

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/226120628>

# Capturing Player Enjoyment in Computer Games

Chapter · June 2007

DOI: 10.1007/978-3-540-72705-7\_8

---

CITATIONS

13

---

READS

131

2 authors, including:



Georgios Yannakakis  
University of Malta

220 PUBLICATIONS 5,360 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



COMnPLAY-Science [View project](#)



C2Learn [View project](#)

# 1. Capturing Player Enjoyment in Computer Games

Georgios N. Yannakakis and John Hallam

Maersk Mc-Kinney Moller Institute  
University of Southern Denmark  
Campusvej 55, Odense M, DK-5230  
{georgios,john}@mip.sdu.dk

The current state-of-the-art in intelligent game design using Artificial Intelligence (AI) techniques is mainly focused on generating human-like and intelligent characters. Even though complex opponent behaviors emerge through various machine learning techniques, there is generally no further analysis of whether these behaviors contribute to the satisfaction of the player. The implicit hypothesis motivating this research is that intelligent opponent behaviors enable the player to gain more satisfaction from the game. This hypothesis may well be true; however, since no notion of entertainment or enjoyment is explicitly defined, there is therefore no evidence that a specific opponent behavior generates enjoyable games.

This chapter introduces a discussion of quantitative entertainment capture in real-time and presents two dissimilar approaches for modeling player satisfaction. Successfully capturing the level of entertainment during play provides insights for designing the appropriate AI methodology for entertainment augmentation in real-time. For this purpose, adaptive on-line learning methodologies are proposed and their strengths and limitations are discussed.

## 1.1 Introduction

Cognitive modeling within human-computer interactive systems is a prominent area of research. Computer games, as examples of such systems, provide an ideal environment for research in artificial intelligence (AI), because they are based on simulations of highly complex and dynamic multi-agent worlds [1.1, 1.2, 1.3]. Moreover, computer games offer a promising ground for cognitive modeling since they embed rich forms of interactivity between humans and non-player characters (NPCs) [1.4]. Being able to capture quantitatively the level of user (gamer) engagement or satisfaction in real-time can grant insights to the appropriate AI methodology for enhancing the quality of playing experience [1.5] and furthermore be used to adjust digital entertainment environments according to individual user preferences.

Motivated by the lack of quantitative models of entertainment, endeavors to measure and augment player satisfaction in real-time are presented in this chapter. More specifically, the open question of modeling entertainment during game play is discussed and the strengths and weaknesses of

previous attempts in the field are outlined. Herein entertainment is defined qualitatively primarily as the level of satisfaction generated by the real-time player-game opponent interaction — by ‘opponent’ we define any controllable interactive feature of the game. We view a game primarily as a learning process, and the level of entertainment is kept high when game opponents enable new learning patterns (‘not too easy a game’) for the player that can be perceived and learned by the player (‘not too difficult a game’) [1.6, 1.7]. On the same basis, according to [1.8] — within the axis of emotions varying from boredom to fascination — learning is highly correlated with interest, curiosity and intrigue perceived. The collection of these emotions is referred to as entertainment (or “fun”) in this chapter.

Two different approaches for quantitatively capturing and enhancing the real-time entertainment value of computer games are presented in this chapter: one based on empirical design of entertainment metrics and one where quantitative entertainment estimates are extracted using machine learning techniques, grounded in psychological studies. The predator/prey game genre is used for the experiments presented here; though it is argued that the proposed techniques can be applied more generally, to other game genres.

In the first, “empirical,” approach, a quantitative metric of the ‘interestingness’ of opponent behaviors is designed based on qualitative considerations of what is enjoyable in predator/prey games. A mathematical formulation of those considerations, based upon data observable during game play, is derived. This metric is validated successfully against the human notion of entertainment.

In the second approach, entertainment modeling is pursued by following the theoretical principles of Malone’s [1.9] intrinsic qualitative factors for engaging game play, namely *challenge* (i.e. ‘provide a goal whose attainment is uncertain’), *curiosity* (i.e. ‘what will happen next in the game?’) and *fantasy* (i.e. ‘show or evoke images of physical objects or social situations not actually present’) and driven by the basic concepts of the Theory of Flow [1.10] (‘flow is the mental state in which players are so involved in the game that nothing else matters’). Quantitative measures for challenge and curiosity are inspired by the “empirical” approach to entertainment metrics. They are represented by measures computed from appropriate game features based on the interaction of player and opponent behavior. A mapping between the quantitative values of these challenge and curiosity measures and the human notion of entertainment is then constructed using evolving neural networks (NNs).

The chapter concludes with a discussion of several remaining open questions regarding entertainment modeling and proposes future directions to answer these questions. The limitations of the presented methodology, and the extensibility of the proposed approaches of entertainment capture and augmentation to other genres of digital entertainment, are also discussed.

## 1.2 Capturing Entertainment

There have been several psychological studies to identify what is “fun” in a game and what engages people playing computer games. Theoretical approaches include Malone’s principles of intrinsic qualitative factors for engaging game play [1.9], namely challenge, curiosity and fantasy as well as the well-known concepts of the theory of flow [1.10] incorporated in computer games as a model for evaluating player enjoyment, namely *GameFlow* [1.11].

A comprehensive review of the literature on qualitative approaches for modeling player enjoyment demonstrates a tendency to overlap with Malone’s and Csikszentmihalyi’s foundational concepts. Many of these approaches are based on Lazzaro’s “fun” clustering which uses four entertainment factors based on facial expressions and data obtained from game surveys of players [1.12]. According to Lazzaro, the four components of entertainment are: hard fun (related to the challenge factor of Malone), easy fun (related to the curiosity factor of Malone), altered states (i.e. ‘the way in which perception, behavior, and thought combine in a collective context to produce emotions and other internal sensations’ — closely related to Malone’s fantasy factor) — and socialization (the people factor). Koster’s [1.7] theory of fun, which is primarily inspired by Lazzaro’s four factors, defines “fun” as the act of mastering the game mentally. An alternative approach to fun measure is presented in [1.13] where fun is composed of three dimensions: endurability, engagement and expectations. Questionnaire tools and methodologies are proposed in order to empirically capture the level of fun for evaluating the usability of novel interfaces with children.

Kline and Arlidge [1.14] support Lazzaro’s findings since their studies on on-line massive multi-player games (e.g. Counter-Strike) identify the socialization factor as an additional component to Malone’s factors for “fun” game play experiences. Their clustering of player styles (‘player archetypes’) corresponds to these four dimensions of entertainment: *Warriors* are those who prioritize combat features and realism (closely related to Malone’s fantasy factor), *Narrators* are those who place priority on the plot, characters, and exploration but they do not like games that are challenging (closely related to Malone’s curiosity factor), *Strategists* are those that prioritize complex strategies, challenging game play and mastery (closely related to Malone’s challenge factor) and *Interactors* for whom competition and cooperation with other players is of the highest importance (socialization factor).

On that basis, some endeavors towards the criteria that collectively make simple on-line games appealing are discussed in [1.15]. The human survey-based outcome of that work presents challenge, diversity and unpredictability as primary criteria for enjoyable opponent behaviors.

Vorderer et al. [1.4] present a quantitative analysis (through an extensive human player survey on the Tomb Raider game) of the impact of competition (i.e. challenge) on entertainment and identify challenge as the most important determinant of the enjoyment perceived by video game players.

They claim that a successful completion of a task generates **sympathetic arousal, especially** when the challenge of the task matches the player’s abilities. In their analysis, social competition (just like Lazzaro’s people factor and Kline’s and Arlidge’s *Interactors*) appears to enhance entertainment. Their survey, however, is based on single game-task enjoyment evaluations rather than comparative evaluations of several scenarios.

The study by Choi et al. [1.16] ranks perceptual and cognitive fun as the top-level factors for designing “**fun**” computer games. For the former, it is the game interface that affects the player’s perception (vividness and imagination) — this corresponds to Malone’s fantasy factor. For the latter, it is the game mechanics (level of interactivity) that affect the player’s cognitive process — which comprises the challenge and satisfaction factors. According to Choi et al. (*ibid.*), challenge and satisfaction appear as independent processes, in contrast to the views of Malone [1.9] and Yannakakis et al. [1.17] where satisfaction derives from the appropriate level of challenge and other game components. Moreover, in Choi et al.’s study, vividness and imagination (perceptual fun) appear more important entertainment factors for players of strategic simulation games and challenge and satisfaction (cognitive fun) appear more important for role-playing games.

