to exist in a more complicated form than as a simple limit for in-Simulation team members. It is necessary, then, to account for a number of different variables which make up the time-performance limit of each of the characters and to ensure that they can be referenced in a meaningful way. As to the particulars of those specific time-performance variables more is discussed in a later section.

However, the benefit of this approach is three-fold: there can be greater variance in the personality types of in-Simulation characters because of the increased number of possible permutations those variables can exist in, performance can be influenced and tracked in a multitude of interactions – both positive and negative, and the relative values of those variables with those of other in-Simulation characters ensures that it is possible to create dynamic team member relationships that mirror the experiences of the interviewed experts. All of which can exist without the need for some invention of a needlessly complicated feud between non-player characters.

One might ask why this is a fair conclusion to come to. After all isn't the possibility of a non-player character with an incorrect skill base that will require mentoring and monitoring an equally complicated event?

The answer to which is yes that is certainly the case, however, and as the expert interviews show the management of a personal feud between team members is a typically hands-off affair. The two team members are expected to quickly and appropriately work out their personal differences before it impacts negatively on their professional lives; actioned by the project manager but not overseen. Mentoring and monitoring a junior member of staff to ensure that their inexperience does not impact negatively on the life of the project is a hands-on affair necessitating time and attention. It would be fundamentally unfair to judge the user of Promasim on the basis of a game of chance rather than by their direct interactions.

This is not to say that inter-personal conflicts could not arise in Promasim. In fact they will be a necessary part of providing the kind of epistemic feedback that the Wrongness Framework has been developed to give. Rather it is to say that inter-personal conflicts – such that they may exist in the guise of approval rating or some such similar variable – will be of the user's own making. Most of the remaining questions asked in the expert interview itself are dedicated to providing the information to ensure that these kinds of events can be managed effectively.

The last topic covered in this expert interview section relates to the treatment and communication with clients within a project lifetime. It is one thing to dictate the relationships of team members who have to report to a manager directly but it is quite another to handle the interactions between client and project team even if that dynamic takes place within the same organisation. Clients as environmental variables within Promasim must exist within the confines of two broad concepts: their confidence in the project team as vendor to their project, and their satisfaction with the efforts of the project team as they have performed so far. Whilst these two concepts are sometimes linked they are not intrinsically linked. It is not enough to say that satisfaction will guarantee confidence as both of these two values are dependent on prior knowledge. The communication principles between project management and client have to account for the normalisation of professional standards, benchmarks, and, lastly, jargon.

For example, suppose that the initial industrial benchmark for a cloud computing system within the first test phase is around a 30% fail rate. If that is so and for the purposes of this hypothetical scenario a cloud computing system project in this phase had a 28% failure rate then it is reasonable to say that the development is performing as expected for the phase. It is only reasonable to make this assertion, however, once one is aware of the industrial benchmark. Otherwise just using the colloquial understandings of what a 28% failure rate means may cause a lack of confidence or satisfaction until it is explained. The question then becomes one of whether or not the project manager should explain the situation specifically using figures or whether a general trend explanation would be sufficient. Naturally the answer to this question is dependent on audience.

I: It's very much dependent who the clients are and how long you've had a relationship with them. [...] when looking at something like the development of software you have to split this into two bits. You've got to say, what is this thing, what that thing does, and deliver - and that's the easy bit to explain. And that may be the level which they need. [...] But if we're dealing with the techies on the bottom level who want to know code, architecture, etc., then you can go through that in fine detail. So it depends again on what are the expectations.

[As to techniques] apart from death by PowerPoint, you have to describe what you're doing in a clear simple way and at the same time bring to the table what you have developed [...] and get them to buy into the specification. [Doing so ensures] that it doesn't come as a great surprise and they understand what is happening and when in two months and twenty-nine days you avoid a situation where they say "what is this you have delivered?". And you should meet with a representative of them on, perhaps, a two weekly basis.

The nature of the client's relationship with the project team is something which should be defined at the outset. As the excerpt above indicates, and as was discussed previously, the presumed history of interactions between client and the project team's organisation will make a difference to the tone and content of communications between the two. Unlike the management of disagreements and personal issues between team members, a variance in client could provide an opportunity to increase or decrease challenge for the user.

Designing the communication protocols for a client point of enquiry character wouldn't be much different than for a team member point of enquiry member but it would require, perhaps, more variety in the tone of that communication. The introduction of those kinds of variables would provide an opportunity for the user to engage in a more nuanced relationship with their clients and increase the range of interaction between these two in-Simulation agents. The excerpt below develops some of the ideas which may be pertinent to achieving this aim and explores the kinds of working communications that should be made possible with the client in Promasim. Additionally this excerpt demonstrates a point that most of the expert interviewees agree upon. It is worth noting, however, that with the exception of the previous excerpt, few interviewees offered an insight into various client working relationship histories but rather on the tone of that communication. This is important because it will inform the design of client-participant communication within the Promasim Simulation not only on grounds of what is said but the manner in which it is said.

Q: When you communicate with clients, who may themselves not have the same technical experience as yourself, how do you go about describing the important phases of the project's progress? Are there any techniques you usually employ?

I: Interesting question, I tend to do that...I probably do that automatically, I don't even think about it. My worry would be if I did think about it I'd start talking down to them and there's nothing that turns a client off so much as being talked down to as though they're an idiot and don't understand these technicalities.

So I tend not to have a script which is "Yeah, you're an idiot client, you have no idea about this. Sorry about that I'm going to give you the idiot's version." I usually give them the version we work with, that the team works with, the reality of what the team's doing. I may, as much as possible, translate whilst talking. So I'll tell them I've completed the feasibility study, I've completed the systems analysis, we've completed the entity relationship diagram, what this means is... We know what the data is. [...] I would rather, because of the time, you must ask me questions and say something when you're thinking "What's that?". You really must come back to me straight away and say "What's happening dude? Spell this out for me. Dumb it down if you have to." So I tell them that, but I wouldn't initiate a dumbing down ever.

This would seem to suggest that, assuming this protocol is implemented, it would be best to leave the matter of tone in communications with the clients themselves. As discussed previously there is scope for client communication needs as a metric to set the difficulty curve of the entire simulation itself. In doing so variables for understanding, engagement, and rapport could be artificially determined based on expected levels and a desired challenge rating all of which could define the kinds of interactions the participant has with the client point of enquiry. Great care and attention would have to be taken to ensure that these interactions remain within the bounds of realistic expectations, as described by the experts, and in terms of designing suitable mechanisms for Cumulative Malpractice Flags and Wrongness Flags to be utilized within them for triggering assessment and feedback with these points of enquiry.

The next task in this design process will be to determine the design remits for the team member and client points of enquiry in terms of their communication protocols and feedback mechanisms, their node threshold functions, and finally their sample dialogue trees for different environmental variables at different stages of the Simulation. This is to be achieved with the ideas expanded upon in this section as well as utilizing other feedback from the expert interviews in a design context.

Once that is achieved these elements can be prototyped and tested as part of the ongoing refinement process of establishing feedback mechanisms from within the Wrongness Framework as a suitable case study.

5.4. Design implementations

To begin it is worthwhile establishing some expectations for this design section. Firstly the focus of this section is to provide the designs for the Point of Enquiry characters that would exist within Promasim. The reason for focusing on these characters rather than on any other element within the Simulation is simply that as Point of Enquiry characters they are responsible for providing direct and indirect feedback to the participant and that feedback mechanisms are the focus of this case study.

Secondly, the scope of the design for these characters is limited insofar as their appearance, the formatting of the their feedback, and the exact user interface means of their interactions is not explored. The study of agency, anthropomorphisation, characterisation and depiction, and user interface design are topics which are far too large to incorporate in any meaningful way in this study and would serve only to pull focus from the demonstration of the background mechanics of the Wrongness Framework. To this end the designs as described in the sections below will, where necessary, account for expected output, accepted inputs, and the sources by which that information is gained – be it variable triggers, direct input, etc. - but, it will not account for the interface designs which will facilitate the transfer of that data. This will not, however, prevent the types of interfaces which would be used from being named so that it might make for a design explanation which is clear and easier to understand.

5.4.1. Deriving general parameters for Promasim non-player character design from the findings of the expert interviews

The findings from these expert interviews do not, themselves, provide for the creation of rules for the knowledge bank in the same way that the findings from various texts did for AdQuest. Instead they are to be used to identify in-game events which ought to take place, the interactions between different in-game non-player characters, and to describe the variety in quality of those interactions; e.g. tone, mannerisms. Properly, and according to questions that were asked in the expert interviews (found in Appendix B), has been the task to describe how these findings should be utilized within the Promasim experience. These descriptions are themselves explored in latter parts of this chapter with respect to specific non-player characters. However, there are some general parameters that have been derived from these findings that should be used to influence all non-player characters in Promasim.

It would be useful to explain this focus on non-player characters before going further with this interview-to-game translation. As a Serious Game, Promasim is to provide participants with some experience in project management from the perspective of the management of people. This being the case the non-player characters in Promasim can be treated in much the same way as described for environmental variables in the Wrongness Framework – that is that small alterations will make changes to what feedback the participant receives and how the environment of the scenario changes. The types of changes that would take place in the non-player characters are behavioural in nature and this position matches the findings of the expert interviews with regard to specific non-player characters. However, mapping behaviours through actor profiling, of the kind described in the later relevant sections on specific non-player characters, is still largely in its infancy although certain techniques have been used to achieve some results; Artificial Neural Networks for learning styles (Bernard et al, 2015) and Bayesian Networks for behavioural profiling (Millán et al, 2015). A proposed system of perceptron arrays that makes use of similar techniques is described later in the chapter with a newly proposed correction mechanism for altering nodal weights.

