Relating Cognitive Models of Computer Games to User Evaluations of Entertainment

by

Paolo M. Piselli

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Computer Science

by

____
Date

APPROVED:

____
Dr. Mark Claypool, Major Advisor

____
Dr. James Doyle, Advisor

____
Prof. Gary Pollice, Reader

____
Dr. Michael Gennert, Head of Department

Abstract

As the interactive entertainment industry matures, we need a better understanding of what makes software entertaining. A natural starting point is the application of traditional Human-Computer Interaction (HCI) tools to interactive entertainment software. Cognitive models are tools that HCI researchers have used to model users' thought processes and evaluate interface design. With this research we investigate the relationship between the complexity of an interaction and the entertainment experienced by the user. We designed a simple computer game, created a normative model for how a user plays this game, built several variations of this game such that normative models of these variants differed across two factors: pace and complexity. User studies were conducted on these variations, and we compared these factors to user performance and self-reported user enjoyment.

Contents

1.Introduction

Over the past few decades, entertainment software has grown from a cottage industry to a multi-billion dollar industry. According to a Jan. 2006 press release by NPD Research, U.S. retail sales of game consoles, handheld devices, games and accessories reached $10.5 billion in 2005. As the industry matures, developers will naturally have a competitive interest in producing games that offer a more entertaining experience to the end-user; however there has been limited research into what makes a computer game entertaining.

Although specialized for the purpose of entertaining one or more human users, a computer game is still software and thus provides the opportunity for study by computer scientists as software. The science of Human-Computer Interaction (HCI) is relevant to such a user-focused domain. But what can HCI methods say of the entertainment value of a given piece of entertainment software? One informal suggestion is that a game must have a certain computational complexity in order to be interesting [E1]. The problem with this theory is that it speaks of the complexity of an algorithm used to compute a solution to a game, rather than the thinking a human player does while playing a game. If humans go about playing games in different ways than computers compute solutions to games; then there may be more accurate measures for what interests a human than computational complexity.

An evaluation of the user experience must look at interest and complexity from the user's point of view. Studies indicate that there is a correlation between cognitive interest and enhanced learning [HM1997]. It is possible that the cognitive features of a game also correlate with enjoyment. Cognitive models, tools that model user thought processes, may take a variety of forms that capture thinking, such as decision trees or production rules. Cognitive Modeling has been used in HCI research for over two decades, but to the best of our knowledge has yet to be applied to the study of user enjoyment of computer games. As a quantifiable representation of a user's thought process, a cognitive model is an artifact that we can measure more directly than the user's actual thoughts. We have investigated the relationship between a cognitive model of an interaction and the entertainment value that a user places on that interaction.

This research explores the use of cognitive models in the evaluation of entertainment software. We built four variations of a simple computer game, created normative rule-based models of how users play these variations, and conducted user studies to determine the relative entertainment value of each variation. The game was made simple enough so that normative models for play could be constructed apriori. The variations were designed such that their models would differ (by the number of rules in the model and by the rate at which these rules are used.) User studies were conducted with these game variants using college-aged men and women. Subjects played a randomly chosen game variant then filled out a post-test questionnaire. The resulting data was then analyzed to determine if the size of the active-set of rules and rate of rule-use in the cognitive models can be related to user self-evaluations of entertainment.

An analysis of our data indicated that complexity variable had a statistically significant relationship with user perceptions of difficulty. Subjects' in-game enjoyment was significantly related to both the subject's performance and the complexity variable. Our data showed evidence of an "inverted-U" phenomenon when comparing enjoyment to performance. We also noted a difference between the enjoyment that subjects reported in-game and the enjoyment that was reported in the post-game questionnaire.

The rest of this document contains several sections. Section 2 discusses work related to this research. Section 3 describes our approach to designing our game and our experiment. Section 4 presents an analysis of our user study data. Section 5 discusses issues that arose and lessons learned over the course of research. Section 6 summarizes our conclusions. Section 7 looks forward to future avenues of investigation.

2.Related Work

2.1.Cognitive Psychology

Nearly a century ago, Yerkes and Dodson's famous experiment predicted an "inverted U" relationship between level of arousal and performance in a memory task [YD1908]. The Yerkes-Dodson law predicts that there is a level of emotional arousal that optimizes a subject's rate of learning on some task. With too little or too much emotional arousal, the subject's rate of learning will be diminished. Our research also investigates emotional arousal, but the specific type of emotional arousal that a subject judges to be "enjoyment". We also seek to investigate this type of arousal as a variable that is dependant on factors of an interactive experience.

Burns proposed the use of Bayesian-information theory to analyze enjoyment in a game of slots [BU2006]. Burns' work models fun as resulting from information gain from a violation of expectations. This model also predicts "inverted-U" functions, referred to as "Goldilocks Functions", relating fun to win-probability. Our research attempts to relate fun to cognitive complexity. If cognitive complexity impacts win-probability, we may see evidence of a Goldilocks Function.

2.2. Cognitive Modeling in HCI

Cognitive Models have been in use as evaluative tools in HCI for over two decades. One of the most popular ways of modeling has been GOMS, which codifies a cognitive task as Goals, Operators, Methods and Selection rules. John and Peck used Soar, a GOMS model, to create a computational model of the task of browsing a database [JP1992]. Their work shows that GOMS models can capture the cognitive elements of even a highly interactive task. John and Kieras, in their overview of GOMS models, describe ten successful real-world applications of a GOMS model to user interface design issues [JK1996].

Another popular cognitive modeling tool is the ACT-R family of cognitive architectures. Belavkin applied ACT-R to model the inverted-U phenomenon described by the classical Yerkes-Dodson experiment [BE2001]. This work modeled emotional activation and found that a medium level of activation resulted in a higher simulated rate of learning. Byrne has applied the ACT-R/PM architecture to modeling the interactive task of menu selection [BY2001]. ACT-R/PM was a cognitive modeling system which had an additional perceptual-motor module to enable modeling interactive tasks to a higher level of detail. Aside from GOMS and ACT-R, Chery and Farrel have described HCI research involving Perceptual Control Theory [CF1998], which may be better suited to modeling tasks of continual adjustment.

2.3. The Science of Entertainment Software

The study of entertainment software is a nascent field. Vorderer, Klimmt and Ritterfeld have presented an integrated theory of media entertainment that accommodates for both the user and the media [VKR2004]. Their model of Complex Entertainment Experiences (CEE) is a conceptual model that defines enjoyment as a user's response to a media product based on the particular user, motives, and the particular media. In this model, it is suggested that all three of these must meet certain prerequisites in order for an experience to be viable as entertainment; however a precise enumeration of these prerequisites is not given.

Vorderer et al. divide the properties of a media product into four categories: technology, design, aesthetics, and content. They state that a product with a certain level of interactivity may bore or interest a user depending on the particular content and the particular user, and that users seek a certain emotional complexity to their experiences. Their survey of other research indicates that human motivations for playing video games range from presence to self-efficacy to competition. Our research focuses on the complexity of the cognitive aspects of an interaction. In the light of Vorderer's integrated theory, this research asks the question: "given the same users and the same content, to what degree do the purely cognitive aspects of entertainment software affect the users' enjoyment?"