As previously mentioned, research in the field of game AI is mainly focused on generating **human-like (believable) and intelligent (see [1.3, 1.18] among others) characters**. Complex NPC behaviors can emerge through various AI techniques; however, there is no further analysis of whether these behaviors have a positive impact to the satisfaction of the player during play. According to Taatgen et al. [1.19], believability of computer game opponents, which are generated through cognitive models, is strongly correlated with enjoyable games. Such implicit research hypotheses may well be true; however, there is little evidence that specific NPCs generate enjoyable games unless a notion of interest or enjoyment is explicitly defined.

Iida’s work on metrics of entertainment in board games was the first attempt in the area of quantitative “fun” modeling. He introduced a general metric of entertainment for variants of chess games depending on average game length and possible moves [1.20]. Other work in the field of quantitative entertainment capture is based on the hypothesis that the player-opponent interaction — rather than the audiovisual features, the context or the genre of the game — is the property that contributes the majority of the quality features of entertainment in a computer game [1.6]. Based on this fundamental assumption, a metric for measuring the real time entertainment value of predator/prey games was designed, and established as efficient and reliable by validation against human judgement [1.21, 1.22]. Further studies by Yannakakis and Hallam [1.23] have shown that Artificial Neural Networks (ANN) and fuzzy neural networks can extract a better estimator of player satisfaction than a custom-designed (or designer-driven) one, given appropriate estimators of the challenge and curiosity of the game and data on human players’

preferences. Similar work in adjusting a game’s difficulty include endeavors through reinforcement learning [1.24], genetic algorithms [1.25], probabilistic models [1.26] and dynamic scripting [1.27]. However, the aforementioned attempts are based on the assumption that challenge is the only factor that contributes to enjoyable gaming experiences while results reported have not been cross-verified by human players.

A step further to entertainment capture is towards games of richer human-computer interaction and affect recognizers which are able to identify correlations between physiological signals and the human notion of entertainment. Experiments by Yannakakis et al. [1.28] have already shown a significant correlation of average heart rate with children’s perceived entertainment in action games played in interactive physical playgrounds. Moreover, Rani et al. [1.29] propose a methodology for detecting anxiety level of the player and appropriately adjusting the level of challenge in the game of ‘Pong’ based on recorded physiological signals in real-time and subject’s self-reports of their emotional experiences during game play. Physiological state (heart-rate, galvanic skin response) prediction models have also been proposed for potential entertainment augmentation in computer games [1.30].

Following the theoretical principles reported from Malone [1.9], Koster [1.7] and Yannakakis [1.21], and to a lesser degree from Lazzaro [1.12] and Kline and Arlidge [1.14], this chapter is primarily focused on the contributions of game opponents’ behavior (by enabling appropriate learning patterns on which the player may train [1.7]) to the real-time entertainment value of the game. This chapter therefore excludes the socialization factor of entertaining game play and investigates instead entertainment perceived in single player scenarios. We argue that among the three dimensions of “fun” (endurability, engagement, expectations) defined in [1.13] it is only *engagement* that is affected by the opponent since both *endurability* and *expectations* are based primarily on the game design *per se*. Given a successful interactive game design that yields high expectations and endurability, we only focus on the level of engagement that generates fun (entertainment).

Rather than being based purely on theoretical assumptions and visual observations of players’ satisfaction, this chapter presents two different approaches that attempt to capture quantitatively the level of player entertainment in computer games. First, a custom-designed quantitative metric of entertainment is proposed, motivated by qualitative considerations of what is enjoyable in computer games. The metric has been validated against humans’ preferences. Second, a mapping between estimators of Malone’s challenge and curiosity entertainment factors and humans’ valuations of entertainment is derived using evolving NNs.

### 1.3 The Test-bed Game

The test-bed studied here is a modified version of the original Pac-Man computer game released by Namco. The player's (*PacMan's*) goal is to eat all the pellets appearing in a maze-shaped stage while avoiding being killed by the four *Ghosts*. The game is over when either all pellets in the stage have been eaten by *PacMan*, *Ghosts* manage to kill *PacMan* or a predetermined number of simulation steps is reached without either of the above occurring. In the last case, the game restarts from the same initial positions for all five characters. In the test-bed game, *PacMan* is controlled by the human player while a multi-layered feedforward neural controller is employed to manage the *Ghosts'* motion.

The game is investigated from the opponents' viewpoint and more specifically how the *Ghosts'* adaptive behaviors and the levels of challenge and curiosity they generate can collectively contribute to player satisfaction. The game field (i.e. stage) consists of corridors and walls. Both dimensions and maze structure of the stage are predefined. For the experiments presented in this chapter we use a  $19 \times 29$  grid maze-stage with corridors 1 grid-cell wide (see [1.31] for more details of the Pac-Man game design).

We choose predator/prey games as the initial genre for this research since, given our aims, they provide us with unique properties: in such games we can deliberately abstract the environment and concentrate on the characters' behavior. Moreover, we are able to easily control a learning process through on-line interaction. Other genres of game (e.g. first person shooters) offer similar properties; however predator/prey games are chosen for their simplicity as far as their development and design are concerned.

### 1.4 Empirical Estimation of Entertainment

As noted in section 1.3, predator/prey games will be our test-bed genre for the investigation of enjoyable games. More specifically, in the games studied, the prey is controlled by the player and the predators are the computer-controlled opponents (non-player characters, or NPCs).

In the approach presented in this section, a quantitative metric of player satisfaction is designed based on general criteria of enjoyment. The first step towards generating enjoyable computer games is therefore to identify the criteria or features of games that collectively produce enjoyment (or else interest) in such games. Second, quantitative estimators for these criteria are defined and combined, in a suitable mathematical formula, to give a single quantity correlated with player satisfaction (interest). Finally, this formulation of player interest is tested against human players' judgement in real conditions using the Pac-Man test-bed (see section 1.4.2).

Following the principles of Yannakakis and Hallam [1.6, 1.22] we will ignore mechanics, audiovisual representation, control and interface contribu-

tions to the enjoyment of the player and we will concentrate on the opponents' behaviors. A well-designed and popular game such as Pac-Man can fulfil all aspects of player satisfaction incorporated in the above-mentioned design game features. The player, however, may contribute to his/her own entertainment through interaction with the opponents of the game and therefore is included implicitly in the interest formulation presented here — see also [1.32] for studies of the player's impact on his/her entertainment.

**Criteria.** By observing the opponents' behavior in various predator/prey games we attempted to identify the key features that generate entertainment for the player. These features were experimentally validated against various opponents with different strategies and redefined when appropriate. Hence, by being as objective and generic as possible, we believe that the criteria that collectively define interest on any predator/prey game are as follows.

1. *When the game is neither too hard nor too easy.* In other words, the game is interesting when predators (opponents) manage to kill the prey (player) sometimes but not always. In that sense, given a specific game structure and a player, highly effective opponent behaviors are not interesting behaviors and *vice versa*.
2. *When there is diversity in opponents' behavior between games.* That is, the game is interesting when NPCs are able to find dissimilar ways of hunting and killing the player in each game so that their strategy is more variable.
3. *When opponents' behavior is aggressive rather than static.* That is, the game is interesting when the predators move constantly all over the game world and cover it uniformly. This behavior gives the player the impression of an intelligent strategy for the opponents.

**Metrics.** These three general criteria must now be expressed in quantitative terms using data observable during game play. We therefore let a group of game opponents — the number of opponents depends on the specific game under examination — play the game  $N$  times (each game for a sufficiently large evaluation period of  $t_{max}$  steps) and record the steps  $t_k$  taken to kill the player in each game  $k$  as well as the total number of visits  $v_{ik}$  opponents make to each cell  $i$  of the game grid.

Given these data, quantifications of the three interest criteria proposed above can be presented as follows.