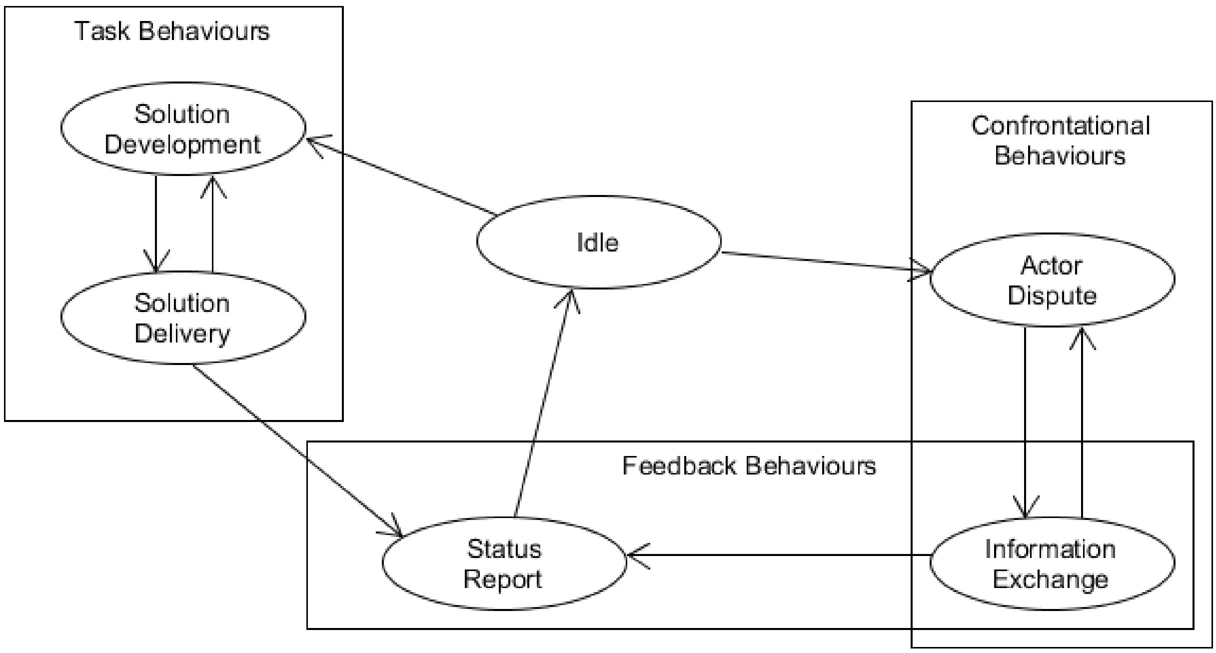
However, there are general parameters or constraint-like rules that might be derived from the expert interviews that are not specific to any particular actor. The first of these is that time-based activities ought to be consistently managed. That is to say that if any regular activity is to take place throughout a project life-cycle, for example team meetings, that there should be some uniformity to how those activities are managed. This parameter is taken directly from the expert answers given to questions one to four and question nine given in Appendix B regarding the timing of certain activities. This shifts the focus from time-based activities to their effects on the various actors within Promasim. Whilst not novel, in that as an attempt at the design of an epistemic Serious Game Promasim must try to reflect reality, the realisation of time-based mechanics for non-player characters does require some consideration.

Indeed it is this sense of time that provides the means by which the Wrongness and Cumulative Malpractice Flags could be created. In order to achieve this it is, therefore, necessary that the non-player characters of Promasim have some means of being influenced by the tracking of time and time disparities. The tracking of time itself can be designed very simply by using modulo arithmetic and the incremental tracking of some agreed unit; hour, day, etc. However the influence of that time on specific actors may be very different, or rather how time should be perceived by specific actors to provide a reasonable epistemic experience. By this it is meant that according to the perspective of certain stakeholders in a project the pressures of time and the management of time may be felt differently. A client may be less inclined to view a delay with a project as reasonable despite protests that the delay was unavoidable. A less-experienced team member may take the view that a period of time provides more opportunities than it truly does. The participant themselves will also, as project manager, bring their own perception of the passage of time to Promasim.

When developing the finite state machines for these non-player characters time-disparities can be used to alter a non-player's effectiveness or timeliness in accomplishing a given task – a specific example is given on how perceptron arrays might be altered to achieve this later. This information can also be used to alter other behavioural variables, such as mood, comfort, etc., which in turn can be used to alter in-game probabilities. The probabilities used in Promasim are used to dictate the success and degree of success a non-player character may have in completion of a task or in the likelihood that they will exhibit some behaviour.

In the former the application is simple to imagine: a given task has an associated success requirement, a non-player character has a base probability of success and that base probability can be altered, either positively or negatively, by a number of associated variables. The latter situation a non-player character may have various different behavioural states triggered based on a given probability. The effects of these probabilities are similar to those found in the work of Millán et al (2015) in that as conditional probabilities they are influenced and also influence other environmental variables. However, it shouldn't be assumed that a Bayesian network approach is being promoted in this instance as there does not exist the frequency statistics present in this case study to support such a network. The example below does provide some demonstration of how these probabilities may manifest in a non-player character's finite state machine.

For example suppose there is a simple non-player character referred to as team member *A*. *A* may have a number of character variables assigned to or accessible by themselves, such as stress_level, base_success_rate, performance_score, Task.expected_receipt, etc. Some of these variables, for example stress_level, may represent some coefficient for calculating the outcomes of a given behaviour or used in the transition between one state in the finite state machine and an other. Others, for example base_success_rate, may not be used by *A*'s finite state machine at all and instead are task interaction variables in the form of probabilities (as discussed above). Others, typically those accessed from other game objects, e.g. Task.expected_receipt, are used only for the calculation of behavioural outcome. This allows us to describe and map the finite state machine for *A* in a way similar to that given below:



In this simple representation of *A*'s possible finite state machine a small number of top-level states are shown in ellipses. These states are themselves confined within rectangle behaviour maps which roughly describe the purpose of each state. Each arrow between these states represents some transition and the direction of that transition between states. What should be clear is that not all states are interconnecting as to do so would require an expansive matrix of transitional calculations. Finally the state “Idle” is given as the default state and should be regarded as it were as the interchange of the states. Unlike typical finite state machine diagrams the transitional triggers have not been given for two reasons; clarity and to maintain abstraction so that the state machine might reasonably apply to any non-player character.

To illustrate suppose that *A* is some project team member assigned with the completion of some task. For the sake of this example suppose also that *A* is an experienced professional and has a success rate of 0.85 for the given task. *A* transitions from Idle to the Solution Development state perhaps through some simple boolean parameter. This boolean parameters may have been calculated by a prescribed algorithm taking account of a task request and time variable.

If *A* has the means to complete the task then the task success and outcome is described from some table of results on the outcome of some calculation involving their success rate probability, a transition is made to the Solution Delivery state to calculate a description of the outcome and then onto the feedback Status Report state to describe this outcome to the participant. At this stage the feedback might not be particularly complex, a simple notification to the participant that the task has or has not been completed and dated using the internal time-based parameters described before. The participant is free to act on this information as they choose, whether through reprimand or praise in which *A*'s internal variables may be altered.

Suppose instead in another example that *A* is unable to complete their task because they are waiting on some information from another non-player character or for that non-player character to complete some other task. Let us call this non-player character *B*. In this instance at the point where *A* would typically be completing their task it is highlighted within the action that they are awaiting some information from *B*, possibly held in some shared variable such as Task.expected_receipt. At the point of feedback and the return to the Idle state, as in the previous example, *A* would then trigger a transition to the confrontation state "Actor Dispute" which would, in effect, raise a complaint with *B*. From this point a transition to "Information Exchange" would occur wherein one of a number of possible things may occur. *B* might oblige by providing the, albeit late, information. They might instead flounder some more, causing their own stress_level variable to increase making *B* yet less likely to succeed. Therefore, it is the case where unless pressures on individual non-player characters are managed carefully it might be possible to overwhelm them and guarantee failure on their part thus impacting on each other team member. This is due in part that the variables of non-player characters are not fixed and altered by state interactions alone but may be affected by factors beyond their control. This illustration then provides a description for how elegantly such a system might be used to manage the performance of non-player characters within Promasim and that it would be possible to allow for complex and dynamic combinations of non-player characters and non-player character abilities within multiple instances of Promasim; in essence achieving one of the goals of the Wrongness Framework.

So far, in this brief initial description, a journey through the findings of the expert interviews, the translation of those findings into conceptual parameters of probabilities and time, and some notion of how these things may interact within Promasim via simple techniques has been presented. To realise these in-game parameters and variables into fully-fledged Wrongness and Cumulative Malpractice flags for feedback and assessment it is necessary to understand the constrictions of the Promasim context. Promasim is a Serious Game for the management of people within project management and, as such, much less attention is given to typical procedural tasks of project management, such as factor triangulation or risk assessment, than is given to the management of people as they complete those tasks. Feedback, as also demonstrated later in the chapter, can thus be provided using simple table systems of stock phrases and feedback compilation algorithms, such that complex feedback can be given if necessary. In the case of assessment, unlike AdQuest, Promasim provides what might be described as more implicit assessment. That is that the participant should, by means of the quality of the feedback they receive from the non-player characters, be able to realise their own success by the performance of those non-player characters and the perceived success of the project completion.

The design of each of these non-player characters must be considered to be a job of individual tailoring rather than cloning a default design for each type of character. This is necessary if only to allow for more complex and interesting experiences. Whilst individual non-player characters are discussed in greater detail later in the chapter it should be understood that each share common features and mechanisms such as those described here so as to provide the means that the Promasim experience might be created from the findings of the expert interviews and that a foundation for understanding each of those non-player characters in light of their shared and unique features can be established.

5.4.2. Project Team Members

The term project team member(s) in this section is used when referring to the project team member characters of the Promasim Serious Game rather than to real-world participants who may play Promasim in a team.