3.Approach

3.1.Overview

This research investigates how the cognitive complexity of playing a computer game influences how entertaining it is for a human user. In order to test this hypothesis, we chose a model for human cognition, developed several variations of a computer game, created simple models of the cognitive processes of playing each variation, and conducted user studies to determine if there is a difference in how entertaining each variation is for users. We chose to model human cognition of a game as a rule-based system. The factors of complexity that we chose to investigate included total number of rules in the model, and frequency of rule activation. These factors were investigated in the context of a simple "Punch-Out!" style game. An original game in this genre was developed, including original artwork. Four variants of the game were made such that the normative cognitive models of these games would vary along the two chosen factors. User studies were then conducted using a between-subjects 2x2 randomized experimental design. Users gave self-evaluations of entertainment both during and after their testing sessions. Finally, data from the user studies was analyzed to see how different levels of game complexity affected user-reported levels of enjoyment.

3.2.Motivation of Design

This research attempts to relate the complexity of thought needed to play a game to how entertaining it is by relating cognitive models to user enjoyment. For our results to be meaningful there needed to be some certainty that the cognitive models matched the thinking of the human subjects. Typically, a cognitive model is developed through Cognitive Task Analysis (CTA), a process which may involve observing subjects as they perform some task, asking subjects to "think aloud" and describe their thoughts as they perform the task, or interviewing subjects about the decisions they make under certain circumstances.

With limited time to complete this work, we chose to investigate a style of computer game which had game-play straightforward enough that game-play cognition could be modeled without lengthy CTA. Such a game would give the user limited choices; otherwise it could involve decision making that varies from subject to subject. A game where the proper action to take is clear for every condition that occurs would challenge the player's memory, dexterity, concentration, and reaction time rather than their decision making ability.

Electronic Games such as *Simon* (© 1978 Milton Bradley) are purely reactive. In the *Simon* game, the player is challenged to repeat a sequence of colors made by the device. For each sequence presented by *Simon*, there is only one proper reaction: to repeat the given sequence. By investigating a game with a reactive style of game-play we could presume that a model for play that captured the proper reactions to all conditions would represent the normative play style.

In addition to being purely reactive, we needed to be fairly certain that subjects had an understanding of the game that matched the normative model. Although the process of learning to play may be a part of the entertainment value of a computer game, we wished to focus our study on other factors. This meant that subjects would need to learn the proper actions for all conditions before testing. By keeping the number of conditions small, we would be able to quickly train subjects and be fairly confident that they had a complete understanding of how to play.

3.3.The Game Design

3.3.1.Punch-Out Style Games

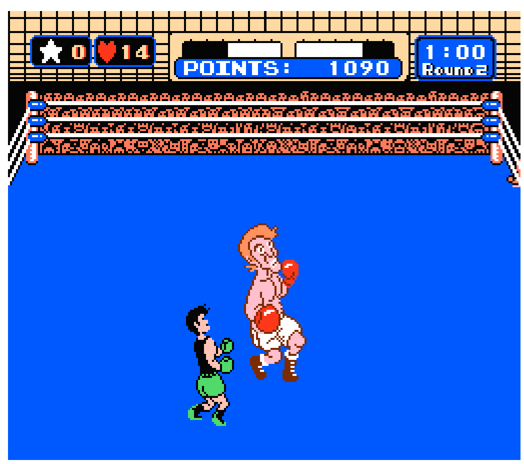

Figure 1: Punch-Out © 1987 Nintendo

We chose to base our game on a reactive play style similar to that of the game *Punch-Out (Figure 1, © 1987 Nintendo)*. *Punch-Out* is a simple abstraction of a one-on-one boxing match. The user has control over one of the boxers, and can execute a limited set of punch, block and dodge actions. The computer controls the actions of the user's opponent. There is very little strategic planning on the part of the user. The game task is mostly reactive - the user must recognize a certain action executed by the opponent and respond with an appropriate sequence of actions within a limited amount of time. For example, if the opponent raises his right arm to punch, the proper action might be for the player to dodge to the opposite side then counter-punch to the mid-section. By using this style of game-play, we were able to control the number of conditions that could arise, and what the proper response to each would be.

3.3.2.Paolo's Kickboxing

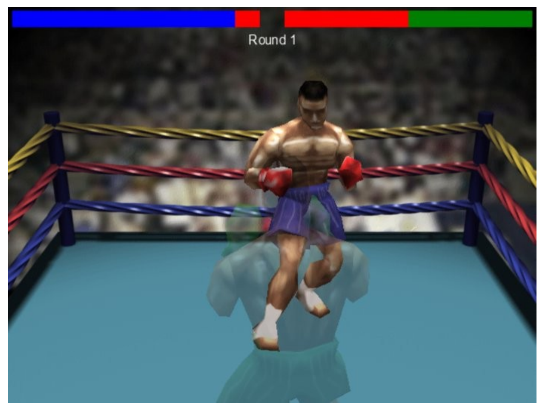

Figure 2: Paolo's Kickboxing

Similar to *Punch-Out*, our game involved two on-screen avatars, one user-controlled and one computer-controlled (Figure 2). The game consisted of ten "rounds" of play, where each round was won by the first boxer to score ten hits in the round. Both avatars started in a neutral state, and were be able to execute any of their available actions from this neutral state. The user interface involved pressing keys that directly correspond to actions. Execution of an action caused the relevant avatar to transition to a different state, such as blocking or punching, followed by a return to the neutral state.

The proper action for the user in the neutral state would be to wait for the opponent to take an action. If the user attacked while the opponent was in a neutral state, the opponent would automatically block the user's attack. When in the neutral state, the computer would delay for a set amount of time then randomly select an attack to execute. The rate of the computer's attacks was varied to control the pace of the game. The number of attacks which the computer could randomly select was varied to control the complexity of the game. When the computer made an attack action, the user had a one-second interval within which to recognize the attack and perform an appropriate block action. If the user successfully blocked the opponent's attack, the opponent would be vulnerable to counter-attack for a one-second interval. These generous response intervals were given because we did not want reaction-time to be a large factor in the difficulty of the game.

Hick's Law predicts that the choice response time given N equally probable choices is logarithmic in N [H1952]. This model of interaction is applicable to our game. However, empirical research is required to estimate the parameter in Hick's equation, and the parameter can differ from individual to individual. For our research, we chose a response interval of one second for all test groups based on an informal pilot study. We chose this interval because trained subjects could easily respond within this time frame.

3.4. The Game Variants

We built four variants of the game such that normative models for game play would vary across two factors. These factors were the "rate of decisions", or pace, and "number of choices", or complexity, of playing the game. The first dimension of variation was the pace at which the computer player made attacks. The second dimension of variation was the total number of possible conditions – i.e. the number of possible attacks executed by the opponent. Informal user testing was conducted with graduate students as the game was developed to determine the two levels for each of these factors.

We hoped to capture a level that was challenging and a level that was not challenging for each factor. For the pace factor we chose to have an opponent that would attack every two seconds and an opponent that would attack every four seconds. For the complexity factor we chose to have an opponent capable of executing four types of attack and an opponent capable of executing eight types of attack. The attacks took the form of punches and kicks. The four-attack opponents had one punch attack and one kick attack on both the left and the right sides. The eight-attack opponents had high and low punch attacks and high and low kick attacks on both the left and right sides.

3.5.The Cognitive Models

3.5.1.Rule Based Systems

A rule-based system represents decision-making as a collection of IF-THEN rules that state that IF certain conditions are true, THEN to perform certain actions. The system has a working memory that maintains assertions about the current state of the world called facts. A rule's conditions make logical statements about these facts, and when these statements become true, the rule activates, and its actions are performed. A rule-based system is sufficient for modeling our game because our game involves the user taking discrete actions in response to actions by the opponent. Rule-based systems may not be good for modeling some types of games because rules are not good at modeling tasks involving continuous control.