1. **Appropriate Level of Challenge.** The game is uninteresting when either the opponents consistently kill the player quickly (game too hard) or when the game consistently runs for long periods (game too easy). This criterion can be quantified by  $T$  in (1.1) below.

$$T = [1 - (E\{t_k\}/\max\{t_k\})]^{P_1} \quad (1.1)$$

where  $E\{t_k\}$  is the average number of simulation steps taken to kill the prey-player over the  $N$  games;  $\max\{t_k\}$  is the maximum  $t_k$  over the  $N$



games —  $\max\{t_k\} \leq t_{max}$ ; and  $p_1$  is a weighting parameter.  $T$  is high when  $E\{t_k\}$  is low compared to  $\max\{t_k\}$ , that is, when games occur that are much longer than average.

$p_1$  is adjusted so as to control the impact of the bracketed term in the formula for  $T$ . By selecting values of  $p_1 < 1$  we reward quite challenging opponents more than near-optimal killers, since we compress the  $T$  value toward 1.  $p_1$  is chosen as 0.5 for the experiments presented here.

The  $T$  estimate of interest demonstrates that the greater the difference between the average and the maximum number of steps taken to kill the player, the higher the interest of the game. Given (1.1), both easy-killing (*‘too easy’*) and near-optimal (*‘too hard’*) behaviors receive low interest estimate values (i.e.  $E\{t_k\} \simeq \max\{t_k\}$ ). This metric is also called ‘challenge’.

- 2. Behavior Diversity.** The game is interesting when the NPCs exhibit a diversity of behavior between games. One manifestation of this is that the time taken to kill the player varies between games. Thus a quantitative metric for this second criterion is given by  $S$  in (1.2) below.

$$S = (\sigma_{t_k} / \sigma_{max})^{p_2} \quad (1.2)$$

where

$$\sigma_{max} = \frac{1}{2} \sqrt{\frac{N}{(N-1)}} (t_{max} - t_{min}) \quad (1.3)$$

and  $\sigma_{t_k}$  is the standard deviation of  $t_k$  over the  $N$  games;  $\sigma_{max}$  is an estimate, based on the range of  $\{t_k\}$  of the maximum value of  $\sigma_{t_k}$ ;  $t_{min}$  is the minimum number of steps required for predators to kill the prey when playing against some ‘well-behaved’ fixed strategy near-optimal predators ( $t_{min} \leq t_k$ ); and  $p_2$  is a weighting parameter which is set so as  $\sigma_{t_k}$  has a linear effect on  $S$  ( $p_2 = 1$ ).

The  $S$  increases proportionally with the standard deviation of the steps taken to kill the player over  $N$  games. Therefore, using  $S$  as defined here, we promote predators that produce high diversity in the time taken to kill the prey.

- 3. Spatial Diversity.** The game is more interesting when opponents appear to move around actively rather than remain static or passively follow the player. A good measure for quantifying this criterion is through entropy of the opponents’ visits to the cells of the game grid during a game, since the entropy quantifies the completeness and uniformity with which the opponents cover the stage. Hence, for each game, the cell visit entropy is calculated and normalized into  $[0, 1]$  via (1.4).

$$H_n = \left[ -\frac{1}{\log V_n} \sum_i \frac{v_{in}}{V_n} \log \left( \frac{v_{in}}{V_n} \right) \right]^{p_3} \quad (1.4)$$

where  $V_n$  is the total number of visits to all visited cells (i.e.  $V_n = \sum_i v_{in}$ ) and  $p_3$  is a weighting parameter.  $p_3$  is adjusted in order to promote very

high  $H_n$  values and furthermore to emphasize the distinction between high and low normalized entropy values. Appropriate  $p_3$  parameter values which serve this purpose are those greater than one ( $p_3 = 4$  in this chapter), since they stretch the value of  $H_n$  away from 1.

Given the normalized entropy values  $H_n$  for all  $N$  games, the interest estimate for the third criterion can be represented by their average value  $E\{H_n\}$  over the  $N$  games. This implies that the higher the average entropy value, the more interesting the game is.

The three individual criterion metrics defined above are combined linearly to produce a single metric of interest (equation 1.5) whose properties match the qualitative considerations developed above.

$$I = \frac{\gamma T + \delta S + \epsilon E\{H_n\}}{\gamma + \delta + \epsilon} \quad (1.5)$$

where  $I$  is the interest value of the predator/prey game;  $\gamma, \delta$  and  $\epsilon$  are criterion weight parameters.

The approach to entertainment modeling represented by equation (1.5) is both innovative and efficient. However, it should be clear from the foregoing discussion that there are many possible formulae such as equation (1.5) which would be consistent with the qualitative criteria proposed for predator/prey games. Other successful quantitative metrics for the appropriate level of challenge, the opponents' diversity and the opponents' spatial diversity may be designed and more qualitative criteria may be inserted in the interest formula. Alternative mathematical functions for combining and weighting the various criteria could be employed.

For example, other metrics for measuring the appropriate level of challenge could be used: one could come up with a  $T$  metric assuming that the appropriate level of challenge follows a Gaussian distribution over  $E\{t_k\}$  and that the interest value of a given game varies depending on how long it is — *very short* ( $E\{t_k\} \approx t_{min}$ ) games tend to be frustrating and *long games* ( $E\{t_k\} \approx \max\{t_k\}$ ) tend to be boring. (However, very short games are not frequent in the experiments presented here and, therefore, by varying the weight parameter  $p_1$  in the proposed  $T$  metric (see (1.1)) we are able to obtain an adequate level of variation in measured challenge.)

To obtain values for the interest formula weighting parameters  $\gamma, \delta$  and  $\epsilon$  we select empirical values based on the specific game in question. For Pac-Man, spatial diversity of the opponents is of the greatest interest: the game no longer engages the player when *Ghosts* stick in a corner instead of wandering around the stage. Thus, diversity in game play ( $S$ ) and challenge ( $T$ ) should come next in the importance list of interest criteria. Given the above-mentioned statements and by adjusting these three parameters so that the interest value escalates as the opponent behavior changes from randomly generated (too easy) to near-optimal hunting (too difficult) and then to following *Ghost* behaviors, we come up with  $\gamma = 1, \delta = 2$  and  $\epsilon = 3$ .

The question remains, however, whether the number produced by such a formula really captures anything useful concerning a notion so potentially complex as human enjoyment. That question is addressed in section 1.4.2 below.

#### 1.4.1 Generality of the Metric

The interest metric introduced in equation (1.5) can be applied in principle to any predator/prey computer game because it is based on generic measurable features of this category of games. These features include the time required to kill the prey and the predators' entropy in the game field. Thus, it appears that (1.5) — or a similar measure based on the same concepts — constitutes a generic interest approximation of predator/prey computer games. Evidence demonstrating the interest metric's generality appears in [1.33] through successful application of the  $I$  value metric to two quite dissimilar predator/prey games.

Moreover, given the two first interest criteria previously defined, the approach can be generalized to all computer games. Indeed, no player likes any computer game that is too hard or too easy to play and, furthermore, any player would enjoy diversity throughout the play of any game. The third interest criterion is applicable to games where spatial diversity is important which, apart from predator/prey games, may also include action, strategy and team sports games according to the computer game genre classification of Laird and van Lent [1.3]. As long as game designers can determine and extract the measurable features of the opponent behavior that generate excitement for the player, and identify observable indices of them that can be computed from data collected during game play, a mathematical formula can be designed in order to collectively represent them.

Finally, a validation experiment like that presented in section 1.4.2 below can be used to assess the performance of the designed formula or test variants of it.

#### 1.4.2 Experiments

Given that the interest measure defined above has been constructed to reflect the designers' intuitions about those features of games that contribute to player interest and satisfaction, one would expect that games with higher values of  $I$  would be judged more satisfying by human players.

To investigate this, the Pac-Man game was used to acquire data on human judgement of entertainment. Thirty subjects (43.3% females, 56.7% males) whose age covered a range between 17 and 51 years participated in this experiment. In addition, all subjects spoke English (language of the survey questionnaire) as a foreign language since their nationality was either Danish (90%) or Greek (10%) [1.21, 1.22].

Subjects in the experiment played against five selected opponents differing in the  $I$  value they generate against a well-behaved computer-programmed player. Each player played several paired sets of games such that all pairwise combinations of the 5 opponents, in each order, were covered by the group of subjects (see [1.6] for details of the experimental design). For each pair, the player indicated which opponent generated the more satisfying game (players were not aware of the computed  $I$  values of the opponents they faced).