The most important aspect of the project team member design which should be tackled is the complex topic of their "learning"; that is to say the means by which the team member's quality variable might be improved. Fundamentally learning in this regard is any decision the participant makes which will result in the optimization of project activities such that there will be a beneficial outcome. This might be the case where there is a particular focus in raising the achievement of a team member or simply employing decisions and techniques which will improve the performance of a team member or team members generally.

Despite this goal it is important to make clear that this is not true learning as one might typically associate with other artificial intelligence techniques such as neural networks but rather the illusion of learning which can be facilitated, as one might expect, with similar node-based systems such as perceptrons. The advantage to using perceptrons in this case is that the there is already an established method by which they can be used to give this illusion of learning. What is of yet more importance is that there already exists a mechanism which can be used to correct the weightings of these perceptron nodes in favour of a desired outcome. Charles et al (2008: 25) describe this mechanism in terms of a simplified differential. The value E stands for the "sum squared error" for the particular nodal pattern and is given in the form:

$$E = \frac{1}{2} \sum_{P} \left(t^{P} - o^{P} \right)^{2}$$

Where $\mathbf{t^P}$ is the target output and $\mathbf{o^P}$ is the given output from the nodal pattern.

By simply taking a gradient (m) of E and the nodal weightings, Charles et al (2009) define the rules for changing the weightings to decrease E as:

$$m \in \mathbb{R},\ m \neq 0 : \begin{Bmatrix} m < 0 \rightarrow w = w + x \\ m > 0 \rightarrow w = w - x \end{Bmatrix}$$

The idea itself is perfectly simple and allows for the node weightings to be changed in such a way that it is possible to, in essence, eliminate error and, as such, raise the value of the team member character's quality variable to its desired and necessary amount. Here it is easy to see how this method of “learning” maps quite nicely to the scenario of mentoring a junior member of staff as described in the earlier section. In this way the value for E can be used to form the basis of feedback to the user on the success of the mentoring process and how the skill base for the junior team member is improving.

This is all well and good but the subtlety of the design must come from how the incremental variable x is determined. It would be tempting to make value of x some percentage of an average of other relevant variables such as the mentor's own quality, experience, and satisfaction variables as this would mean that when, in the scenario, the user is to assign the junior member of the team to a mentor there could be some variety and an opportunity for the user to demonstrate knowledge and forethought. This approach is also tempting as it is a particularly easy design to implement once values for those variables are set.

Indeed this implementation would suffice but whether or not it is an optimal solution is still yet to be shown. Equally there are some design decisions which would have to be made in light of its implementation. If a percentage approach was taken would a lower threshold of accuracy have to exist and would desired output have to exist within a larger range of values rather than a smaller, more precise range or single value? Assuming that to be the case could a junior team member undergoing this mentoring process ever grasp even the simplest task with 100% accuracy or is that a reasonable expectation at all? In any case would it guarantee fair and useful feedback for the participant's decisions?

One other way to settle this design issue would be to use an function derived asymptotic curve to determine the value of the incremental x variable. This could take the general form:

$$\text{Let:}$$

$$x \in \mathbb{R}; x > 1$$
$$a \in \mathbb{R}; a > 1$$

$$f(x,a) = \frac{x}{x \log_a x}$$

There is one distinct advantage to using this kind of value setting function and that is that through changing the logarithmic base via the variable a, the nature of the curve itself can be altered to a gentler or steeper incline as necessary. From a design perspective this enables Promasim to account not only for the summation of those other mentor character variables described above but also to account for the passage of time during the mentoring process. As such it would be possible to affect the effectiveness of a mentoring process not only on the basis of the capabilities of the mentor and mentee but also the circumstances surrounding that decision. Leave the decision too late and the effectiveness of the process is reduced. This would appear to satisfactorily account for the timing of decisions – with even the best will and talent available to the participant, only so much can be completed in a given amount of time.

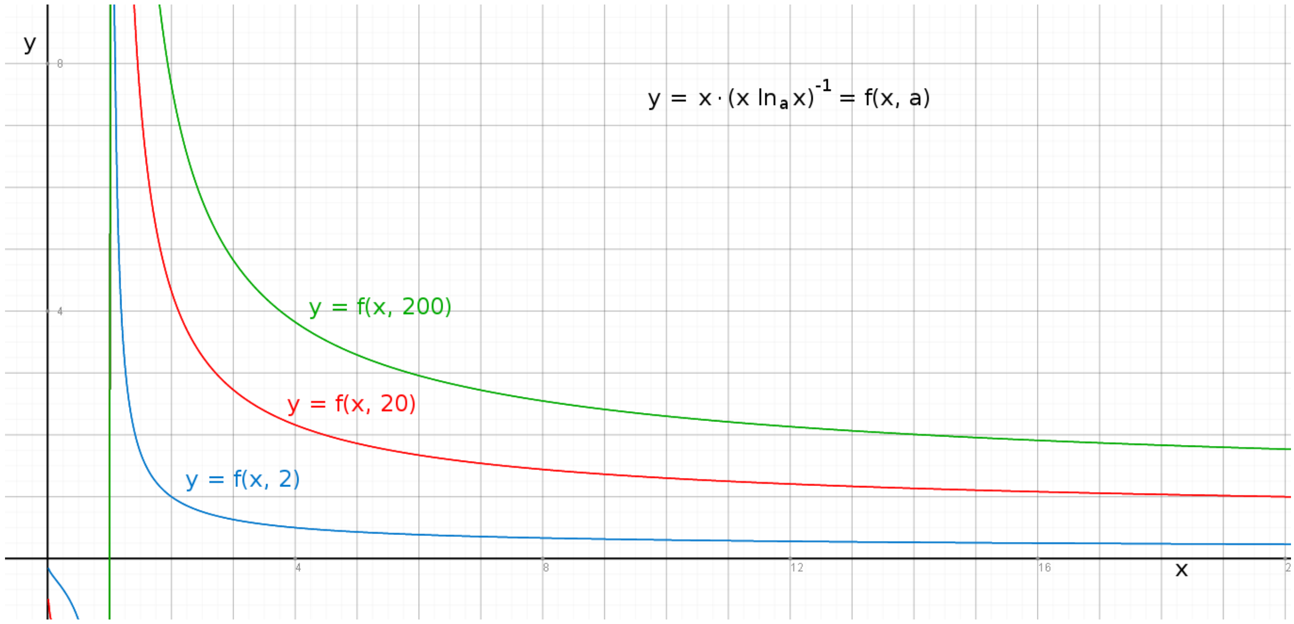


Fig. 5.1., *Graph of X-Increment Function at log$_a$, a = 2; 20; 200*

The graph in Fig. 5.1. shows the curves of the x-increment function, as described previously, when the base of the logarithm of that function is set to 2, 20, and 200 respectively. There is nothing particularly special in those figures themselves but they do show by increasing magnitude the steepness of the curve reducing as values of the variable "a" increase. This means that if one wished to apply more pressure to the function to give an illusion of time then a smaller number should chosen or a larger number if one wished to decrease the pressure.

This method is, of course, not limited only to the particular kind of mentoring scenario mentioned – though it provides a good example - but can be utilised throughout the team member non-player characters wherever variable feedback is required within a period of illusory progression. This could be anything from assessing and formatting for presentation system performance in the testing phase of software development in Promasim. Another possibility is that this method might even provide a suitable method for handling satisfaction management and project team interactions over periods of in-Simulation time.

In all of these instances the important innovation is the provided X-increment/decrement function as without the application of that function to perceptron alteration method the important time based factors which would effect in-Simulation feedback could not be achieved with such flexibility nor with such variance.

The manifestation of feedback to the participant using this method can take place in one of two ways. Firstly, setting aside the particulars of the feedback itself – which is to say its content; be that auditory, written, or pictorial -, the arrangement of that feedback can be easily managed by tabulating feedback consequences against an index. This is a satisfactory approach as it allows the index of the feedback, depending on context, to act as the desired target output for the sum squared error function and all other associated protocols. As such this approach can be standardised which is useful for making the case for a feedback mechanism in the Wrongness Framework with broad applicability.

The table below given below is a presentation of a hypothetical feedback table for the mentoring scenario as it is described above. Within this table there are two main columns: the index and the feedback consequences. As, for the purposes of illustration, the variables mentioned or functions called are for demonstrating the concept of this feedback mechanism then the content of the feedback consequences column should not be taken as being accurate or necessarily even appropriate feedback. Rather these changes in content are given in a way which is distinct enough that it is possible to discuss a collection of potential and hypothetical outcomes of this process and are not, necessarily, representative of the variables or functions which would be involved in the scenario.

Index	**Consequence**
0	var self.satisQuotient - = thetaThresholdVar var self.qualityVariable = self.qualityVariableVar func mentorIteration(x, y, z) → var timeAccount var iteration + = timeAccount
0.2	var self.satisQuotient - = thetaThresholdVar - X var self.qualityVariable = self.qualityVariableVar + Y func mentorIteration(x, y, c) → var timeAccount var iteration + = timeAccount
. . .	

. . .	
0.9	var self.satisQuotient + = thetaThresholdVar var self.qualityVariable = self.qualityVariableVar + Z func mentorIteration(x, b, c) → var timeAccount var iteration + = timeAccount
1	(Desired) var self.satisQuotient + = thetaThresholdVar + Y var self.qualityVariable = self.qualityVariableVar + Z func mentorIteration(a, b, c) → var timeAccount var iteration + = timeAccount

The tables themselves, as one can see, are fairly straightforward. As the perceptron array provides an output that output is checked against the table index for that interaction and a number of instructions are followed. For simplicity sake this example table shows a little pseudo-code as the consequence of certain outputs and it is arranged in a linear fashion where there are general and small improvements to the kinds of variables and functions used in performing the hypothetical interaction. Of course it needn't always be the case that the first row of the table is the least desirable result and the last row is the most desirable as certain interactions may require less optimal solutions for various reasons; e.g. sub-optimal working conditions, alterations to relationships, variations in work ethic.