By representing normative game play cognition as a rule-based system, we can quantify the game play task in various ways. We can count the number of rules there are in the entire model. We can determine how many facts must be kept in working memory. We can tell how frequently rules must be activated, and how many rules might apply at any given time. For each rule, we can time how long the model has for working memory to be updated, conditions to be matched, and the rule executed. If our model accurately captures human thought, then these quantities represent the demands the game places on the user's cognitive faculties.

3.5.2.Modeling Assumptions

We assume that rules in a model that are always executed in sequence would collapse into a single rule with a concatenated sequence of actions. Because the opponent always becomes vulnerable after a successful block action, the two separate rules "if I am attacked, then block" and "if the opponent is vulnerable, then punch" would collapse into the single rule "if I am attacked, then block and then punch". Thus we assume that all "block, punch" action sequences are the result of the activation of a single rule.

If we are going to measure features of our rule-based model, then the way in which we write our rules will influence the resulting numbers. For instance, the single rule “if A or B, then do C and D” is functionally equivalent to having the two rules “if A then do C and D” and “if B then do C and D”. For a more complicated model, we would need a well-defined method of counting rules based on unique sets of conditions resulting in unique sequences of actions. However, our game is so simple that the arising conditions are mutually exclusive. We therefore assume that there is exactly one rule in the model for each condition that results in a specific sequence of actions.

Our game also has the property that there is only one proper sequence of actions to take for each condition. Thus there is exactly one rule in the model for each condition. Finally, the only conditions that arise that are appropriate for the user to respond to are attacks by the opponent; therefore there is exactly one rule in the model for each of the opponent’s attacks.

3.5.3. Characterization of Paolo’s Kickboxing as a Rule-Based System

Given the stated game design, the game play in Paolo’s Kickboxing can be described by a rule-based model which has a certain number of rules, and time constraints for how quickly and how often these rules must operate. In all cases, the firing of a rule must take place within one second in order for its action to be successful. The number of rules in the model for a given variant is equal to the number of attacks the computer opponent can execute in that version of the game.

For the four-attack variants, there are four rules in the model:

```
IF the opponent punches towards my left side
      THEN press '7', press SPACE-BAR.

IF the opponent kicks towards my left side
      THEN press '1', press SPACE-BAR.

IF the opponent punches towards my right side
      THEN press '9', press SPACE-BAR.

IF the opponent kicks towards my right side
      THEN press '3', press SPACE-BAR.
```

For the eight-attack variants, there are eight rules in the model:

```
IF the opponent punches high towards my left side
      THEN press '7', press SPACE-BAR.

IF the opponent punches low towards my left side
      THEN press '4', press SPACE-BAR.

IF the opponent kicks high towards my left side
      THEN press '1', press SPACE-BAR.

IF the opponent kicks low towards my left side
      THEN press '0', press SPACE-BAR.

IF the opponent punches high towards my right side
      THEN press '9', press SPACE-BAR.

IF the opponent punches low towards my right side
      THEN press '6', press SPACE-BAR.

IF the opponent kicks high towards my right side
      THEN press '3', press SPACE-BAR.

IF the opponent kicks low towards my right side
      THEN press '.', press SPACE-BAR.
```

Our models also differ across the rate at which these rules are activated. The slower variants have a delay of four seconds between attacks and the faster variants have a delay of two seconds between attacks. With these four variants, we have the conditions for our 2x2 experimental design:

	"Simple" Four Rules	**"Complex" Eight Rules**
"Slow" **0.25 attacks/second**	4 rules at a rate of 0.25 rules/sec	8 rules at a rate of 0.25 rules/sec
"Fast" **0.5 attacks/second**	4 rules at a rate of 0.5 rules/sec	8 rules at a rate of 0.5 rules/sec

3.6.Tools and Techniques

3.6.1.Production Considerations

With one developer generating all of the design, artwork and code, the primary consideration in production was time. In order to complete the game in a timely fashion, the development process had to be streamlined. A low-cost game development platform, GameMaker, was used to avoid the task of building a game engine from scratch. Simple 2D, sprite-based rendering was chosen to avoid coding logic for a 3D scene-graph and avoid coding exporters and importers for 3D geometry and animations. The sprites were generated by pre-rendering animated 3D models as 2D image sequences in order to speed up the process of creating and revising the many frames of animation. The 3D models were created and animated using 3D Studio MAX. Textures for the 3D models and additional interface graphics were created using Photoshop and The GIMP. Audio files were downloaded from the free online library, then edited using VirtualDub.

3.6.2.Producing the Game

We selected Mark Overmars' GameMaker software as the development platform for our game. This software provided a convenient platform for developing a game for the PC platform. GameMaker provides a visual interface for importing 2D images, managing game content, and building the various stages of the game. The scripting features of GameMaker gave us access to libraries for 2D graphics and file I/O.

The game logic was parameterized such that a new variant of the game could be created by simply changing a constant in the code. This facilitated randomized testing, as the game would pick one of the four variants at random at the start of a new gaming session. The post-test questionnaire was implemented in the game as well, which gave the user's testing experience a continuous flow and allowed for convenient logging of questionnaire answers to file.

3.6.3.Producing the Graphics

3D Studio MAX was used to develop the pre-rendered sprites for both players, as well as the pre-rendered background. The use of pre-rendered 3D gave the game a crisp, 3D look without the difficulty of programming a full 3D scene-graph rendering engine or programming routines for importing 3D geometry and animations. Since the game had a point-of-view fixed behind the player's avatar, and combatants that are unable to move about the ring, we were able to create 3D scenes in MAX with the boxers and the ring rendered from a fixed camera.

Pre-rendering the graphics greatly sped up the process of creating and editing the animations. Upon completion, the game had over 150 frames of animation. Using hand-drawn techniques, each of these frames would have been drawn individually by hand. Any changes to the appearance of the boxers would have necessitated redrawing many frames by hand. With pre-rendered 3D, only key poses in the animation needed to be specified, with in-between frames being interpolated automatically. Also, if the appearance of a boxer, a boxer's texture, or the ring needed to be changed, the process of re-rending all frames of animation required little additional labor.

One boxer was modeled, texture mapped, and rigged with skeletal system for both the user and the opponent in 3DS MAX. One base-texture with two variations in the color of the trunks was created for this mode in Adobe Photoshop. 21 animations were created in 3DS MAX and rendered to image sequences for conversion into sprites. Various interface graphics and iconic representations of levels of enjoyment were created in The GIMP.

To match the game design, the attack animations were designed to be two-seconds long each, moving through several “key poses”. The opponent begins each animation in a neutral pose. Over 0.5 seconds, the opponent then transitions through a “tell” pose – a pose that allows the user to clearly recognize the incoming attack. The opponent then transitions to a “strike” pose at the midway point of the animation – a pose which indicates to the user that the interval to respond to the attack is over and that the opponent has made contact with the user. In the final second of the animation has the opponent returning to the “neutral” pose.