The degree of correlation between human judgement of entertainment and the computed  $I$  value is found by matching the entertainment preferences between the five opponents recorded by the human players and the  $I$  value ranking. According to the subjects' answers the  $I$  value is correlated highly with human judgement ( $r = 0.4444$ ,  $p\text{-value} = 1.17 \cdot 10^{-8}$  — see [1.21, 1.22]). These five opponents are used as a baseline for validating all entertainment modeling approaches presented in this chapter (see section 1.5.3).

In addition, players completed survey questions designed to elicit information about their likeliness for Pac-Man as a computer game, which allowed them to be grouped into three types: the ones that conceptually did not particularly like Pac-Man, the ones believed that Pac-Man is an interesting game and the ones that liked Pac-Man very much. It was found that neither the type of player nor the order in which opponents were encountered had any significant effect on the player's judgement of the relative interest of the games in a pair (see [1.21, 1.22] for full details of the analysis).

### 1.4.3 Conclusions

Given observations, intuitions and informal empirical studies on the predator/prey genre of games, generic criteria that contribute to the satisfaction for the player and map to measurable characteristics of the opponent behavior were defined.

According to the hypothesis of Yannakakis and Hallam [1.6], the player-opponent interaction — rather than the audiovisual features, the context, the mechanics, the control, the interface or the genre of the game — is the property that contributes the majority of the quality features of entertainment in a computer game. Given this fundamental assumption, a metric for measuring the real-time entertainment value of predator/prey games was motivated and designed. This value is built upon three generic entertainment criteria: appropriate level of challenge, opponents' behavior diversity and opponents' spatial diversity.

By testing predator/prey games with human players, it was confirmed that the interest value computed by equation (1.5) is consistent with human judgement of entertainment. In fact, the human player's notion of interest in the Pac-Man game seems to correlate highly with the computed interest metric, independently of player type or order of play [1.22].

Thus, we demonstrate the first approach to quantitative modeling of entertainment: the idea of using a custom-designed (but nevertheless quite

generic) mathematical expression that yields a numerical value well-correlated with human judgement. The metric formula rests on measurable features of opponent behavior computable from data collected during game play.

### 1.5 Quantitative Analysis of Entertainment Factors Derived from Psychological studies

The second approach to entertainment capture is in a sense inverse to the empirical, designer-driven, estimation of entertainment presented in section 1.4. In that approach, generic criteria describing interesting games were identified by intuition and consideration of game experiences on the part of the entertainment metric designer.

Although shown to yield a result that correlates well with human judgement, a more satisfying approach might be to start with the theoretical work on qualitative factors of entertainment reviewed in section 1.2 — specifically, Malone’s qualitative factors of entertainment: challenge and curiosity [1.9] — and attempt to construct a quantitative measure using them.

Quantitative measures for challenge and curiosity are inspired by previous work on entertainment metrics [1.6] and computed from data gathered during the real-time player-opponent interaction. The measured challenge and curiosity values are combined with human judgements on the relative entertainment of games using machine learning techniques to construct an expression analogous to equation 1.5 which is highly correlated with human choices. Again, the Pac-Man game is employed as a test-bed for the approach.

Two NN types, namely a feedforward NN and a fuzzy-neural network (fuzzy-NN), are used as alternatives to represent the entertainment metric. In each case, the inputs to the network are the measures of challenge and curiosity for a particular game (opponent). The output is an analogue of the  $I$  metric. The networks are thus used as function approximators for the expression defining the interest metric. Training of the approximators is done using artificial evolution.

The procedures used are described in more detail below. A comparison between the two alternatives is presented and the results are validated against and compared with the custom-designed metric of equation (1.5). The results of the study demonstrate that both NNs represent functions — possible interest metrics based on challenge and curiosity measures — whose qualitative features are consistent with Malone’s corresponding entertainment factors. Furthermore, the evolved feedforward NN provides a more accurate model of player satisfaction for Pac-Man than the custom-designed model (the  $I$  value), presented in the previous section, for this genre of games [1.33].

### 1.5.1 Experimental Data

As a final part of the experiment presented in section 1.4.2, subjects played a set of 25 games against each of two well-behaved opponents ( $A$  and  $B$ ). After the 50 games, the player records whether the first set or the second set of games was the more interesting, i.e. whether  $A$  or  $B$  generated a more interesting game. (The preference of  $A$  over  $B$  is also written as  $A \succ B$ .) To minimize order effects, each subject plays the aforementioned sets both orders during the experiment.

Given the recorded values of human playing times  $t_k$  over the 50 ( $2 \times 25$ ) games against a specific opponent, either  $A$  or  $B$ , the average playing time ( $E\{t_k\}$ ) and the standard deviation of playing times ( $\sigma\{t_k\}$ ) for all subjects are computed. We suggest that the  $E\{t_k\}$  and  $\sigma\{t_k\}$  values are appropriate measures to represent the level of challenge and the level of curiosity respectively [1.9] during game play. The former provides a notion for a goal whose attainment is uncertain (winning the game) — the lower the  $E\{t_k\}$  value, the higher the goal uncertainty and furthermore the higher the challenge — and the latter effectively portrays a notion of unpredictability in the subsequent events of the game — the higher the  $\sigma\{t_k\}$  value the more variable the game duration, so the higher the opponent unpredictability and therefore the higher the curiosity.

### 1.5.2 Tools

Two alternative neural network structures (a feedforward NN and a fuzzy-NN) for learning the relation between the challenge and curiosity factors and the entertainment value of a game have been used and are presented here. The assumption is that the entertainment value  $y$  of a given game is an unknown function of  $E\{t_k\}$  and  $\sigma\{t_k\}$ , which the NN will learn. The subjects' expressed preferences constrain but do not specify the values of  $y$  for individual games. Since there is no *a priori* target  $y$  values for any given game, the output error function is not differentiable, and ANN training algorithms such as back-propagation are inapplicable. Learning is achieved through artificial evolution [1.34] and is described in Section 1.5.2.

**Feedforward NN.** A fully-connected multi-layered feedforward NN has been evolved [1.34] for the experiments presented here. The sigmoid function is employed at each neuron, the connection weights take values from -5 to 5 and both input values are normalized into  $[0, 1]$  before they are entered into the feedforward NN. In an attempt to minimize the controller's size, it was determined that single hidden-layered NN architectures, containing 20 hidden neurons, are capable of successfully obtaining solutions of high fitness (network topology is not evolved, however).

**Fuzzy-NN.** A fuzzy [1.35] Sugeno-style [1.36] inference neural network is trained to develop fuzzy rules by evolving the memberships functions for both the input ( $E\{t_k\}$ ,  $\sigma\{t_k\}$ ) and the output variable  $y$  of the network as well as each fuzzy rule's weight. Each of the input and output values is presented by five fuzzy sets corresponding to very low, low, average, high and very high. The membership functions for the input values are triangular and their center  $\alpha$  and width  $\beta$  are evolved whereas the output fuzzy sets use singleton membership functions [1.36] — only the center  $\alpha$  of the spike membership function is evolved. The centroid technique is used as a defuzzification method.

**Genetic Algorithm.** A generational genetic algorithm (GA) [1.37] is implemented, which uses an “exogenous” evaluation function that promotes the minimization of the difference in matching the human judgement of entertainment. The feedforward NNs and fuzzy-NNs are themselves evolved. In the algorithm presented here, the evolving process is limited to the connection weights of the feedforward NN and the rule weights and membership function parameters of the fuzzy-NN.