Taking all of these points into account it is possible to design an iterative flow diagram in which feedback interactions are refined making small changes to the input values of the interaction perceptron array and reassessing the output. The small changes can occur as part of the expected consequence results from the feedback table or as part of a change in background environmental variables as appropriate. This approach is very similar to the more abstracted version of the assessment and feedback loop as detailed in Fig. 2.2. The diagram given below (Fig. 5.2.) shows this loop as it exists with all of the feedback components derived and discussed so far:

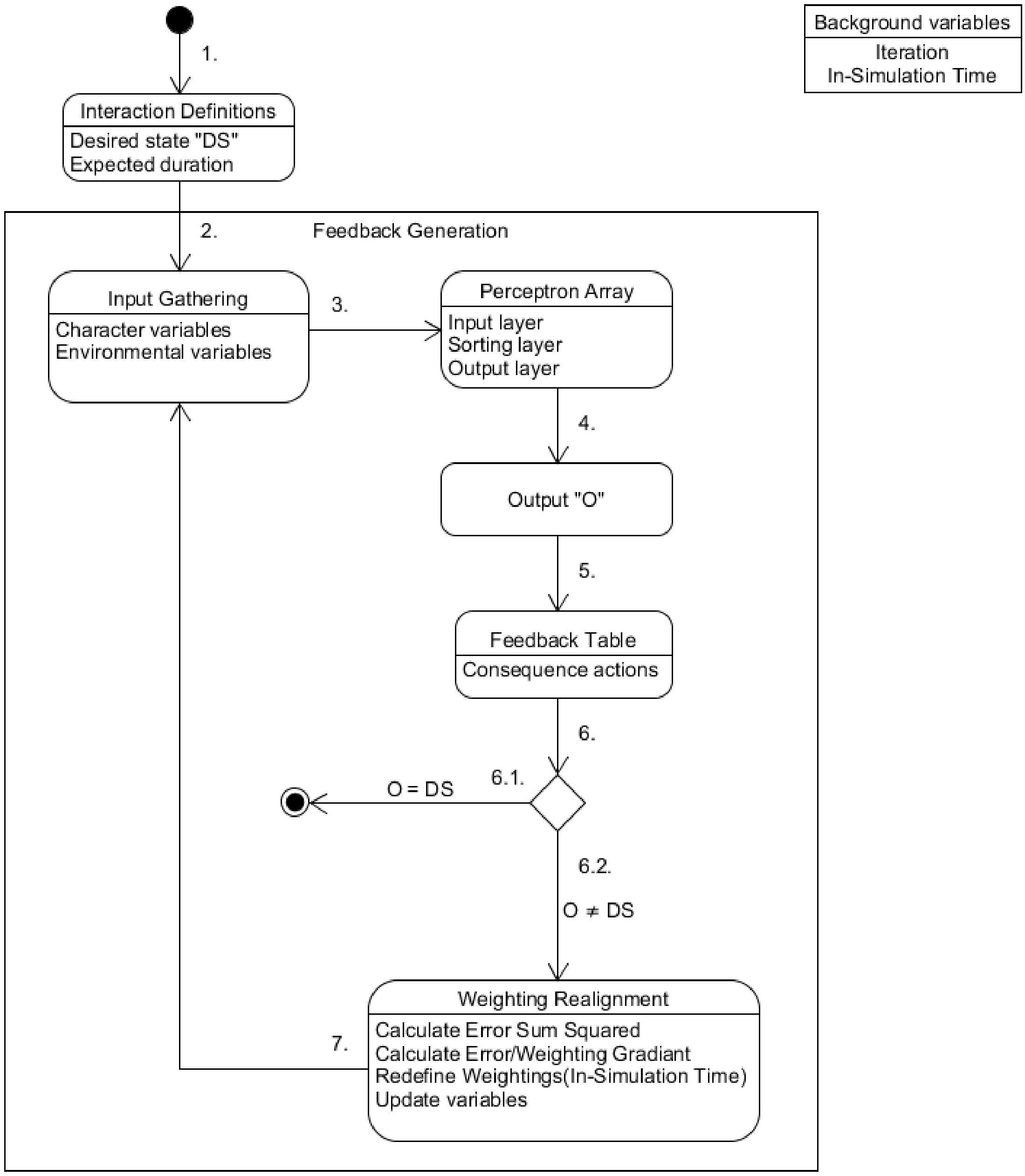


Fig. 5.2., *Iterative Feedback Generation Loop*

The diagram in Fig. 5.2. clearly demonstrates each of the independently and previously described elements of the proposed feedback mechanism in a single looping process and their place and dependencies in that process. The looping of these elements is a procedure that, with the exception of part 5 “Consequence Actions”, will largely be performed within the background mechanics of Promasim; its effects manifesting on screen only when called upon to do so during a general update of the Simulation state – apropos Fig. 2.2.

The background variables “Iteration” and “In-Simulation Time” serve as time keeping metrics for the entire process. As each interaction is undertaken as part of its definitions it will include an expected duration and a value of the maximum iterations of this process permissible within that expected duration. In this way when time updates occur within the flow process, values of the current iteration of the process can be used to affect other character based variables. This might be something as simple as a positive change in satisfaction when a task is achieved in less or equal time to the expected duration – demonstrating a quick learner, or a negative change in satisfaction for both mentee and mentor when a task requires more iterations than the expected duration would allow for – demonstrating frustration. All of this data can then be used to define and redefine impact variables which will demonstrate the concept of success or failure factors for the project through ongoing risk factors. If this approach is used it would be a simple matter to log the identification of this impact to a specific cause as it happens and then in a summary at the end of the Promasim experience, provide a full break down where failure or success factors were triggered.

Ultimately what should be taken away from this design is that, despite it being grounded within the solving of a particular team member training scenario for Promasim, the same technique can be used for other time-based activity feedback processes throughout the Simulation. In essence this covers all of the typical interactions of project team members within the experience of the project. All that is required is to determine what activities are to be included, to define through expert interviews the details and consequences of various possibilities of those activities, and to apply a time-based alteration parameter along the general form described in the previous section. Once these parameters are defined and used in this process then it will be possible to provide dynamic and participant-defined experience with the project team member characters during each experience of Promasim.

5.4.3. Project Clients

In general a similar approach can be taken with defining, triggering, and managing client-based feedback as it is in the manner described for project team members. As with the feedback table/Iterative Feedback Generation approach environmental and internal variables can be altered, outward Simulation appearances can be altered, and specific messages can be relayed to the participant in whichever form is deemed most appropriate for the situation. The fundamental difference between applying this technique to clients in the Promasim context rather than project team members is the difference in typical participant as project manager interactions with both parties.

Using time-based weighting iterations with client interactions directly would not be advisable. The time-based changes to perceptron array weightings represents the time constraints of an activity itself. Clients are, for the most part, not responsible for activities which have time constraints in the project life-cycle. This means that interactions with clients, instead, are mostly limited to ensuring that the client's agreed wishes are being fulfilled and within the given parameters of the contractual agreement by the project vendor. Clients are thus, for the most part, passive participants within a project. Of course this is not to say that all of their interactions are independent of time. For example, it would be difficult for a project manager who was undertaking or supervising the undertaking of a specifications analysis to complete a project within an agreed time period or budget if the client refused to assist in a timely manner.

At this point however it is necessary to make a design decision similar to the one made for conflicting team members. A client can, of course, be less than helpful in assisting where they are required to do so. No doubt this eventuality happens on a regular basis within project management. The question of whether including this type of interaction in Promasim would be beneficial is to ask if its inclusion would add any value to the Promasim participant experience. From an internal project perspective a client failing to assist where necessary is hurting only the success of the project which they are paying for and is ultimately damaging only to themselves.

Introducing the notion of self-destructive clients may have some grounding in the reality of project management but it is dubious as to whether that reality would train participants to be better project managers which is the ultimate design goal of a Simulation like Promasim.

Of course, it could be argued that such a situation is a valuable Simulation scenario as it would provide the user with the opportunity to make judgements based on when and how a project should end, penalty management, and other similar contractual difficulties to navigate. However, in doing so for a gradual learning experience it would be necessary to change the focus of the Promasim Simulation to one which is focused on that particular project management risk as it rather pushes the other project management activities to one side and reduces their impact. Instead it would be better, for the sake of activity variety, to instead assume that the client character of Promasim is as helpful as is necessary to achieve the goals of the project as they themselves, no doubt, would wish to see developments done on their behalf be successful rather than not. This must form the first client position for Promasim.

Referring to the transcripts of the expert interviews in the earlier section, it is plain to see that the satisfaction of a client, their confidence in the project management and individual project team members, and their willingness to comply with requests is itself based on the tone of communication established with the project manager. The particulars of the interaction between project manager and client are not overly important to the feedback mechanism but provided the interaction itself provides a number of different possible ways that the interaction could take place then the perceptron network approach can be used again in this context.