3.6.4.Data Analysis

During user testing, data was logged to a text file marked with a unique, non-identifying session ID. The post-test questionnaire was implemented in-game, and questionnaire responses were also logged to this file. Following testing, data from each subject's text files was aggregated into a single table via a Perl script. This table of user data was then brought into Microsoft Excel for manual filtering of bad data, and preliminary analysis. Finally, the table was imported into SAS for more advanced analysis.

3.7. The User Studies

3.7.1. Subjects

Once this research was passed by the WPI Institutional Review Board (WPI IRB) as an exempt activity, approval was secured with several WPI professors to offer extra credit to students in their classes in exchange for participation in the study (see appendices B and C for materials submitted to the WPI IRB). Most subjects were undergraduates at WPI who agreed to participate in exchange for the aforementioned extra credit, with an additional number of undergraduate and graduate students who volunteered in response to an announcement of the study. These students were then given sign-up sheets where they could sign up for a specific half-hour time slot. 107 subjects were tested; however input lockup occurred during three of the testing sessions (see Section 5.3.3) and one testing session was aborted early due to program failure. The data used in our analysis is from the remaining 103 subjects. Of these 103 subjects, 82 were male and 21 were female. With respect to computer experience, 82% of subjects reported 20 or more hours of computer use per week. With respect to gaming experience, 72% of subjects reported computer game play of 10 hours or less per week, and 50% of subjects reported 5 hours or less per week.

3.7.2.Test Setup

Testing took place in a computer lab on the WPI campus. The lab was not reserved for our exclusive use, but conditions were quiet and each subject was adjacent to vacant computer stations. Subjects wore ear-covering headphones to hear in-game sounds and muffle background noise. Subjects were tested three at a time, positioned at computer stations in three corners of the lab, with subjects not facing each other. The tester was present in the room during testing, but sitting in the fourth corner of the room, not facing the subjects.

Instances of the game were executed remotely over the campus Microsoft Windows Network. The remote execution of the game centralized the recording of log files to one location on the network without the use of additional networking code within the game itself. Remote execution also enabled testing to easily take place on any free PC in the lab, as no executable files, configuration files, or log files needed to be moved back and forth from the local machine to the centralized storage.

3.7.3.Testing Procedure

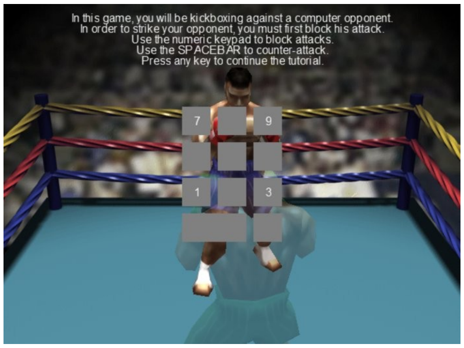

Figure 3: In-game tutorial

When each subject arrived for their time-slot, they were given a waiver that explained the testing procedure and the voluntary nature of their participation. After reading and signing the waiver an instance of the game was started on a free lab machine and the student was left alone to undergo the in-game tutorial (Figure 3), play the game, then complete the in-game questionnaire. After each round in the game, the user was asked how much they had enjoyed that round (Figure 4). After the completion of ten rounds in the game, a 20 question in-game questionnaire was administered. Once the questionnaire was completed, the subject was informed by the game that their testing session was over. The entire testing session took from 10 to 20 minutes depending on the speed of the variant and the proficiency of the subject.

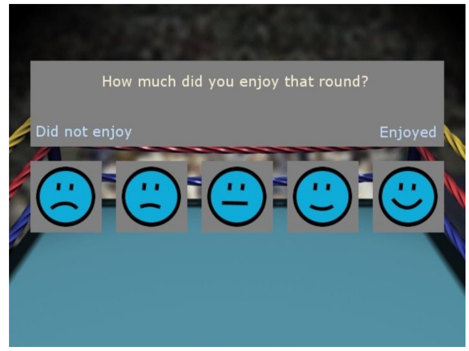

Figure 4: Post-round asessment

The questionnaire that followed the game asked them demographics questions regarding their age, gender, and computer usage and asked them to evaluate their level of interest, immersion and enjoyment of the game (see appendix for full questionnaire). The answers to these questions allowed us to relate features of the user's experience with the features of the specific game variant that they had played.

4.Results

4.1.Questionnaire

Subject questionnaire responses were scaled between -1 and 1, values representing strong disagreement and strong agreement respectively. Standard deviations for most questions were high, generally above 0.5. Figure 5 shows mean questionnaire responses to questionnaire questions 5 through 13. The height of each bar in this graph represents the mean response of one question for one of the four variants. Note that for most questions, the mean differences are not very large across the four variants. This reflects the fact that the control variables of pace and complexity were not found to have a significant effect on most responses, especially in the question related to enjoyment. It is interesting to see that pace and complexity did not significantly affect post-game enjoyment considering the results seen in section 4.4 with respect to in-game reported enjoyment. This may indicate a difference between reflective recollection of enjoyment and in-the-moment experience of enjoyment.

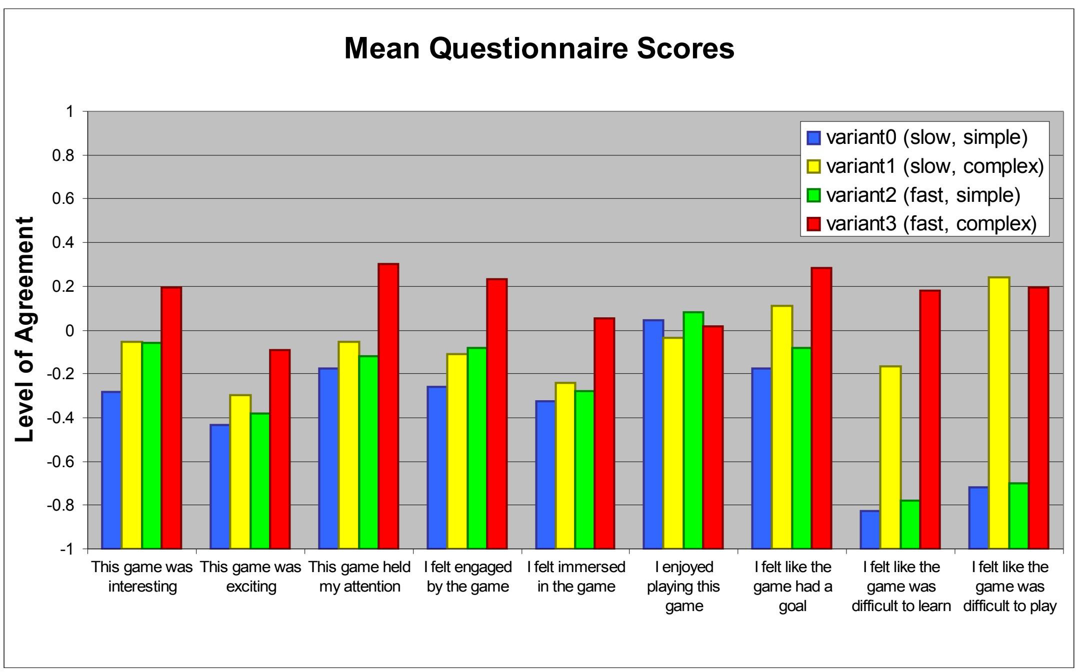

Figure 5

We conducted a two-way analysis of variance (2-way ANOVA) on each questionnaire response using pace and complexity as the independent variables.