The evolutionary procedure used can be described as follows. A population of  $N$  networks is initialized randomly. For feedforward NNs, initial real values that lie within  $[-5, 5]$  for their connection weights are picked randomly from a uniform distribution, whereas for the fuzzy-NNs, initial rule weight values equal 0.5 and their membership function parameter values lie within  $[0, 1]$  (uniformly distributed). Then, at each generation:

Step 1 Each member (neural network) of the population is evaluated with two pairs of ( $E\{t_k\}$ ,  $\sigma\{t_k\}$ ) values, one for  $A$  and one for  $B$ , and returns two output values, namely  $y_{j,A}$  (interest value of the game set against opponent  $A$ ) and  $y_{j,B}$  (interest value of the game set against opponent  $B$ ) for each pair  $j$  of sets played in the survey ( $N_s = 30$ ). If the numerical relationship between  $y_{j,A}$ ,  $y_{j,B}$  matches the preference of subject  $j$  then we state that: ‘the values agree with the subject’ throughout this chapter (e.g.  $y_{j,A} > y_{j,B}$  when subject  $j$  has expressed a preference for  $A$  over  $B$ :  $A \succ B$ ). In the opposite case, we state that: ‘the values disagree with the subject.’

Step 2 Each member  $i$  of the population is evaluated via the fitness function  $f_i$ :

$$f_i = \sum_{j=1}^{N_s} \begin{cases} 1, & \text{if } y_{j,A}, y_{j,B} \text{ agree with subject } j; \\ \left( \frac{1-D(A,B)}{2} \right)^2, & \text{if } y_{j,A}, y_{j,B} \text{ disagree with subject } j. \end{cases} \quad (1.6)$$

where  $D(A, B) = |y_{j,A} - y_{j,B}|$

Step 3 A fitness-proportional scheme is used as the selection method.

Step 4 Selected parents clone an equal number of offspring so that the total population reaches  $N$  members or reproduce offspring by crossover.

The Montana and Davis [1.38] crossover and the uniform crossover operator is applied for feedforward NNs and fuzzy-NNs respectively with a probability 0.4.

Step 5 Gaussian mutation occurs in each gene (connection weight) of each offspring's genome with a small probability  $p_m = 1/n$ , where  $n$  is the number of genes.

The algorithm is terminated when either a good solution (i.e.  $f_i > 29.0$ ) is found or a large number of generations  $g$  is completed ( $g = 10000$ ).

### 1.5.3 Results

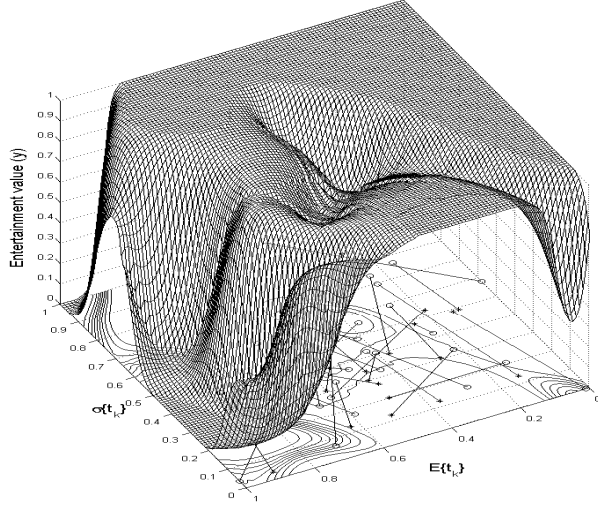
Results obtained from both feedforward NN and fuzzy-NN evolutionary approaches are presented in this section. In order to control for the non-deterministic effect of the GA initialization phase, each learning procedure (i.e. GA run) for each NN type is repeated ten times — we believe that this number is adequate to illustrate clearly the behavior of each mechanism — with different random initial conditions.

**Evolved Feedforward NN.** For space considerations, only the two fittest solutions achieved from the evolving feedforward NN approach are illustrated in Fig. 1.1. The qualitative features of the surfaces plotted in Fig. 1.1 appeared in all ten learning attempts. The most important conclusions derived from the feedforward NN mapping between  $E\{t_k\}$ ,  $\sigma\{t_k\}$  and entertainment are that:

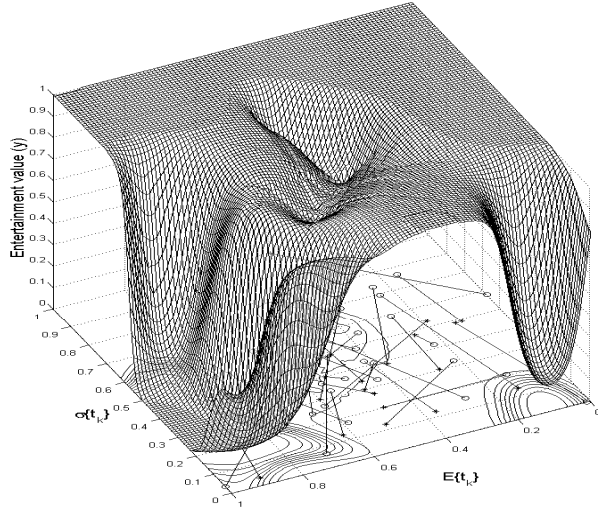
- Entertainment has a low value when challenge is too high ( $E\{t_k\} \approx 0$ ) and curiosity is low ( $\sigma\{t_k\} \approx 0$ ).
- Even if curiosity is low, if challenge is at an appropriate level ( $0.2 < E\{t_k\} < 0.6$ ), the game's entertainment value is high.
- If challenge is too low ( $E\{t_k\} > 0.6$ ), the game's entertainment value appears to drop, independently of the level of curiosity.
- There is only a single data point present when  $\sigma\{t_k\} > 0.8$  and the generalization of the evolved feedforward NNs within this space appears to be poor. Given that only one out of the 60 different game play data points falls in that region of the two-dimensional  $(E\{t_k\}, \sigma\{t_k\})$  space, we can hypothesize that there is low probability for a game to generate curiosity values higher than 0.8. Thus, this region can be safely considered insignificant for these experiments. However, more samples taken from a larger game play survey would be required to effectively validate this hypothesis.

The fittest evolved feedforward NN is also tested against the custom-designed  $I$  metric for cross-validation purposes. The feedforward NN ranks the five different opponents previously mentioned in section 1.4.2 in the order  $I_1 = I_2 < I_4 < I_3 < I_5$  (where  $I_i$  is the entertainment value the  $i$  opponent generates) which yields a correlation of 0.5432 (p-value =  $3.89 \cdot 10^{-12}$ )





**Fig. 1.1a.** The fittest feedforward NN solution ( $f = 29.95$ )



**Fig. 1.1b.** The second fittest feedforward NN solution ( $f = 29.67$ )

**Fig. 1.1a–b.** Circles ('o') and stars ('\*') represent  $E\{t_k\}$ ,  $\sigma\{t_k\}$  values obtained by playing against opponents  $A$  and  $B$  respectively. Straight lines are used to connect the sets of games that humans played in pairs

of agreement with human notion of entertainment expressed by the subject choices in the original experiment. Given this ranking of entertainment against these five opponents, the feedforward NN approach appears to model human entertainment better than the custom-designed interest metric proposed in [1.6, 1.22] and described above ( $r = 0.4444$ ,  $p\text{-value} = 1.17 \cdot 10^{-8}$ ).

The relationship between entertainment, challenge and curiosity expressed by the evolved feedforward NNs appears to follow the qualitative principles of Malone’s work [1.9] and the human-verified interest metric developed in our previous work [1.6] for predator/prey games. According to these, a game should maintain an appropriate level of challenge and curiosity in order to be entertaining. In other words, too difficult and/or too easy and/or too unpredictable and/or too predictable opponents to play against make the game uninteresting.