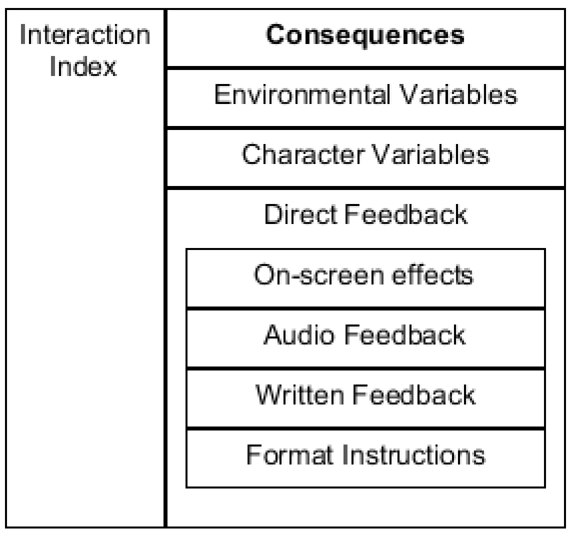
Not much needs to change to use the same iterative feedback generation loop as found in Fig. 5.2. or rather any such changes would be superficial changes. The only alteration that would need to be made would be that the direction of the output of that loop would instead go more or less directly into the background environmental variables which feed into it, making it more recursive than iterative, and that the connection between the time-constraint for the loop and the expected duration would be much more disparate. In this way it would be possible to take a metaphorical step-back from the Simulation activity in an attempt to view it from the outsider perspective; which would mirror that of the client's reasonably well.

In essence this is simply a case of splitting the effects of the alterations in design between the client variables and other environmental variables rather than having them dictate directly to a Simulation non-player character as is the case in team member designs.

5.4.4. Consequence Tables

Now that the mechanisms for providing feedback have been described it is worth devoting some time to exploring the design of the feedback, or consequence, tables. The description of them above as effectively a row with two columns, the first being the index value which is triggered by the output of the perceptron array, and the second containing all the instructions is useful enough as a general abstraction. However, the form of the second column – which contains instructions for the Simulation – can be standardised such that they could be applied to any situation.

The form of the consequence table row should be as the following:



Interaction Index

This is a simple numerical value with which is linked a number of consequence actions. The output from the output layer of the appropriate perceptron array for the activity is compared and, in the event that the index value and that output matches, or is within an accepted range, the consequences of that index are performed. As one might imagine the values possible for each interaction index for each feedback table are limited by the range of possible outputs available to the specific interaction array.

Environmental Variables

These are the background scenario variables which are either present in almost all activities or, if that is not the case, affect other character variables which are present in those activities. Their alteration must come first as recalculation of character variables will be a priority and anything which might affect those character variables must be accounted for before those new calculations take place. The environmental variables also include any logging procedures by which each iteration may provide further detailed feedback to all users.

Character Variables

Much like environmental variables, character variables are the alterations which will occur to the in-Simulation non-player characters for each output and each iteration of the feedback generation loop. As each change and new iteration of the loop occurs we should expect to see different results at points where $E \neq 0$.

Direct Feedback: On-screen effects

Direct feedback in the consequence tables are rare for each row as not every iteration of the generation loop will warrant direct feedback. Rather it is most likely that any severe error will be logged by the mechanism in the environmental variables. However there are instances, not necessarily within the desired index alone, where direct feedback will be given.

In this case on-screen effects refer simply to any change in animation, appearance, background, or indeed any visual change to the Simulation's appearance that might take place as a result of some activity.

Direct Feedback: Audio Feedback

Instances where recorded dialogue, sound effects, musical cues, and the like will be triggered by Simulation activities. This is much like the on-screen effects described above.

Direct Feedback: Written Feedback

Occasionally it may be necessary to provide more detailed and written instructions/activity analysis to the participant. Scripts can be pre-devised, selected, and processed in this manner.

Direct Feedback: Format Instructions

The last of the direct feedback consequences contains a simple set of instructions, where necessary, for the direct feedback given. As each Simulation and Simulation scenario is different so to will its feedback presentation have to be different in order to provide the epistemic approach to Simulations that the Wrongness Framework is based on. This means that there may be a context which would lead to a preferred presentation style for the direct feedback; such as in a case from Promasim where an update report should be emailed to the participant in the in-Simulation email system. Whilst text-formatting is a simple example to use to demonstrate this premise it is not limited to it and there is no conceivable reason as to why any of the other direct feedback consequences could not contain formatting instructions to achieve a specific goal within the Simulation provided that they were included in a way which was possible to achieve within the limitations of that kind of feedback.

5.4.5. Breaks, defaults, and desired states

The final aspect of the consequence rows is not defined within themselves or their consequence actions. That is that any index can be identified as causing a break in the feedback generation loop, or might be the default state for the activity, or may – as discussed previously in the loop mechanic section – be identified as the desired state for the activity; against which all error checking is performed. Here is where the application of Wrongness Flags and Cumulative Malpractice Flags come to the fore.

Breaks in the generation will be triggered when the reasoning managed by the Wrongness Flags dictates that an effect has occurred from which there can be no other recourse. For example, a misdemeanour or ethical transgression may result in a complete break in the current Simulation activity and a beginning of another to deal with that transgression. Likewise the default reasoning as determined by the development of the knowledge bank (Fig. 2.1.) will have to be identified so as to affect the initial state of the weights in the input layers where necessary. The desired state is as one might expect and its use is covered fairly extensively in previous sections.

5.5. Summary

The use of Promasim as a hypothetical Simulation to explore feedback mechanisms has been something of a tricky path to follow. It was established at the beginning of the chapter that the entire Promasim Simulation was used only to serve as specific examples of a more standardised feedback mechanism for the Wrongness Framework. As such references to the inclusion or not of those specific examples serve only as an attempt to provide examples which can demonstrate the elements of the feedback mechanism in as clear a way as possible.

These feedback mechanisms themselves are not new in their application with the Wrongness Framework. Indeed the user tested AdQuest Serious Game used the same mechanisms in a much simpler form – necessitated by the inclusion of feedback assessment by the participants themselves in the AdQuest results. The purpose of the Promasim hypothetical Simulation was to provide an opportunity to A) focus solely on the concept of these feedback mechanisms themselves, and B) to use a more complex example to explore and identify where these feedback mechanisms needed to be expanded upon. As a result of this exploration, with the assistance of expert interviews to frame the context of the Promasim experience and expected behaviours, new elements such as time-based and iterative adjustment mechanisms have been introduced which might not have been considered but now can form the standard model of application for feedback using the Wrongness Framework.

The advantages of this Promasim design study are that with an increase in dynamic complexity possibilities the outputs and feedback of Simulations designed with the Wrongness Framework can represent more complex training scenarios and the remit for learning can be likewise increased. This means that, in light of the conclusions of the AdQuest scenario – which itself fed into the design decisions of the Promasim design study -, it would be possible to go back to the AdQuest Serious Game and extend and improve not only the direct feedback mechanisms found within it to be as complex as its assessment mechanisms but also to extend the number and type of interactions a participant could experience with it. This would enable a case to be made to increase the number of challenges for the Wrongness Framework which are identified in its application methodology.

* * *

Chapter Six:
A Discussion of the Wrongness Framework

The purpose of giving challenges for the Wrongness Framework in chapter two of this thesis was itself three-fold in purpose. Firstly, it provided an opportunity to demonstrate a gap in knowledge and a position for the Wrongness Framework to fill. Secondly, it provided some criteria by which the Wrongness Framework could fairly be judged. Lastly, it provides an opportunity to assess where future work should be directed. It is, therefore, the aim of this, the last chapter of this thesis, to discuss the Wrongness Framework in reference to those challenges and to present issues which have come up during its initial conception and performance. However, It would be unjust to discuss the concept of the Wrongness Framework in terms of either failure or success against these challenges. After all the Wrongness Framework was not pitched as a design solution to these challenges as if they were some hypothesis for testing. Rather, it is the case that the design of the Wrongness Framework has been an attempt to bring together the understandings of various existing models and practices in the realm of Simulation and Serious Game design with the specific focus on providing new assessment and feedback mechanisms to incorporate, expand upon, and attempt to improve those practices.

Throughout the process of presenting and developing those assessment and feedback mechanisms there have been some issues or points of interest which have arisen which have, up to this point, not been discussed in the specific case studies of those mechanisms. This was not an attempt to dishonestly deal with issues and omit them from their relevant conclusions but rather to provide a platform where they might be discussed fully in more general terms and without the unnecessary complications of a case study context. It is thus the intent of this chapter to explore those issues generally, rather than the specific conclusions or summaries of the existing case studies, so that they might provide a focus point for further research with the Wrongness Framework and an opportunity to identify areas which might be refined. Ultimately this is to be achieved by identifying the various design considerations that should be taken at this time.

6.1. Assessing against the challenges

The first part of this evaluation is dedicated to discussing the applications of the Wrongness Framework and its general design in relation to the challenges which were defined for it in chapter two. The simplest approach has been to respond each of these challenges and present a position on each of them to describe the state of the Wrongness Framework as it currently exists.

The first two of these four challenges are similar in many respects and have, for the most part, validated by the designs found in chapters four and five. To summarise challenge A says that it should be possible to use the Wrongness Framework to incorporate the typical behaviours one would expect to see for a given scenario into a simulated version of that scenario. Challenge B is that it should be possible to use the Wrongness Framework to translate these typical behaviours into a form of symbolic logic for the sake of defining a rules-based system by using Wrongness and Cumulative Malpractice Flags as weightings so as to properly balance the consequences of those behaviours from within the Simulation. Now, it is to be argued that the two case studies of the previous two chapters demonstrate that those challenges have been met and the designs of the specific mechanics in those chapters show the method of meeting those challenges using the Wrongness Framework. However, that it is possible to meet these challenges in two specific cases does not mean that it is optimal to use the Wrongness Framework as the means to do so. To discuss that issue it is pertinent to regard these challenges in light of the others.

Challenge C is something which is more difficult to demonstrate. The case as to whether the Wrongness Framework has or has not increased the complexity, dynamism, range of possibilities for Simulation scenarios cannot be easily shown. Certainly the Wrongness Framework does not, as it currently exists, prohibit any level of complexity and, in fact, the consequence mechanisms described in the last chapter can be applied in such a way that any level of detail could be possible provided that the technology existed to support it. However, this is no proof by contradiction. What would be needed to clearly demonstrate the completion of this challenge would be to design, with the Wrongness Framework, a Simulation or Serious Game in response to an existing one, noting its limitations and expanding upon them. Only then a case might be made that this challenge was met.