The slow pace and low complexity levels of each variable were assigned numerical values of -1, while the fast pace and high complexity levels were assigned values of 1. The questionnaire responses were scaled on a -1 (strong disagreement) to 1 (strong agreement) scale. In almost all cases, there was not a significant difference across the two levels of the pace variable (at α 0.05). This possibly indicates a poor choice in the attack delays used to control the pace of our game (see later discussion). The pace did have a small effect on response to the question of whether the user was engaged by the game ($F=4.36, p=0.0386, R^2=0.076$).

Across levels of complexity, users showed a marked difference in responses to question number 12 “this game was difficult to learn” ($F=63.69, p<0.0001, R^2=0.41$), and question number 13 “this game was difficult to play” ($F=77.2, p<0.0001, R^2=0.44$). Complexity also had significant difference in responses to question number 7 “this game held my attention” ($F=4.31, p=0.04, R^2=0.077$) and question number 11 “this game had a goal” ($F=5.88, p=0.0171, R^2=0.066$).

In no cases did the interaction between pace and complexity explain a significant amount of the variance in questionnaire responses. This would likely be explained by the chosen levels of pace not having significant effect on their own. We also tried analyses using questionnaire responses for age, gender, computer usage and game usage as controls (question numbers 1 through 4, respectively), but did not find significant results.

4.2. User Performance

Subjects showed evidence of improving performance over the course of their ten-round play sessions. We used the user's margin of victory over the opponent as a measure of individual round performance. This value had a maximum of 10 (perfect victory) and a minimum of -10 (absolute defeat). When considering all rounds played by all users, the complexity of the variant showed a significant effect on the margin of victory ($F=661$, $p<0.0001$, $R^2=0.392$), and mean margin of victory increased based on round number ($F=3.17$, $p=0.0009$, $R^2=0.0272$). Figure 6 illustrates the difference in average performance between the simple and complex variants over the course of 10 rounds. In this figure, the horizontal axis represents the round of the game and the vertical axis represents the point difference, or margin-of-victory, at the end of the round. There are four data series which represent the four game variants, and each data point is the mean margin-of-victory in that round for subjects who played that variant.

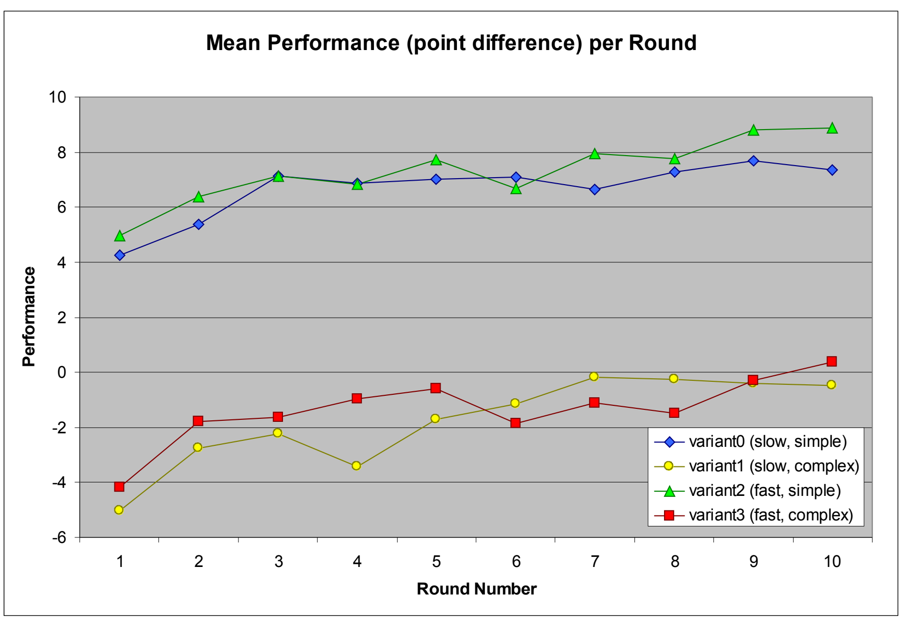

Figure 6

This graph indicates that there was a large performance gap between players of the simple and complex variants. On average, players of all variants improved in their margin of victory by roughly 4 points over the course of testing. Notice that although improvement in performance was similar, average performance of players in the simple variants begins and ends in "winning territory", while average performance in the complex variants begins in "losing territory" and approaches the win-loss threshold.

4.3. Enjoyment Over Time

Figure 7 shows a graph of in-game enjoyment over the course of the ten rounds

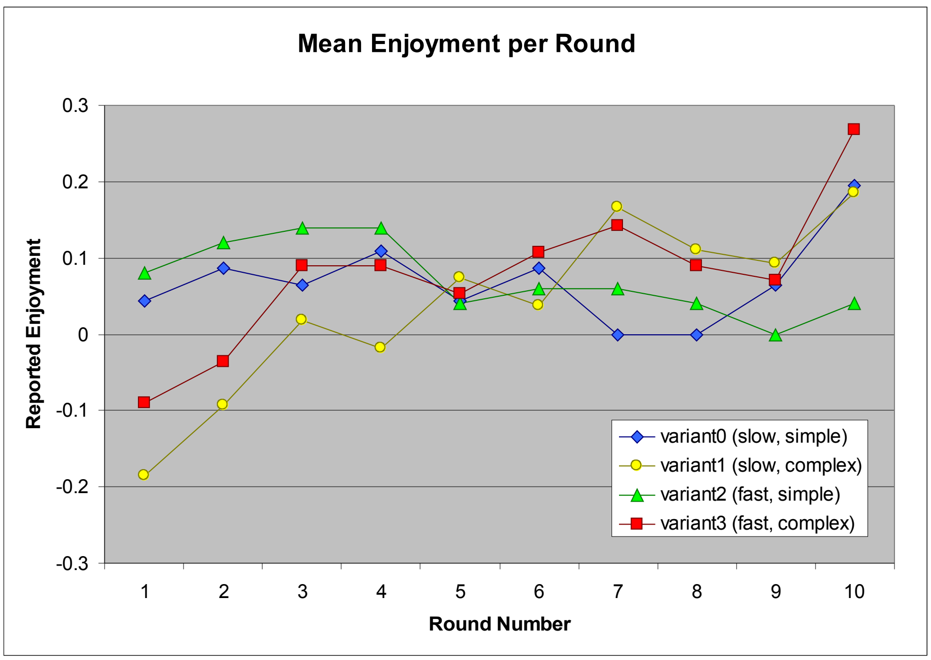

Figure 7

for each variant. The horizontal axis represents the round number, and the vertical axis represents the mean reported enjoyment in the [-1,1] range. In this graph, enjoyment appears flat with the simple game variants, but enjoyment appears to increase over time with the complex variants.

One could assume that the more complex variants were more difficult and therefore as the user's performance improved, they felt a greater sense of accomplishment in overcoming the challenge. However, as observed in section 4.2, the absolute improvement in mean margin-of-victory from round 1 to round 10 was similar for all variants. If the users were simply judging their increasing mastery of the game, enjoyment over time should be similar for all variants. There is another possible explanation of the difference in enjoyment over time, which we will discuss in section 4.4.

4.4.Enjoyment of Success

With our chosen factors of pace and complexity not proving to have much effect on user enjoyment, we decided to look at reported enjoyment with respect to user performance. First we compared the post-game enjoyment reported in the questionnaire to measures of overall performance. For measures of overall performance, we considered the fraction of rounds that the user had won and the user's average margin of victory over all rounds. We performed a regression on enjoyment with respect to pace, complexity, fraction-won and average-victory. In this model, fraction-won showed a significant effect ($p=0.044$, parameter estimate 1.33) on the user's post-game enjoyment.