#### 1.5.4 Evolving Fuzzy-NN

The evolutionary procedure for the fuzzy-NN approach is also repeated ten times and only the fuzzy-NN that generates the highest fitness ( $f = 29.81$ ) is presented here for reasons of space. Twenty five fuzzy rules are initially designed based the conclusions derived from the evolved feedforward NNs. The fittest fuzzy-NN generates 19 fuzzy rules in total — rules with weight values less than 0.1 are not considered significant and therefore are excluded from further consideration — which are presented here with their corresponding weight values  $w$ :

- Entertainment is *very low* if (a) challenge is *very high* and curiosity is *low* (Rule 1;  $w_1 = 0.4440$ ) and (b) challenge is *low* and curiosity is *average* (Rule 2;  $w_2 = 0.3617$ ).
- Entertainment is *low* if (a) challenge is *very low* and curiosity is *average* (Rule 3;  $w_3 = 0.9897$ ) or *low* (Rule 4;  $w_4 = 0.7068$ ); (b) challenge is *low* and curiosity is *high* (Rule 5;  $w_5 = 0.7107$ ); (c) challenge is *high* and curiosity is *very low* (Rule 6;  $w_6 = 0.5389$ ) and (d) challenge is *very high* and curiosity is *very low* (Rule 7;  $w_7 = 0.9520$ ) or *high* (Rule 8;  $w_8 = 0.9449$ ).
- Entertainment is *average* if challenge is *very low* and curiosity is *high* (Rule 9;  $w_9 = 0.5818$ ).
- Entertainment is *high* if (a) challenge is *low* and curiosity is *very low* (Rule 10;  $w_{10} = 0.8498$ ) or *very high* (Rule 11;  $w_{11} = 0.2058$ ); (b) challenge is *average* and curiosity is *low* (Rule 12;  $w_{12} = 0.5$ ); (c) challenge is *high* and curiosity is *low* (Rule 13;  $w_{13} = 0.2824$ ) or *average* (Rule 14;  $w_{14} = 0.25$ ) and (d) challenge is *very high* and curiosity is *average* (Rule 15;  $w_{15} = 0.2103$ ).
- Entertainment is *very high* if (a) challenge is *very low* and curiosity is *very high* (Rule 16;  $w_{16} = 0.7386$ ); (b) challenge is *average* and curiosity is *very low* (Rule 17;  $w_{17} = 0.5571$ ) or *very high* (Rule 18;  $w_{18} = 0.8364$ ) and (c) challenge is *high* and curiosity is *high* (Rule 19;  $w_{19} = 0.2500$ ).

The quantitative measures of entertainment achieved through the neuro-fuzzy approach and the majority of the fuzzy rules generated appear consistent with Malone’s principles of challenge and curiosity, the empirical contributions of the interest metric from the literature [1.22] and the fittest feedforward NN presented in section 1.5.3. However, the fittest fuzzy-NN (being less fit than the fittest feedforward NN) generates some few fuzzy rules that are not consistent with the aforementioned principles — e.g. Rule 10: entertainment is *high* if challenge is *low* and curiosity is *very low*. It is not clear whether the poorer performance is intrinsic to the method or a result of unlucky initialization; further tests are needed to distinguish these alternatives.

The fuzzy-NN is tested against the  $I$  metric of section 1.4 as in the evolved feedforward NN approach. The evolved fuzzy-NN ranks the five opponents in the order  $I_2 < I_1 < I_3 < I_4 = I_5$ . This ranking demonstrates a correlation of 0.3870 (p-value =  $1.74006 \cdot 10^{-6}$ ) of agreement with human notion of entertainment, which appears to be lower than the correlation achieved through the  $I$  value of section 1.4 ( $r = 0.4444$ , p-value =  $1.17 \cdot 10^{-8}$  [1.6]). However, as in the feedforward NN approach, the generalization of the evolved fuzzy-NNs appears to be poor when  $\sigma\{t_k\} > 0.8$  due to the presence of a single data point within this region of the  $(E\{t_k\}, \sigma\{t_k\})$  space. Even though we consider this non-frequent region as insignificant as far as this work is concerned, it may be sampled from a more extensive human game experiment in a future study.

### 1.5.5 Conclusions

This section introduced an alternative approach to constructing a quantitative metric of entertainment motivated by the qualitative principles of Malone’s intrinsic factors for engaging game play [1.9]. More specifically, the quantitative impact of the factors of challenge and curiosity on human entertainment were investigated in the Pac-Man game.

The two neuro-evolution approaches for modeling entertainment examined demonstrate qualitative features that share principles with the interest metric ( $I$  value) presented in section 1.4. This (second) approach replaces the hand-crafted mathematical formulation of the interest metric with a more general process of machine learning applied to neural network models. Both obtained models manage to map successfully between the measures of entertainment factors such as challenge and curiosity and the notion of human game play satisfaction.

Validation results obtained show that the fittest feedforward NN gets closer — in the sense of statistical correlation — to the human notion of entertainment than both the  $I$  value [1.22] and the fittest fuzzy-NN. Therefore, it appears that the average and the standard deviation of a human’s playing time over a number of games are in themselves adequate, and in fact

more effective than the  $I$  value (as reported in [1.22]), for capturing player entertainment in real-time in predator/prey games.

The reported work on this approach is most significantly limited by the number of participants in the game survey we devised. Therefore, not all regions of the challenge-curiosity search space were sampled by human play which therefore yielded poor NN generalization for these regions. Limited data also restricted the sensible number of inputs to the learning system.

Malone’s entertainment factor of fantasy is omitted here since the focus is on the contribution of the opponent behaviors to the generation of entertainment; however, experiments on interactive physical predator/prey games with children have shown that entertainment increases monotonically with respect to the fantasy factor [1.39].

This second entertainment modeling approach presented here demonstrates generality over the majority of computer game genres since the quantitative means of challenge and curiosity are estimated through a generic feature of game play which is the playing time of humans over a number of games. Thus, these or similar measures could be used to measure player satisfaction in any genre of game. However, each game possesses additional idiosyncratic entertainment features that might need to be extracted and added to the proposed generic measures used as input to the machine learning tools — therefore, more games of the same or other genres need to be tested to evaluate the true generality of this approach.

## 1.6 Discussion

The foregoing has proposed and demonstrated a pair of methods for deriving a quantitative estimate of the level of entertainment experienced by a player of a computer game, using data that can be derived from or during game play. In this section, we discuss some of the questions raised by the approach and the assumptions on which it is based.

An immediate and natural question is whether the techniques described really capture “entertainment” which, after all, is a complex mental phenomenon depending on the player, the game and (probably) a number of external factors in rather involved ways. We acknowledge that the definition of a quantitative metric for entertainment in this sense is almost certainly infeasible. However, we take a practical approach here: it is sufficient for our purposes if a quantity exists that can be computed from observable data from the player-game interaction and that correlates well with players’ expressed preferences. In other words, a numerical value which orders games in the same way as players’ judgement of entertainment is sufficient for our purposes.

The foregoing material illustrates two ways to construct such a metric: by design, using the insights of a skilled game player; and by using machine learning to explore the space of possible evaluation functions whose values are consistent with human judgement of entertainment value. The resulting

metric is specific to the particular game under consideration, but the general method of construction is applicable to a wide variety of game instances and genres.

To summarize, therefore, the proposed approach does not capture details of the complex mental states associated with enjoyment of computer games, but it does provide a way to evaluate different instances of game play in terms of how entertaining they are for a player. Such knowledge can be used, for example, for tuning the game to suit the player (see below).

In addition to this general question concerning the approach, there are a number of assumptions (and hence limitations) associated with the methods presented; these are discussed below.

### 1.6.1 Assumptions and Limitations of the $I$ value

The interest metric described in section 1.4 is based on specific assumptions about the features of a game that generate enjoyment for the player.

- The approach assumes that interaction with opponent behavior is the primary source of variability in the entertainment value of a given game. That is, the enjoyment generated by the graphical, multimedia and story-line components of the game design is disregarded. This is a reasonable assumption for comparisons between instances of play of a single given game, but means that the interest metric is specific to a certain game and cannot be used for meaningful comparison between different games (e.g. to answer “Is Space Invaders more entertaining than Quake?”).
- The interest metric is calculated using data obtained from a sample of  $N$  games. This is consistent with human cognition since it appears that human players require a significant number of games (or else playing time) to classify a specific computer game according to their perceived satisfaction. However, this assumption constitutes a limitation of the method. A further investigation of the relationship between the  $I$  value and the  $N$  played games might reveal that fewer games are needed for an estimate that is still consistent with human notion of perceived entertainment.
- The interest value definition assumes that players of the game have average-level playing skills. By ‘average-level’ we only exclude the two following extreme player types: (1) those who have never played the specific game before and furthermore do not know the rules and how to play it — these players perform poorly against almost any type of opponent; (2) those who have an excellent knowledge of the specific game, can easily predict the opponent behavior and have mastered the game controls. These players can usually beat even the most effective opponents designed for the game. In each case, the interest value might not be very well estimated since the challenge criterion  $T$  approximates a zero value regardless of the opponent, in the first case because the game is much too hard and in the second because it is too easy (recall that  $T$  is designed to be maximum for a

‘reasonably difficult’ game). This appears to be consistent with human notion of interestingness since we believe that neither extreme player type will find the game interesting.