Challenge D, on the other hand, has a more mixed analysis. The Wrongness Framework certainly does provide a means by which all of Aldrich's (2004) slates of educational experience – described in the first chapter – could be incorporated into a Simulation or Serious Game design. Much of that demonstration was shown during the AdQuest user tests. However, one such aspect has not yet been explored and that is the longer term Simulation/Serious Game without direct and in-room instructor contact.

This, fundamentally, is the end game goal of the Wrongness Framework but, though one could be designed and indeed the hypothetical Promasim Simulation does provide an opportunity to design a Simulation/Serious Game with no in-room instructor necessary, as this has still not been achieved. This should not necessarily be seen as a failure to meet the challenge. Rather, that in proposing a new design framework, the primary goal of this thesis is to create and develop the Wrongness Framework and establish the means by which Simulations could be designed to meet the gap identified in the initial research. Given the complexities of creating the foundation of the Wrongness Framework and the work which has been undertaken to ensure that the mechanisms are in place to achieve such complex Simulation/Serious Games, this, perhaps, should been seen as an area which should be further explored in a dedicated and independent study whereby such design issues might be tackled and further refinement to the basic structures of the Wrongness Framework can be undertaken.

6.2. The Devil on the Shoulder Problem

One of the more interesting design issues that has arisen during both of the case studies and in general development of the Wrongness Framework is, what is referred to in this thesis as, the devil on the shoulder problem. The devil on the shoulder problem is analogous to the legal concept of “leading the witness”. Simply it is the premise that certain design decisions when developing a Simulation/Serious game might be made that could influence the participant to make decisions or to interact with this Simulation/Serious Game in a way which they not have done so had those influences not been present. It is, in essence, the idea of suggestive design.

For example, suppose there exists some Simulation for a high-risk activity - like directing or monitoring air traffic – which is controlled by the participant through a monitor. Suppose still further that that Simulation is designed in such a way that mimics the environment of that activity, the types of interactions that activity has, and can fully simulate the sensation of that high-risk. If a person commissioning such a Simulation wished to use it as a means of training individuals and testing their reactions for stress, then the question becomes how those feedback and assessment mechanics might be incorporated into such a Simulation and how they would be represented. If, for example, the Simulation were to be designed with a large “press if you're stressed” button, the Devil on the Shoulder Problem would be to ask whether or not providing that button would be beneficial. In other words, would this button's inclusion facilitate the measuring of stress instances as they naturally occur within the participant or, by the presence of this button alone, does it suggest to the user that they should find aspects of the Simulation stressful. Unless this can be answered then the validity of the data taken from such a design component must be considered questionable.

It has been the practice of the application of the Wrongness Framework in the case studies discussed in this thesis to, where possible, hide these assessment and feedback requirements from the participant where possible. However, this has much to do with the subject matter of the case studies themselves and the way in which this is permissible and even practical. Other subjects when translated into Simulation form using the Wrongness Framework and with a focus on acquiring specific data may not enjoy that luxury. As such it would be beneficial for future phases of research with the Wrongness Framework to focus on answering the questions raised by the Devil on the Shoulder Problem and addressing any design concepts which may have to be employed in developing front-end interactions that can be used to tackle this issue of data validity. This would be a powerful future addition to the Wrongness Framework and would go some way toward further standardising its approach to Simulation/Serious Game design.

6.3. The Wrongness Framework and its place in games-based learning

In chapter one section five a case for further refinement was put forward on the basis of the surrounding literature in light of the slates of educational experience as defined by Aldrich (2004). In that case for further refinement the foundations of the Wrongness Framework was presented as a means to incorporating the various salient design points of some of the more well known positions on Simulations and Serious Games and framing them within these slates of educational experience. In this way the Wrongness Framework was to – as per the original research question in section 1.5 -, firstly, provide a design framework by which the all of these design decisions could be made by someone designing a Simulation or Serious Game for games-based learning and in a way which would allow that designer to benefit from this collated expertise. Secondly, by providing standardised assessment and feedback mechanisms the Wrongness Framework, within an expanded model of an intelligent tutoring system in the style of Charles et al (2011) and historically Sleeman and Brown (1982), that it would be possible for a designer to provide a level of autonomy to the Simulation/Serious Game itself in managing the progression of its users. What is more, that through the Wrongness Framework, that autonomy should be complex, dynamic, and refined to the user based on epistemic expectations of a scenario context, non-tokenistic, and reactive. The current position of the Wrongness Framework can be illustrated in the table below:

Design Feature	**Present? (Y or N)**	**Demonstrated where? [Chapter.Section]**	**Challenge Met**			
			A	**B**	**C**	**D**
Assessment mechanisms	Y	4.2 – 4.5	✓	✓		✓
Cumulative Malpractice Flags	Y	4.2, 5.4	✓	✓		
Wrongness Flags	Y	4.2, 5.4	✓	✓		
Feedback mechanisms	Y	5.2 – 5.5		✓	✓	✓
Knowledge bank development	Y	4.2, 5.2 – 5.3		✓		
Point of Enquiry Feedback	Y	4.1, 4.4, 5.3				✓
Directed learning	Y (In part)	4.2, 5.3 – 5.4			✓	✓
Tutor-free experiences	N					

The Wrongness Framework was designed and developed solely to act as a response to these needs. Through its presentation, analysis, and evaluation in the case studies presented here, the mechanisms for achieving these goals have been created and demonstrated to be suitable within some reasonable degree of limitation. It is from these foundations that future work can be undertaken to improve upon and validate the Wrongness Framework as a design approach.

6.3.1 Future projects

Part of the positioning of the Wrongness Framework in terms of both its position within the literature and its contribution to this topic is to describe the kinds of future projects which should be undertaken to further refine and improve the Wrongness Framework. It is necessary that the projects described below would be future projects rather than current case studies, as one might suppose, as to tackle these future projects would suppose that the foundations of the Wrongness Framework as explored in this thesis be established. Indeed, without having made the case for the current assessment and feedback mechanism designs it would be neither practical nor wise to consider the kinds of issues which are present in these future projects. As such their listing and descriptions here should not be seen as formal proposal of a specific project but rather, in general terms, an analysis of the kinds of topics which could be handled using the Wrongness Framework and to improve it.

Simulation/Serious Game with minimal tutor support

The first of these future project descriptions is really an expansion of what has already been achieved. One of the major reasons that the Wrongness Framework was developed was that the literature revealed that assessment and feedback which took place in the real-world rather than within the Simulation/Serious Game itself caused a break in immersion for the participants which, in turn, would affect their motivation, agency, and overall performance. As such the purpose of providing the Wrongness Framework, as the previous two case studies outline, was to help develop and refine automated mechanisms so that assessment and feedback can occur within the Simulation/Serious Game, rather than out of it.

This being said, even the AdQuest case study with its actual user tests and analyses required some tutor presence during its test run. Now as the feedback and assessment mechanisms of the Framework have been established it should be possible to create a Simulation/Serious Game within the guidelines of an intelligent tutoring system which can feedback, assess, and guide the learning of its participants without real-world tutor support. This does not necessarily have to be a Simulation/Serious Game based on particularly complex subject matter but it should be an autonomous experience. Even better would be to create such an experience that could be remotely assessed so that there is no possibility of interference from a real-world tutor during the experience.

Being able to make the case for the Wrongness Framework in a remote user tested setting would go some way to demonstrating its use in providing training and educational Serious Games/Simulations without direct reference to a tutor participant. This is advantageous not just because of the reasons of immersion, flow, and motivation as given above but also because it may provide the opportunity to provide educational experiences for participants where no localised expertise exists or is available. This type of project, therefore, could set a precedent for using the Wrongness Framework for designing Simulations/Serious Games for a user demographic which could benefit from such experiences in much the same way that early expert systems are reported to have done in non-training circumstances (Luger, 2005; Russell and Norvig, 2010).

Simulation/Serious Game design with specific cultural context

The AdQuest case study provided something of an exploration of this topic already as its localised assessment mechanisms, GUI, and feedback from the Point of Enquiry character required expert knowledge to incorporate into the Chinese language version of the game. However, the matter of cultural differences in presentation, tone of feedback, feedback styles, assessment styles, and the like are not the only interesting issues which might be addressed when designing with the Wrongness Framework. The context of the design itself may have to include specifics pertaining to all kinds of different demographic needs and an exploration of those needs may lead to more general refinement within the Wrongness Framework itself if a common factor can be found amongst all of them which is not already accounted for.

Testing the validity of subject predicates

Perhaps the most indirect use of the Wrongness Framework is to use its own feedback and assessment mechanisms as a way of validating the expert knowledge used to create the knowledge bank. In this way supposed ideas on typical interactions whereby the expected reaction to those interactions is known, it would be possible through user testing with a user group with some agreed level of expertise to use the Simulation/Serious Game and test as to which of the expected reactions are most likely or valid. This couldn't formally be used as a proof by negation or, dependent on design, proof by contradiction but any anomalies which might occur from the expected user behaviour would provide an interesting discussion point for further research in a given area. Thus, the focus of the Wrongness Framework moves from being a design framework to an analytical tool.

The creation of an engine for Simulation/Serious Games

The last of the possible future research projects for the Wrongness Framework listed here would be the creation of a game engine based on the Wrongness Framework mechanisms such that other designers could create and release their own Simulations or Serious Games. The creation of such an engine could provide the means by which the problem of growing educational needs and remote expert locales, as described in previous sections and notably in 1.5., could be addressed and some provisions could be made. Potential in this area could be quite large particularly if adopted by some industry in conjunction with a standards agency for that industry. To do so would ensure that training for workers in that industry at a given level could be standardised and monitored centrally and in such a way that expansion could be promoted and new markets explored easily and with fewer regulation and legislative issues.