Next we looked at reported enjoyment on the individual-round level. We assigned the variable "won" the value of 1 for a round that the user won and 0 for a round that the user lost. We performed a regression on the reported enjoyment for each round across all users with respect to pace, complexity and “won”. In this model, “won” showed a significant effect (p<0.0001, parameter estimate 0.6) and complexity also showed a significant effect (p<0.0001), parameter estimate 0.14) on the user’s enjoyment of an individual round.

Finally, with success seeming to be a significant factor in user enjoyment, we performed a regression on individual round enjoyment with respect to the user’s margin of victory in that round. In this model, margin of victory showed a significant effect on round enjoyment (p<0.0001). Visually, this relationship can be seen by looking at a graph of the mean reported enjoyment for each margin of victory (Figure 8). The horizontal axis of this figure represents the margin of victory; the vertical axis represents reported enjoyment. Each data point in this graph represents the mean reported of enjoyment for all rounds that ended in the given margin of victory, with the vertical lines depicting the standard deviation for that statistic. The histogram above the graph represents the number of rounds that ended in the respective margin-of-victory (these numbers fall in the range [26,147]). The most interesting feature of this graph is the large jump in reported enjoyment between a margin of -1 (barely lost) and 1 (barely won). Also interesting is the falloff of enjoyment close to the rounds that were perfect victories. Here we see empirical evidence of an “inverted-U” phenomenon. Unlike the Yerkes-Dodson function, which plots learning versus arousal, this inverted-U appears in a graph of performance versus enjoyment. This graph is similar to the Goldilocks Functions described by Burns that relate fun to win-probability [BU2006].

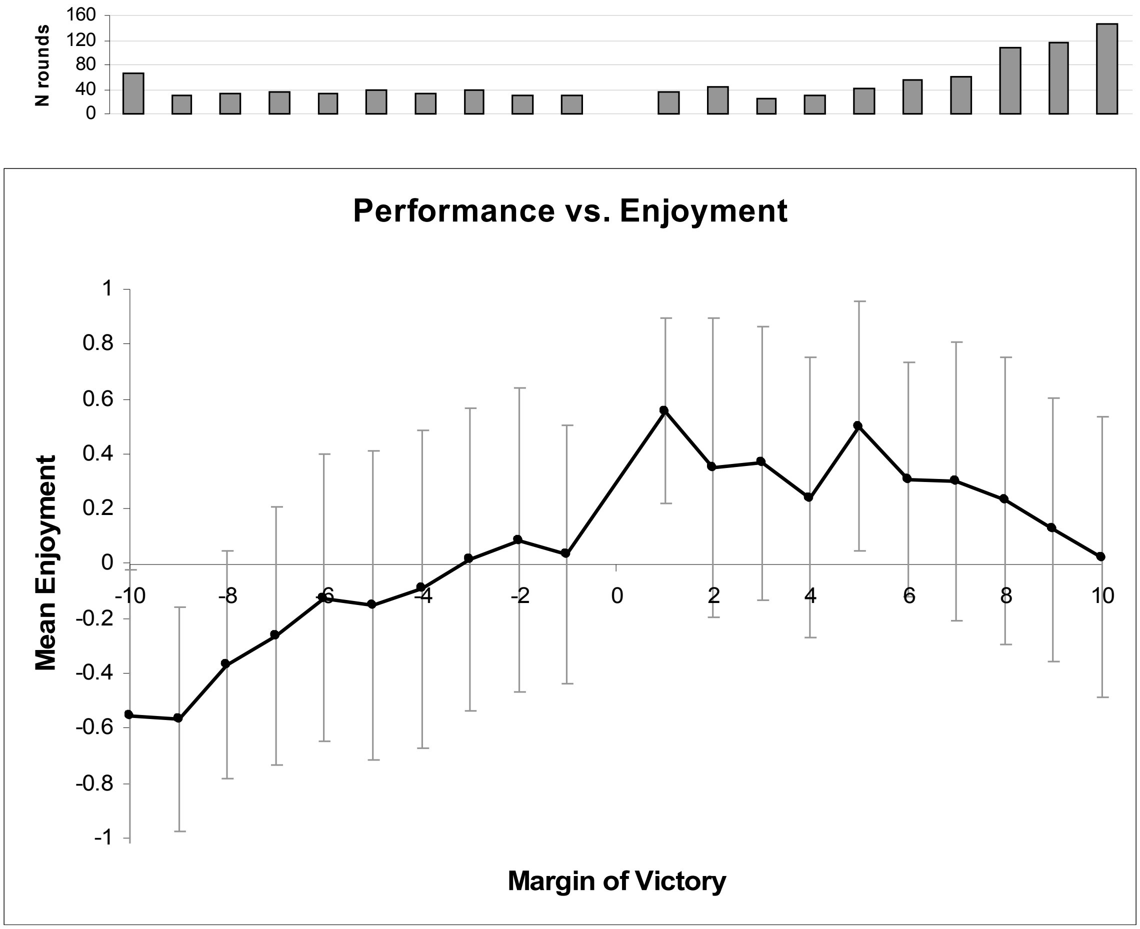

Figure 8

This exploration of mean enjoyment also elucidates the differences in enjoyment over time between the simple and complex game variants. Referring back to mean performance per round, one can see that over the course of testing, the average play session involving a simple variant moves from moderate success to near perfect success, whereas the average play session involving a complex variant moves from moderate failure to the barely-lost/barely-won threshold. Thus the players of the simple variants are moving down the right decline of the inverted-U as their play improves, while the players of the complex variants are moving up the left incline of the inverted-U as they improve. Both groups' performance may be improving to a similar degree, but what appears to be important is not how much performance improves, but how close the user's experience was to the "sweet spot" on the inverted-U of performance.

5.Discussion

5.1.Development Post-mortem

5.1.1.Consequences of Sprite-Based design

Although using pre-rendered 3D art saved us a great deal of effort, the consequence was that the game had a very large memory footprint. Because we were using a middleware development platform, we did not have the ability to optimize the memory usage of the sprite engine. It is likely that images for all frames of animation were held uncompressed in memory simultaneously, with the memory usage of the game becoming excessive. In order to save memory, we halved the frame-rate of our animations, thus halving the number of images per animation. Even after this change, memory usage often exceeded 250 MB. We were unable to conduct user studies on machines with slow hard-disks or insufficient RAM because disk caching effects would cause a good deal of slowdown, slowing the pace of the game and lengthening the response interval for reacting to a punch. Fortunately, we had access to a lab with machines that was able to play the game smoothly.

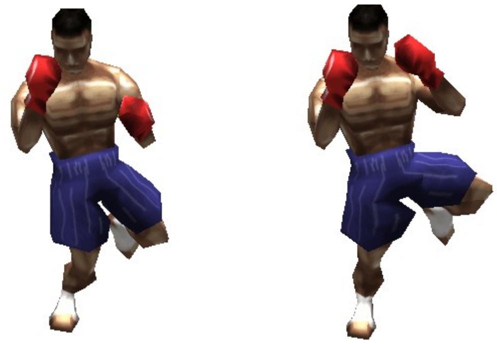

Figure 9: Differences in "tell" poses between left-low kick and left-high kick.