- The interest value depends primarily on the opponents’ behavior. Implicitly, through that, it depends on the player’s behavior since the opponent behavior is elicited in interaction with the player. (The previous point concerning playing skills is a specific instance of this more general observation). If the players of a game can be divided into classes with quite different playing styles, for example “aggressive” vs. “defensive”, then it may be necessary to design a separate  $I$  metric formula for each player type, because of the difference in opponent behavior elicited by their differing styles of play. For a comprehensive discussion of this assumption in [1.32] where the interest value dependence on the player is investigated through experiments in the Pac-Man test-bed game.
- A factor that may contribute to enjoyable games is the match between the real-time speed of the game and the reaction time of the player. Gradually decreasing the required player reaction time is a standard and inexpensive technique used by a set of games (i.e Pac-Man) to achieve increasing challenge for the player as the game proceeds. This is not considered in the work in this chapter since changing the demanded reaction time does not alter the qualitative properties of the opponent behavior (except through the implicit dependence on the player’s style of play). Note that in the extreme case, an unintelligent opponent may generate an exciting game just because of the unrealistically fast reaction time required of the player.

### 1.6.2 Assumptions and Limitations of the Machine Learning Approach

The second approach to constructing a metric uses machine learning rather than designer insight to build a quantitative evaluation of a game. This method is based on the same fundamental assumptions as the  $I$  value approach: that the opponents’ behavior is the primary determinant of the entertainment value of a given instance of game play. Many of the comments of the previous section also apply to this approach. However, there are a few observations specific to this methodology.

- The issue of playing style is arguably less relevant here than in the designer-insight method. If it is necessary to determine player type to be able to evaluate games, then the requirement of a consistent metric will in principle force the machine learning technique to perform an implicit player-type classification, assuming that the function representation technology (here, a neural network) is powerful enough to do so. In other words, because this approach can construct much more complex (and therefore less scrutable) mappings from raw measurements to entertainment metric value, it can

cope with more complex relationships between the data and the metric than a designer is likely to be able to manage.

- However, the effectiveness of a machine learning method depends strongly on the quantity and quality of data available. For the case considered here, this data comprises two kinds: raw measurements derived from game play, that represent aspects of features such as challenge, curiosity, fantasy, etc.; and expressed preferences or rankings between instances of game play. The former provide the observables on which the metric is built, the latter determine the degree of consistency with human judgement of any given proposal for the metric function. Obtaining the latter kind of data involves experimentation in which players are asked to say which of several (here two) instances of game play they prefer; collecting such data is time- and player- consuming. This limits the complexity of metric that can be considered, since it limits the feedback available to the machine learning process during the exploration of the space of metrics.
- The former kind of data also presents certain problems: specifically, how do we know what measurements to include as a basis for the metric? The work presented here uses one designer-chosen measurement for each relevant feature — challenge and curiosity — but one could in principle devise many measurements correlated with either feature, and allow the machine learning system to use all of them or to make a selection of the best measurements for the purpose. This approach removes a certain designer bias in the method at the expense of complicating the machine learning task and approaching earlier the limits imposed by the difficulty of collecting preference data from players.
- The issue of what value of  $N$  to choose can be finessed in this second approach, as can the question of game speed and demanded player reaction time, by appropriate choice of measurements from which to build the evaluation metric. For instance, game speed can be included directly as a measurement and measurements requiring different numbers of games to compute can be included in the process.

### 1.6.3 Making Use of Entertainment Metrics

Given a successful metric of entertainment for a given game, designed and evaluated using one of the methods proposed above, the final question we consider here is how such knowledge might be used. As previously noted, opponents which can learn and adapt to new playing strategies offer a richer interaction to entertain the player. An obvious use of the entertainment metric is therefore to adapt the game so that the metric value increases.

Two possible strategies for this might be:

- to use a machine learning mechanism for the game studied which allows opponents to learn while playing against the player (i.e. on-line). The entertainment metric can be used to guide the learning process.

Such an on-line learning mechanism is comprehensively described in [1.6, 1.33, 1.22]. Its basic steps are presented briefly here as follows. At each generation of the algorithm:

- Step 1: Each opponent is evaluated every  $e_p$  simulation steps via an individual reward function that provides an estimate of the  $I$  value of the game, while the game is played.
- Step 2: A pure elitism selection method is used where only a small percentage of the fittest opponents is able to breed. The fittest parents clone offspring.
- Step 3: Mutation occurs in each opponent (varying neural network controller connection weights) with a suitable small probability.
- Step 4: Each mutated offspring is evaluated briefly in off-line mode, that is, by replacing the least-fit member of the population and playing a short off-line game of  $e_p$  simulation steps against a selected computer-programmed player. The fitness values of the mutated offspring and the least-fit member are compared and the better one is deployed in the game.

The algorithm is terminated when a predetermined number of games has been played or a game of high interest (e.g.  $I \geq 0.7$ ) is found.

Results reported in [1.21, 1.22] demonstrate that 50 on-line learning games are not enough (in Pac-Man) for the on-line learning mechanism to cause a difference in the  $I$  value which is noticeable by human players. The on-line learning period of 50 games is an empirical choice to balance efficiency and experimental time. The duration of the on-line learning procedure in this experiment lasted 20 minutes on average while the whole human survey experiment presented in section 1.4.2 and section 1.5.1 exceeded 65 minutes in many cases, which is a great amount of time for a human to concentrate.

- to use a metric evaluation function constructed using the machine learning technique directly to enhance the entertainment provided by the game. The key to this is the observation that the models (feedforward NN or fuzzy-NN) relate game features to entertainment value. It is therefore possible in principle to infer what changes to game features will cause an increase in the interestingness of the game, and to adjust game parameters to make those changes. For the feedforward NN, the partial derivatives of  $\partial y / \partial E\{t_k\}$  and  $\partial y / \partial \sigma\{t_k\}$  indicate the change in entertainment for a small change in an individual game feature. One could use gradient ascent to attempt to improve entertainment with such a model. The fuzzy-NN approach provides qualitative rules relating game features to entertainment, rather than a quantitative function, but an analogous process could be applied to augment game entertainment.

Such a direction constitutes an example of future work within computer, physical and educational games. The level of engagement or motivation of the user/player/gamer of such interactive environments can be increased by



the use of the presented approaches providing systems of richer interaction and qualitative entertainment [1.5],

## Acknowledgements

This work was supported in part by the Danish Research Agency, Ministry of Science, Technology and Innovation (project no: 274-05-0511).

## Resources

### Key Books

- Koster, R., *A Theory of Fun for Game Design*, Paraglygh Press, 2005.

### Key Papers

- Andrade, G., Ramalho, G., Santana, H., Corruble, V. “Challenge-Sensitive Action Selection: an Application to Game Balancing,” in *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT-05)*, pp. 194–200, Compiegne, France. IEEE Computer Society, 2005.
- Iida, H., Takeshita, N., Yoshimura, J., “A metric for entertainment of boardgames: its implication for evolution of chess variants,” in *Nakatsu, R., Hoshino, J., eds.: IWEC2002 Proceedings*, pp. 65–72, Kluwer, 2003.
- Malone, T. W., “What makes computer games fun?,” *Byte*, vol. 6, pp. 258–277, 1981.
- Lazzaro, N., “Why we play games: Four keys to more emotion without story,” Technical report, XEO Design Inc. 2004.
- Yannakakis, G. N., Hallam, J., “Evolving Opponents for Interesting Interactive Computer Games,” in *Proceedings of the 8th International Conference on the Simulation of Adaptive Behavior (SAB’04); From Animals to Animats 8*, pp. 499–508, Los Angeles, CA, USA, July 13–17, 2004. The MIT Press.
- Yannakakis, G. N., Hallam, J., “Towards Optimizing Entertainment in Computer Games,” *Applied Artificial Intelligence*, 2007 (to appear).