6.4. Concluding thoughts regarding the Wrongness Framework

Within the bounds of this thesis the Wrongness Framework has been set forth as a reaction to a need identified in the literature of the topic of e-learning, its components developed to answer those needs, the mechanisms to achieve the Wrongness Framework have been designed, discussed, tested, analysed, and evaluated, and its position has been defined within past research findings and potential future research topics.

By and large the Wrongness Framework has met the challenges set for it and in the few instances where those challenges were not met in totality, proposals for refinement and expansion have been made such that those challenges might be met at a later date. The important task of providing the foundations for the Wrongness Framework has been achieved and demonstrated within this thesis and provides the foundations for the more difficult future research topics described previously. In light of this, it is justifiable to regard the Wrongness Framework, as it is proposed here, as being a successful design in terms of its research question proposed in section 1.5., meeting the remit of specifications which make up its challenges as a newly proposed and original design framework for Simulations and Serious Games and one which will provide new and interesting positions for tackling complex design issues in that field. This work provides a useful contribution to knowledge in gamification, by providing valuable lessons in the design of games with the use of Intelligent Tutoring System mechanisms in the background. So far this area has been little explored and this thesis provides a framework for more useful studies, elaborating on what has been demonstrated here.

WF wrongness framework

* * *

Appendix A: Rules for the AdQuest Knowledge Bank

The collection of rules given here are the rules which were used to form the knowledge bank of the AdQuest game and against which the assessment mechanisms checked the participant's interactions. For simplicity's sake the rules are given, grouped in order of the source from which they were taken. The citation for each group is given first and has been indented and made bold for clarity.

(Felten, 1947)

1. Objects should be well proportioned.
2. Mathematically obvious dimensions should be avoided.
3. Correct proportion is the regulation of the dimensions of an object such that the eye does not readily fathom the mathematical proportion.
4. In badly proportioned areas you should reduce the heigh or width with some ornamented device.
5. Unrelated sizes and shapes are to be avoided.
6. Balance should be in relation to the optical centre rather than the mathematical centre.
7. Groups of images may have one larger image for contrast but should not be enlarged such that relative proportions are maintained.
8. The range of tonal values should be reduced to a select small range to reduce unnecessary and distracting complexity for the eye.
9. Vertical lines draw the eye down.
10. Horizontal lines draw the eye across in direction the viewer would read.
11. Irregular shapes point toward the area it best emphasises.
12. A design is at its most logical, rhythmically, when each of its components conforms to the direction of reading for that culture (i.e. left-right-top-bottom in the Western world).
13. Alterations to components used in repetition should be made in such a way that is not distracting to the eye.
14. Component fundamentals - such as shape, tone, size, etc. - should be complimentary throughout the entire layout promoting unity of that layout.

Appendix A

15. Ornamented devices should make use of subtle shades of tones of existing elements of the layout.
16. Ornamented devices should also be relatively small in terms of comparative size for maximum effect.
17. There should be a minimum number of ornamented devices used in a single layout.
18. Coloured objects should be weighted heavier than black coloured objects.

(Landa, 2010)

19. Designs will follow a natural line curvature downwards.
20. Text should be of a readable size.
21. Information should flow from most to least important.
22. Highly ornamental typefaces should be avoided.
23. Images and text should maintain clear margins.
24. Either image or text should hold the "starring role" in a layout. Uniformity in both is to be avoided.
25. Image and text are subordinate to background image. Their component variables should not exceed 30% of those of the background image.
26. Product image and text should be placed together in the copy.
27. Images and text should occupy dead spaces which promote spatial depth.

(Checchinato et al, 2013)

28. Photos to use: Caucasian mostly – but it could be Asian as well.
29. Message/Text: No price information – if it is the case.
30. Chinese cosmetics advertisings mention the name of the model in the picture.
31. The ads are colourful - in the jewellery ads there are only coloured pictures because of the characteristics of the product.

32. Generally, ads show only one model and a female one.
33. Concerning language, it is interesting to note that brand name is translated in Chinese only once in the clothing, while in the cosmetics about 70% of brands are translated - for example, Chanel appears only as “Chanel”in the clothing, while in the cosmetics it appears both as “Chanel” and “香奈儿” (read xiang nai’er, which is the phonetic translation of the brand name).

(Hung et al, 2011)

34. Model ethnicity is considered as one of the experimental factors largely due to the fact that the use of foreign models is a distinctive feature of most magazine advertising in modern China today.

Appendix B: Promasim Expert Interview Questions

The following list of questions are those which were asked of all of the experts interviewed for the Promasim analytical case study:

Q1: How often would you prefer to meet/communicate with members of your team?

Q2: If you would meet/communicate more than once within a given period of time, would you arrange those meetings to discuss specific topics uniformly? For example, Tuesdays for progress reports, Fridays for scheduling.

Q3: Do you favour a specific time period – such as mornings, lunchtime, or evenings – for meeting/communicating with team members, and, if you do, could you expand upon the reason for that position?

Q4: When you document your meetings/communications with members of your team, either as a whole or singularly, do you typically provide copies of those documents to those involved or do you rely on your team members to account for their own documentation?

Q5: Suppose you have a junior team member with little working experience of a specific task whom you suspect needs additional training. How would you go about confirming that suspicion and what would be your ideal method for providing that training before that team member's inexperience had an opportunity to impact on the success of the project?

Q6: Suppose there are some personal difficulties between members of the team. Has your experience innovated any way of dealing with and resolving personal difficulties quickly and effectively?

Q7: Do you have a code of conduct that you expect your team members to work within? What would that code of conduct contain? (e.g., due diligence, fair presentation of skills and abilities, …)

Q8: Do you have a favoured delegation method? Is the task always given to the person with the greatest experience in that area or are there other important factors in your experience?

Q9: Are there specific moments within the lifetime of a project where you would call together a general meeting of all team members? If so what would those moments be?

Q10: Do you discuss team member aspirations as part of your duty as a project manager? Do you provide opportunities for team members to document their achievements as part of a portfolio for furthering their own careers?

Q11: When you communicate with clients, who may themselves not have the same technical experience as yourself, how do you go about describing the important phases of the project's progress? Are there any techniques you usually employ?

Q12: Assuming client dissatisfaction with the performance of your team, even to the point of withdrawing from the project, how do you manage the repercussions of failure amongst team members? Are there review procedures you would use? What would the outcome of such reviews be?

Q13: In your own words, how would you describe your leadership style?

Bibliography

Aldrich, C., 2004, “The four slates of educational experiences” in *On The Horizon*, 12(1), Emerald, pp. 14 – 17.

Aldrich, C., 2009a, *Learning Online with Games, Simulations, and Virtual Worlds: Strategies for Online Instruction*, John Wiley & Sons, Inc.

Aldrich, C., 2009b, *The Complete Guide to Simulations and Serious Games: How the Most Valuable Content Will Be Created in the Age Beyond Gutenberg and Google*, Pfeiffer.

Aleven, V., McLaren, B., Roll, I., and Koedinger, K., 2006, “Toward Meta-Cognitive Tutoring: A Model of Help-Seeking with a Cognitive Tutor”, *International Journal of Artificial Intelligence in Education,* 16(2), IOS Press, pp. 101-128.

Amory, A., 2007, “Game object model version II: a theoretical framework for educational game development” in *Educational Technology Research and Development*, 55(1), Springer, pp. 51 – 77.

Anetta, L. A., 2010, “The “I's” Have It: A Framework for Serious Educational Game Design” in *Review of General Psychology*, 14(2), APA, pp. 105 – 112.

Bernard, J., Chang, T., Popescu, E., and Graf, S., 2015, “Using Artificial Neural Networks to Identify Learning Styles”, *Artificial Intelligence in Education*, Springer, pp. 541 – 544.

Bouchet, F., Harley, J. M., Trevors, G. J., and Azevedo, R., 2013, “Clustering and Profiling Students According to the their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning” in *Journal of Educational Data Mining*, 5(1), JEDM, pp. 104 – 146.

Cen, H., Koedinger, K., and Junker, B., 2006, “Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement” in *Intelligent Tutoring Systems*, Springer, pp. 164 – 175.

Charles, D., Fyfe, C., Livingstone, D., and McGlinchey, S., 2008, *Biologically Inspired Artificial Intelligence for Computer Games*, IGI.

Charles, M., Bustard, D., and Black, M., 2009, “Game Inspired Tool Support for e-Learning Processes” in *Journal of e-Learning*, 7(2), Academic Publishing, pp. 101 – 110.

Charles, D., Charles, T., McNeill, M., Bustard, D., and Black, M., 2011, "Game-based feedback for educational multi-user virtual environments" in *British Journal of Technology*, 42(4), Blackwell, pp. 638 – 654.

Checchinato, F., Colapinto, C., and Giusto, A., 2013, "Advertising in a luxury fashion magazine: a comparison between Italy and China", *Paper 5*, *Department of Management*, Università Ca'Foscari Venezia. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2324443. [Accessed: Jan 2015].

Connolly, T. M., Stansfield, M. H., and Hainey, T., 2008, "Development of a general framework for evaluating games-based learning", *Proceedings of the 2nd European conference on games-based learning*, Universitat Oberto de Catalunya, pp. 105 – 114.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., and Boyle, J. M., 2012, "A systematic literature review of empirical evidence on computer games and serious games" in *Computers & Education*, 59, Elsevier, pp. 661 – 686.

Cowley, B., Charles, D., Black, M., and Hickey, R., 2013, "Real-time rule-based classification of player types in computer games" in *User Modeling and User-Adapted Interaction*, 23, Springer, pp. 489 – 526.