5.1.2.Clarity of Poses

After testing, a few subjects reported difficulty in distinguishing between the high- and low-kick conditions until it was too late. We had hoped to make each animation such that as the opponent transitions into the "tell" poses the task of recognizing each condition would be easy. In the case of the kick animations, the "tell" poses were similar, and early frames of the animations differed only slightly (Figure 9). If the subject was unable to detect this subtle difference, then the condition would not be recognized until a later point in the animation, and the user would have effectively less time to react with a block action. It is possible that this problem was caused in part by the frame-rate reduction of the animations; a step which was taken to save memory. To correct this, the kick animations should be reworked such that the difference is clear at an earlier time.

5.1.3.Problems with Pace

Our original intent was to have the enemy attack at a rate of one attack every two or four seconds, depending on the variant. During informal testing, we realized that how quickly a user countered the opponent affected how fast they perceived the pace of the game to be. For example, assuming a two-second attack interval, if the user blocks and counters in 0.5 seconds, the opponent's next attack would occur 1.5 seconds after the counter attack. If the user were less proficient in recognizing the condition and responding, their block and counter could take as long as 1.5 seconds, leaving only 0.5 seconds before the opponent's next attack and making the pace of the game seem much faster.

We chose to change the pace from a rate of attack to a delay until the next attack. This way, no matter how long the user took to execute their actions, the duration until the opponent's next attack would be constant. In doing this, we made the mistake of keeping two-seconds and four-seconds as time values. Following the user study, it became clear that the chosen levels for the pace factor resulted in little difference between questionnaire responses – especially in reported level of difficulty, where the complexity factor showed a large effect.

We believe the poor values of the pace variable resulted form not adjusting for the change made to the implementation of the game. Prior to the change, the effective delay before the opponent's next attack following the completion of the user's response would have been the attack interval minus the time for the user to recognize and respond to each attack. This delay would have been close to one-second for the "fast" variants, and close to three seconds for the "slow" variants. In the change from implementing pace as a rate to pace as a delay we should have adjusted the levels of the pace variable to accommodate for this.

5.2.Modeling

An underlying assumption of this research was that the user would recognize a certain condition, make a decision on how to respond and then respond within a certain amount of time. Within this interaction, we only considered the cognitive task of decision making; however this process also involves the perceptual task of recognizing the conditions and the motor task of executing the responses. It became clear when certain users had difficulty differentiating between a few of the attacks, and when users overshot or missed the appropriate key, that the perceptual and motor tasks are also important factors in at least the difficulty of the game, if not the entertainment value.

In order to complete game development and cognitive modeling in a short amount of time, and ensure that the game was easily learned, a very simple game design was chosen. The decision making process with this simple game was degenerate because there were no strategic decisions to be made, only reactions. It would not be practical to model user cognition for more complicated games, or that a game would have flexible game play and lack a normative style of play.

5.3.Testing

5.3.1.Sample Imbalance

For most tests, the software picked one of the four game variants at random for the subject to play. Although the game's random selection process had an equal probability of selecting each variant, as the study neared its end it became apparent that the fast pace, low complexity test group was under-populated. A fixed-variant version of the game was created, and the final nine subjects were all tested using the fast pace, low complexity variant. The final sample sizes were 23 subjects for the slow and simple variant, 27 for the fast and simple variant, 25 for the slow and complex variant, and 28 for the fast and complex variant.

5.3.2.Data Loss over Network

For convenient aggregation of logging, it was decided to run the game executables remotely over a Microsoft Windows network from a central location. This convenience ultimately cost us a good deal of data. The logging functionality of the game had only been tested in a single-machine environment, and when the game was run across the network the last two of the twenty questionnaire questions were lost.

5.3.3.NUMLOCK

In a few cases, a problem arose due to the location of input keys on the keyboard. The buttons for blocking were located on the numeric keypad of the PC keyboard. If the user tried to press the ‘7’ key, but overshot their target by a full row of keys, they would hit the NUMLOCK key, which would cause the game to stop recognizing key presses on the numeric keypad. The data for the few cases where this issue interrupted testing was discarded.

6.Conclusions

We originally set out to show that the complexity of a game impacts its entertainment value. However, our measures of pace and complexity did not show a direct relationship to post-game evaluations of enjoyment. The complexity variable did show a significant effect on subject performance, perception of difficulty, and reported in-game enjoyment. This result supports our hypothesis that complexity affects enjoyment, but also indicates an important difference between in-game reporting of enjoyment and post-game reporting of enjoyment.

Subject performance improved similarly across all variants; however in-game enjoyment did not follow improving performance. Post-game enjoyment was significantly related to the fraction of rounds won and in-game enjoyment was significantly related to whether an individual round was won. An analysis of in-game enjoyment with respect to margin-of-victory showed evidence of an inverted-U phenomenon. This differed from previous research in that performance was treated as the independent variable and enjoyment as the dependant variable. The difference in enjoyment over time across game complexity can be explained by observing that players of the simple variants were approaching perfectly victorious performance, while players of complex variants were approaching barely victorious performance. The experimental evidence showed that mean enjoyment peaked at levels of performance near barely victorious, and fell off towards perfectly victorious, thus players of complex variants were climbing up the enjoyment curve while players of simple variants were sliding down the side. This result is significant in that proximity to the victory-threshold is an important factor in the enjoyment of a game such as ours. This result also contradicts the idea that mastery of a game leads to enjoyment, because it indicates that an increase in performance can lead to a decrease in enjoyment.

7.Future Work

We believe that motivational factors will help explain user enjoyment of an interactive game. With success seeming to play heavily into user enjoyment, further research could investigate how goals play into enjoyment. What are the differences between enjoyment of a game with one clear victory condition and a game with multiple goals? Are goals necessary for enjoyment, or will a user create goals to enjoy overcoming? How do things such as frequency of goal-satisfaction, and perceived probability of goal-satisfaction play into user enjoyment? Social goals such as competition and cooperation with other people could also be important factors in user enjoyment.

Although the differences in our cognitive models did not explain user enjoyment very well, we feel that they proved to at least be good ways of characterizing a task and speaking of its difficulty. Cognitive models provide us with tools to speak of the equivalence of games. Once we can speak of structural similarities in the game-play of different games, we can research factors outside of game-play that influence enjoyment. Would two games that have identical cognitive models be enjoyed differently by users if they are framed in different contexts (e.g. a competitive activity versus a cooperative activity) or had a different style of artwork (cartoon 2D sprites versus realistic 3D models)? If we can isolate game-play, then we can ask whether there are demographic differences in enjoyment, and if so determine to what degree things such as game-play, context and aesthetics contribute to these differences.

References

[BE2001] Belkavin, R.V., “Modelling the Inverted-U Effect in ACT-R”, *Proceedings of the 2001 Fourth International Conference on Cognitive Modelling*, 275-276.

[BU2006] Burns, K., (in press) “Fun in Slots”, *Proceedings of CIG-06: IEEE Conference on Computational Intelligence in Games*. Reno, NV, May 22-24, 2006.

[BY2001] Byrne, Michael D., “ACT-R/PM and menu selection: Applying a cognitive architecture to HCI.” *International Journal of Human-Computer Studies.* 55. 2001, 41-84.

[CF1998] Chery, Sandra, Farrell, Philip S.E., “A Look at Behaviourism and Perceptual Control Theory in Interface Design”, *Defence and Civil Institute of Environmental Medicine*, Report Number 98-R-12. 1998.