## Discussion Groups, Forums

- Optimizing player satisfaction in games discussion group:  
<http://groups.google.com/group/optimizing-player-satisfaction-in-games>
- Player experience special interest group of DIGRA; list url:  
<http://mail.digra.org/mailman/listinfo/player-experience>

## Key International Conferences/Workshops

Workshop series on Optimizing Player Satisfaction. In conjunction with

- Simulation of Adaptive Behaviour (SAB) Conference, Rome, Italy, in 2006 and
- Artificial Intelligence and Interactive Digital Entertainment (AIIDE) Conference, Stanford, US, in 2007.

## References

- 1.1 Champandard, A.J.: AI Game Development. New Riders Publishing (2004)
- 1.2 Funge, J.D.: Artificial Intelligence for Computer Games. A. K. Peters Ltd, (Wellesley, Massachusetts, USA)
- 1.3 Laird, J.E., van Lent, M.: Human-level AI's killer application: Interactive computer games. In: Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI). (2000) 1171–1178
- 1.4 Vorderer, P., Hartmann, T., Klimmt, C.: Explaining the enjoyment of playing video games: the role of competition. In Marinelli, D., ed.: ICEC conference proceedings 2003: Essays on the future of interactive entertainment, Pittsburgh, (Carnegie Mellon University Press) 107–120
- 1.5 Yannakakis, G.N., Hallam, J.: A scheme for creating digital entertainment with substance. In: Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI). (2005) 119–124
- 1.6 Yannakakis, G.N., Hallam, J.: Evolving Opponents for Interesting Interactive Computer Games. In Schaal, S., Ijspeert, A., Billard, A., Vijayakumar, S., Hallam, J., Meyer, J.A., eds.: From Animals to Animats 8: Proceedings of the 8<sup>th</sup> International Conference on Simulation of Adaptive Behavior (SAB-04), Santa Monica, LA, CA, The MIT Press (2004) 499–508
- 1.7 Koster, R.: A Theory of Fun for Game Design. Paraglyph Press (2005)
- 1.8 Kapoor, A., Mota, S., Picard, R.: Towards a Learning Companion that Recognizes Affect. In: Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium . (2001)
- 1.9 Malone, T.W.: What makes computer games fun? Byte **6** (1981) 258–277
- 1.10 Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. New York: Harper & Row (1990)
- 1.11 Sweetser, P., Wyeth, P.: GameFlow: A Model for Evaluating Player Enjoyment in Games. ACM Computers in Entertainment **3** (2005)
- 1.12 Lazzaro, N.: Why we play games: Four keys to more emotion without story. Technical report, XEO Design Inc. (2004)
- 1.13 Read, J., MacFarlane, S., Cassey, C.: Endurability, engagement and expectations. In: Proceedings of International Conference for Interaction Design and Children. (2002)
- 1.14 Kline, S., Arlidge, A.: Online Gaming as Emergent Social Media: A Survey. Technical report, Media Analysis Laboratory, Simon Fraser University (2003)
- 1.15 Crispini, N.: Considering the growing popularity of online games: What contributes to making an online game attractive, addictive and compelling. Dissertation, SAE Institute, London (2003)

- 1.16 Choi, D., Kim, H., Kim, J.: Toward the construction of fun computer games: Differences in the views of developers and players. *Personal Technologies* **3** (1999) 92–104
- 1.17 Yannakakis, G.N., Lund, H.H., Hallam, J.: Modeling Children’s Entertainment in the Playware Playground. In: *Proceedings of the IEEE Symposium on Computational Intelligence and Games*, Reno, USA, IEEE (2006) 134–141
- 1.18 Nareyek, A.: Intelligent agents for computer games. In Marsland, T., Frank, I., eds.: *Computers and Games, Second International Conference, CG 2002*. (2002) 414–422
- 1.19 Taatgen, N.A., van Oploo, M., Braaksma, J., Niemantsverdriet, J.: How to construct a believable opponent using cognitive modeling in the game of set. In: *Proceedings of the fifth international conference on cognitive modeling*. (2003) 201–206
- 1.20 Iida, H., Takeshita, N., Yoshimura, J.: A metric for entertainment of boardgames: its implication for evolution of chess variants. In Nakatsu, R., Hoshino, J., eds.: *IWEC2002 Proceedings*, Kluwer (2003) 65–72
- 1.21 Yannakakis, G.N.: AI in Computer Games: Generating Interesting Interactive Opponents by the use of Evolutionary Computation. Ph.d. thesis, University of Edinburgh (2005)
- 1.22 Yannakakis, G.N., Hallam, J.: Towards Optimizing Entertainment in Computer Games. *Applied Artificial Intelligence* (2007) to appear.
- 1.23 Yannakakis, G.N., Hallam, J.: Towards Capturing and Enhancing Entertainment in Computer Games. In: *Proceedings of the 4<sup>th</sup> Hellenic Conference on Artificial Intelligence, Lecture Notes in Artificial Intelligence*. Volume 3955., Heraklion, Greece, Springer-Verlag (2006) 432–442
- 1.24 Andrade, G., Ramalho, G., Santana, H., Corruble, V.: Extending reinforcement learning to provide dynamic game balancing. In: *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)*. (2005) 7–12
- 1.25 Verma, M.A., McOwan, P.W.: An adaptive methodology for synthesising Mobile Phone Games using Genetic Algorithms. In: *Congress on Evolutionary Computation (CEC-05)*, Edinburgh, UK (2005) 528–535
- 1.26 Hunicke, R., Chapman, V.: AI for Dynamic Difficulty Adjustment in Games. In: *Proceedings of the Challenges in Game AI Workshop, 19<sup>th</sup> Nineteenth National Conference on Artificial Intelligence (AAAI’04)*. (2004)
- 1.27 Spronck, P., Sprinkhuizen-Kuyper, I., Postma, E.: Difficulty Scaling of Game AI. In: *Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-ON 2004)*. (2004) 33–37
- 1.28 Yannakakis, G.N., Hallam, J., Lund, H.H.: Capturing Entertainment through Heart-rate Dynamics in the Playware Playground. In: *Proceedings of the 5<sup>th</sup> International Conference on Entertainment Computing, Lecture Notes in Computer Science*. Volume 4161., Cambridge, UK, Springer-Verlag (2006) 314–317
- 1.29 Rani, P., Sarkar, N., Liu, C.: Maintaining optimal challenge in computer games through real-time physiological feedback. In: *Proceedings of the 11<sup>th</sup> International Conference on Human Computer Interaction*. (2005)
- 1.30 McQuiggan, S., Lee, S., Lester, J.: Predicting User Physiological Response for Interactive Environments: An Inductive Approach. In: *Proceedings of the 2<sup>nd</sup> Artificial Intelligence for Interactive Digital Entertainment Conference*. (2006) 60–65
- 1.31 Yannakakis, G.N., Hallam, J.: A generic approach for generating interesting interactive pac-man opponents. In Kendall, G., Lucas, S., eds.: *Proceedings*

- of the IEEE Symposium on Computational Intelligence and Games, Essex University, Colchester, UK (2005) 94–101
- 1.32 Yannakakis, G.N., Maragoudakis, M.: Player modeling impact on player's entertainment in computer games. In: Proceedings of the 10<sup>th</sup> International Conference on User Modeling; Lecture Notes in Computer Science. Volume 3538., Edinburgh, Springer-Verlag (2005) 74–78
  - 1.33 Yannakakis, G.N., Hallam, J.: A Generic Approach for Obtaining Higher Entertainment in Predator/Prey Computer Games. *Journal of Game Development* **1** (2005) 23–50
  - 1.34 Yao, X.: Evolving artificial neural networks. In: Proceedings of the IEEE. Volume 87. (1999) 1423–1447
  - 1.35 Zadeh, L.: Fuzzy sets. *Information and Control* **8** (1965) 338–353
  - 1.36 Sugeno, M.: *Industrial Applications of Fuzzy Control*. North-Holland (1985)
  - 1.37 Holland, J.H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI (1975)
  - 1.38 Montana, D.J., Davis, L.D.: Training feedforward neural networks using genetic algorithms. In: Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89), San Mateo, CA, Morgan Kaufman (1989) 762–767
  - 1.39 Yannakakis, G.N., Hallam, J., Lund, H.H.: Comparative Fun Analysis in the Innovative Playware Game Platform. In: Proceedings of the 1<sup>st</sup> World Conference for Fun 'n Games. (2006) 64–70