Dede, C., 2009, "Immersive Interfaces for Engagement and Learning" in *Science*, 323, AAAS, pp. 66 – 69.

de Freitas, S., 2006, *Learning in Immersive worlds: A review of games-based learning*, JISC.

de Freitas, S., and Jarvis, S., 2009, "Towards a development approach for serious games" in Connolly, T., Stansfield, M., and Boyle, E., (eds.), *Games-based learning advancements for multi-sensory human-computer interfaces: Techniques and effective practices*, IGI.

Dondlinger, M. J., 2007, "Educational Video Game Design: A Review of the Literature" in *Journal of Applied Educational Technology*, 4(1), pp. 21 – 31.

Dunwell, I., de Freitas, S., and Jarvis, S., 2011, "Four-dimensional Consideration of Feedback in Serious Games" in de Freitas S., (ed.), *Digital Games and Learning*, Bloomsbury, pp. 42 – 60.

Echeverria, A., Garcia-Campo, C., Nussbaum, M., Gil, F., Villalta, M., Amestica, M., Echeverria, S., 2011, "A framework for the design and integration of collaborative classroom games" in *Computers & Education*, 57, Elsevier, pp. 1127 – 1136.

Egenfeldt-Nielsen, S., 2004, "Practical barriers in using educational computer games" in *On the Horizon*, 12(1), Emerald, pp. 18 – 21.

Frank, A., 2007, “Balancing Three Different Foci in the Design of Serious Games: Engagement, Training Objective and Context” in *Situated Play, Proceedings of DiGRA 2007 Conference*, DiGRA, pp. 567 – 574.

Felten, C. J., 1947, *Layout*, Felten: New York.

Garris, R., Ahlers, R., and Driskell, J. E., 2002, “Games, motivation, and learning: A research and practive model” in *Simulation & Gaming*, 33 (4), Sage, pp. 441 – 467.

Gee, J. P., 2003, *What Video Games Have to Teach Us About Learning and Literacy*, Palgrave.

GMSA Intelligence, 2013a, *Sub-Saharan Africa Mobile Economy 2013*, GMSA.

GMSA Intelligence, 2013b, *Mobile Economy Europe 2013*, GMSA.

Habgood, M. P. J., and Ainsworth, S. E., 2011, “Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games” in *The Journal of Learning Sciences*, 20, Routledge, pp. 169 – 206.

Hung, K., Chan, K. W., and Tse, C. H., 2011, “Assessing Celebrity Endorsement Effects in China: A Consumer-Celebrity Relational Approach”, *Journal of Advertising Research*, Warc, pp. 6 – 21.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A., 2008, “Measuring and defining the experience of immersion in games” in *International Journal of Human-Computer Studies*, 66, Elsevier, pp. 641 – 661.

Johnson, W. L., Vilhjalmsson, H., and Marsella, S., 2005, “Serious Games for Language Learning : How Much Game, How Much AI?” in Looi, C., (ed.), *Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology*, IOS Press, pp. 306 – 313.

Ke, F., and Grabowski, B., 2007, “Gameplaying for maths learning: cooperative or not?” in *British Journal of Educational Technology*, 38(2), Blackwell Publishing, pp. 249 – 259.

Ke, F., 2010, “Examining online teaching, cognitive, and social presence for adult students” in *Computers & Education*, 55, Elsevier, pp. 808 – 820.

Kebritchi, M., Hirumi, A., and Bai, H., 2010, “The effects of modern mathematics computer games on mathematics achievement and class motivation” in *Computers & Education*, 55, Elsevier, pp. 427 – 443.

Kiili, K., 2005, “Digital game-based learning: Towards an experiential gaming model” in *Internet and Higher Education*, 8, Elsevier, pp. 13 – 24.

Kirkland, K., Ulicsak, M., and Harlington, M., 2010, *Games-Based Learning Experiences: Testing principals with teachers and students*, Futurelab.

Landa, R., 2010, *Advertising by Design: Generating and Designing Creative Ideas Across Media*, 2nd ed., Wiley.

Lopes, R., and Bidarra, R., 2011, "Adaptivity Challenges in Games and Simulations: A Survey" in *IEEE Transactions on Computational Intelligence and AI in Games*, 3(2), IEEE, pp. 85 – 99.

Lucas, K., and Sherry, J. L., 2004, "Sex differences in video game play: a communication-based approach", *Communication Research*, 31(5), Sage, pp. 499 – 523.

Luger, G. F., 2005, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 5th ed., Addison-Wesley.

Malone, T. W., 1980, "What Makes Things Fun to Learn?: Heuristics for Designing Instructional Computer Games" in *Proceedings, SIGSMALL '80 Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, ACM, pp. 162 – 169.

McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., and McVay, A., 2012, *A Literature Review of Gaming in Education: Research Report*, Pearson.

Millan, E., Jimenez, G., Belmonte, M-V., and Perez-de-la-Cruz, J-L., 2015, "Learning Bayesian Networks for Student Modeling", *Artificial Intelligence In Education,* Springer, pp. 718 – 721.

Mitrovic, A., Martin, B., Suraweera, P., 2007, "Intelligent tutors for all: Constraint-based modelling methodology, systems, and authoring" in *IEEE Intelligent Systems*, 22(4), IEEE, pp. 38 – 45.

Mitrovic, A., and Suraweera, P., 2000, "Evaluating an animated pedagogical agent", *Intelligent Tutoring Systems*, Springer, pp. 73 – 82.

Papastergiou, M., 2009a, "Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation" in *Computers & Education*, 52, Elsevier, pp. 1 – 12.

Papastergiou, M., 2009b, "Exploring the potential of computer and video games for health and physical education: A literature review" in *Computers & Education*, 53, Elsevier, pp. 603 – 622.

Pierce, N., Conlan, O., and Wade, V., 2008, “Adaptive Educational Games: Providing Non-invasive Personalised Learning Experiences” in *Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*, IEEE, pp. 28 – 35.

Robertson, J., and Howells, C., 2008, “Computer game design: Opportunities for successful learning” in *Computers & Education*, 50, Elsevier, pp. 559 – 578.

Roll, I., Aleven, V., McLaren, B. M., Koedinger, K. R., 2011, “Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system”, *Learning and Instruction*, 21, Elsevier, pp. 267 – 280.

Russell, S. J., and Norvig, P., 2010, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall.

Salen, K., and Zimmerman, E., 2003, *Rules of Play: Game Design Fundamentals*, MIT Press.

San Pedro, M., Baker, R., and Rodrigo, M., 2011, “Detecting Carelessness through Contextual Estimation of Slip Probabilities among Students Using an Intelligent Tutor System for Mathematics” in *Artificial Intelligence in Education*, Springer, pp. 304 – 311.

Santosh, A. M., and Koedinger, K. R., 2010, “Fostering the Intelligent Novice: Learning From Errors With Metacognitive Tutoring”, *Educational Psychologist*, 40(4), Routledge, pp. 257 – 265.

Shaffer, D. W., 2007a, *How Computer Games Help Children Learn*, Palgrave.

Shaffer, D. W., 2007b, “In Praise of Epistemology” in Shelton, B. E., and Wiley, D. A., (eds.), *The Design and Use of Simulation Computer Games in Education*, Sense, pp. 7 – 28.

Shelton, B. E., 2007, “Designing Educational Games for Activity-Goal Alignment”, Shelton, B. E., and Wiley, D. A., (eds.), *The Design and Use of Simulation Computer Games in Education*, Sense, pp. 101 – 127.

Sleeman, D., and Brown, J. S., 1982, *Intelligent Tutoring Systems*, Academic Press.

Steenbergen-Hu, S., and Cooper, H., 2014, “A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning” in *Journal of Education Psychology*, 106(2), APA, pp. 331 – 347.

Syngenta Foundation, 2011, “Mobile Applications in Agriculture”, Syngenta Foundation [online]. Available at: http://www.syngentafoundation.org/__temp/Report_on_mAgriculture_abridged_web_version.pdf [Accessed: March 2014].

Thackray, L., Good, J., and Howland, K., 2010, “Learning and Teaching in Virtual Worlds: Boundaries, Challenges and Opportunities” in Peachy, A., Gillen, J., Livingstone, D., and Smith-Robbins, S., (eds.), *Researching Learning in Virtual Worlds*, Springer, pp. 139 – 158.

Thomas, J. M., and Young, R. M., 2010, “Annie: Automated Generation of Adaptive Learner Guidance for Fun Serious Games” in *IEEE Transactions on Learning Technologies*, 3(4), IEEE, pp. 329 – 343.

Ulicsak, M., and Wright, M., 2010, *Games in Education: Serious Games*, Futurelab.

United Nations, 2013, *Millennium Development Goals and Beyond 2015*, UN [online]. Available at: http://www.un.org/millenniumgoals/ [Accessed: February 2014].

Van Eck, R., 2007, “Building artificially intelligent learning games”, *Games and simulations in online learning: Research and development frameworks*, IGI, pp. 271 - 307

VanLehn, K., 2006, “The Behavior of Tutoring Systems” in *International Journal of Artificial Intelligence in Education*, 16(3), IOS Press, pp. 227 – 265.

van Staalduinen, J.P., and de Freitas, S., 2011, “A Game-Based Learning Framework: Linking Game Design and Learning Outcomes” in Khyne, M. S., ed., *Learning to Play: Exploring the Future of Education with Video Games*, Springer.

Walonoski, J. A., and Heffernan, N. T., 2006, “Detection and Analysis of Off-task Gaming Behavior in Intelligent Tutoring Systems” in *Intelligent Tutoring Systems*, Springer, pp. 382 – 291.

Whitton, N., 2010, *Learning with Digital Games: A Practical Guide to Engaging Students in Higher Education*, Routledge.

Yaman, M., Nerdel, C., and Bayrhuber, H., 2008, “The effects of instructional support and learner interests when learning using computer simulations” in *Computers & Education*, 51, Elsevier, pp. 1784 – 1794.