[DH2000] Douglas, Yellowlees; Hargadon, Andrew. “The Pleasure Principle: Immersion Engagement, Flow”, *Proceedings of the Hypertext 2000 Conference*, 2000, 153-160.

[E1] Eppstein, D., “Computational Complexity of Games and Puzzles”, URI: http://www.ics.uci.edu/~eppstein/cgt/hard.html

[GA2001] Gray, Wayne D., Altmann, Erik M., “Cognitive Modeling and Human-Computer Interaction”, *International Encyclopedia of Ergonomics and Human Factors*. Oct 2001, 387-391.

[HM1997] Harp, Shannon F.; Mayer, Richard E. “The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest”, *Journal of Educational Psychology*. 89(1), Mar 1997, 92-102.

[H1952] Hick, W. E. “On the Rate of Gain of Information”, *Quarterly Journal of Experimental Psychology* 4. 1952, 11-26.

[JK1996] John, Bonnie E.; Kieras, David E. “Using GOMS for User Interface Design and Evaluation: Which Technique?” *ACM Transactions on HCI* 3(4). Dec 1996, 287-319.

[JP1992] John, Bonnie E.; Peck, Virginia A. “BROWSER-SOAR: A Computational Model of a Highly Interactive Task”, *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1992, 165-172.

[SR2001] St. Amant, Robert; Riedl, Mark O. “A perception /action substrate for cognitive modeling in HCI”, *International Journal of Human-Computer Studies* 55. 2001, 15-39.

[ST2004] Steiner, Karl; Tomkins, Jay. "Narrative Event Adaptation in Virtual Environments", *Proceedings of the 9th International Conference on Intelligent User Interfaces*. 2004, 46-53.

[VHK2003] Vorderer, Peter; Hartmann, Tilo; Klimmt, Christoph. “Explaining the Enjoyment of Playing Videogames: The Role of Competition”, *Proceedings of the Second International Conference on Entertainment Computing*. 2003, 1-9.

[VKR2004] Vorderer, Peter; Klimmt, Cristoph; Ritterfeld, Ute. “Enjoyment: At the Heart of Media Entertainment”, *Communication Theory* 14(4). 2004, 388-408.

[YD1908] Yerkes, Robert M.; Dodson, John D. “The Relation of Strength of Stimulus to Rapidity of Habit-Formation”, *Journal of Comparative Neurology and Psychology* 18. 1908, 459-482.

Appendix A: Post-Game Questionnaire

The following questions are for demographic purposes only, and will not be used to identify you individually

What is your age, in years?	**Under 17**	**17-19**	**20-22**	**23-25**	**Over 25**
What is your Gender?		Male		Female	
About how many hours per week do you use a computer?	0-10	10-20	20-30	30-40	40+
About how many hours per week do you play computer games?	0-5	5-10	10-15	15-20	20+

How much do you agree with the following statements:

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
This game was interesting.	❍	❍	❍	❍	❍
This game was exciting.	❍	❍	❍	❍	❍
This game held my attention.	❍	❍	❍	❍	❍
I felt engaged by the game.	❍	❍	❍	❍	❍
I felt immersed in the game.	❍	❍	❍	❍	❍
I enjoyed playing this game.	❍	❍	❍	❍	❍
I felt like the game had a goal.	❍	❍	❍	❍	❍
I felt like the game was difficult to learn.	❍	❍	❍	❍	❍
I felt like the game was difficult to play.	❍	❍	❍	❍	❍
I identified with the character that represented me in the game.	❍	❍	❍	❍	❍
I liked the character that represented me in the game.	❍	❍	❍	❍	❍
I like to play computer games.	❍	❍	❍	❍	❍
I like to play action-oriented computer games.	❍	❍	❍	❍	❍
I like to play fighting-oriented computer games.	❍	❍	❍	❍	❍
I like to play boxing games such as Punch-Out.	❍	❍	❍	❍	❍

Appendix B: Consent Form

Consent to Participate in Research

Relating Cognitive Models of Computer Games to User Evaluations of Entertainment

Purpose: This study involves research into the entertainment value of computer games.

Procedure: If you volunteer to participate in this study, you will be asked to play a simple computer game for approximately ten minutes, and then answer a series of questions about yourself and your game play experience. The entire testing session should take from twenty to thirty minutes.

Potential Risks: There will be no more risk or discomfort in participation than in a typical ten minute session of computer game playing.

Potential Benefits: You will not directly benefit from participating in this research outside of practice playing video games.

Academic Credit: If you are participating in this research for academic credit, please notify the test monitor of your course and instructor.

Confidentiality: Any information collected during this research that can be identified with you will remain confidential. Although you will be asked for some demographic information, the testing session will collect no information that could identify you personally. Test data will be stored in locations only accessible to the principal investigator and the system administrators of WPI's Computing & Communications Center.

Voluntary Participation: Your participation in this experiment is voluntary, and you are free to refuse to participate or discontinue participation at any time without loss of benefits to which you have been entitled.

Investigators: If you have any questions or concerns regarding this research, please contact the Principal Investigator, Paolo M. Piselli, Computer Science Department, Worcester Polytechnic Institute, 100 Institute Rd., Worcester, MA 01609-2280, 508-207-7373, ppiselli@wpi.edu; Faculty Advisor Mark Claypool, claypool@cs.wpi.edu; or Faculty Advisor James Doyle, doyle@wpi.edu.

Signature of Subject: I understand the test procedures, and my questions regarding the present research have been answered to my satisfaction. I have been given a copy of this form.

____

Name of Subject

____ ______________________

Signature of Subject **Date**

Appendix C: Protocol Document

WPI Human Research Protocol Documentation

with regards to "Relating Cognitive Models of Computer Games to User Evaluations of Entertainment", Paolo M. Piselli, Principal Investigator

<u>Description of Research</u>

The aim of this research is to investigate the relationship between the cognitive complexity of a computer game and its entertainment value. The goal of this research is to prove the viability of using cognitive models as tools for evaluating computer games. This research will be carried out by comparing normative models for playing variations of simple kickboxing games to evaluations of entertainment given by human subjects who play these games.

<u>Participant Population</u>

Participants will be college-aged men and women recruited from several courses related to games and cognition at Worcester Polytechnic Institute. Participants will receive a small amount of bonus academic credit in exchange for their voluntary participation. All volunteers will be included in the experimental population.

<u>Testing Procedures</u>

Human testing will involve each subject playing a video game for approximately ten-minutes. Periodically during the play session, the subject will be polled to rank their level of enjoyment of the game. Following the play session, the subject will be asked to answer a few demographic questions, a few questions regarding their experience playing computer games, and a series of questions relating to their enjoyment of the game. The entire testing session for a subject should last between twenty and thirty minutes.

<u>Risks and Benefits to Participants</u>

Participation in this research will expose subjects to no more risk than a typical ten minute session of playing a video game. Although some games that expose the user to flashing or strobing imagery may pose an epilepsy risk, the games used for this research do not use any such intense visual stimulation. Subjects will receive little benefit from the research aside from practice at playing games similar to those used in testing.

<u>Informed Consent</u>

Informed consent will be secured via an informed consent form. (see attached form)

<u>Questionnaire</u>

The attached questionnaire lists the questions that subjects will be asked to answer regarding themselves and their game play experiences. These questions may ultimately be presented to the subjects in digital form.