Measuring entertainment and automatic generation of entertaining games

Zahid Halim*, A. Rauf Baig and
Hasan Mujtaba

FAST-National University of Computer and Emerging Science,
Islamabad, Pakistan
E-mail: zahid.halim@nu.edu.pk
E-mail: rauf.baig@nu.edu.pk
E-mail: hasan.mujtaba@nu.edu.pk
*Corresponding author

Abstract: Over the period of time computer games have became a major source of entertainment for humans. From the point of view of game developers there is a constant demand of writing games which are entertaining for the end users but entertainment itself is of subjective nature. It has always been difficult to quantify the entertainment value of the human player. The two factors which mainly influence the entertainment value are the type of the game and the contents of the game. In this paper we address the issues of measuring entertainment and automatic generation of computer games. We present some quantitative measures for entertainment in a genre of computer game and apply them as a guide for the evolution of new interesting games.

Keywords: predator/prey games; automatic game creation; genetic algorithms; entertainment metrics for games.

Reference to this paper should be made as follows: Halim, Z., Baig, A.R. and Mujtaba, H. (2010) 'Measuring entertainment and automatic generation of entertaining games', *Int. J. Information Technology, Communications and Convergence*, Vol. 1, No. 1, pp.92–107.

Biographical notes: Zahid Halim is working as an Assistant Professor at FAST-National University of Computer and Emerging Sciences, Islamabad and also a PhD student there. His research interests include computational intelligence in games, content evolution for games and improving replay value of games using computational intelligence based techniques.

Rauf Baig is a Professor at FAST-National University of Computer and Emerging Sciences, Islamabad. His areas of interest include computational intelligence (particularly swarm intelligence), machine learning, intelligent systems, applications of CI in games and applications of CI in data mining.

Hasan Mujtaba is a HEC sponsored, PhD student at FAST-National University of Computer and Emerging Sciences, Islamabad. His research interests include computational intelligence, machine learning and intelligent systems in games.

1 Introduction

Computer games have nowadays captured the interest of every one in general and young generation in particular. This means entertainment has become one of the major sources of relaxation and fun for children and adults. There can be many reasons for the popularity of such games some of which include the higher amusement capacity of these games due to their graphics, increasing levels of challenge and choice. According to a survey conducted in Olson et al. (2007) on 1,254 subjects, only 80 were found playing no electronic games in the last six months. The results in Olson et al. (2007) show the popularity of computer and video games in young generation. The responsibility of defining the environment and the rules of the game lies on the shoulders of the game creator. Recently efforts have been made to evolve the rules of the game (Togelius and Schmidhuber, 2008). Even though still in its infancy, this direction of research seems to have much potential and can lead to exciting applications.

In Togelius and Schmidhuber (2008) the evolution of games is guided by a fitness function based on 'learning ability'. It gives low scores to games that do not require any skill to play and also to those which are hard and impossible. It assigns high fitness to games which can be learnt quickly. In our present work we propose and utilise some other metrics for measuring entertainment and interestingness and utilise them for guiding the evolution of games.

From the point of view of game developers it has always been a tough task to define and measure entertainment, as entertainment is very subjective. There have been many studies on entertainment in computer games (Lucas and Kendall, 2006; Malone, 1980; McQuiggan et al., 2006; Rani, 2006; Roberts et al., 2007; Sweetser and Wyeth, 2005; Takeshita and Yoshimura, 2003; Togelius et al., 2006; Togelius and Schmidhuber, 2008; Yannakakis and Hallam, 2004, 2005, 2007a, 2007b; Yannakakis et al., 2008). Value of entertainment depends on the genre of game, among other things, as different genre of games have different environment, players, number of players and other features. To classify computer games at a higher level of abstraction following are the four categories:

a classical perfect information games

b classical imperfect information games

c video games

d real world games (Lucas and Kendall, 2006).

A recently added genre is that of physically activating games (Yannakakis et al., 2008). There are many games under these categories. All this makes it difficult to make a single quantitative criterion to measure entertainment in any genre of games.

Three representative papers of relevant research work on measuring entertainment are Takeshita and Yoshimura (2003), Togelius et al. (2006) and Yannakakis and Hallam (2007a). In Takeshita and Yoshimura (2003) the authors present their work on entertainment metrics for board based games. In Togelius et al. (2006) work on automatic construction of tracks to increase the value of entertainment in car racing games is presented. In Yannakakis and Hallam (2007a) entertainment metrics to measure entertainment for predator/prey type of games are presented.

In this paper we present entertainment metrics for a predator-prey genre of game developed in Togelius and Schmidhuber (2008). Our entertainment metrics considers four quantitative measures of entertainment:

a duration of play

b level of challenge

c diversity in artefact's behaviour

d usability of the play area.

Based on these factors we propose a combined metrics of entertainment. We further evolve a game's rules based on the proposed entertainment metrics. These metrics are proposed based on different theories on entertainment, which are covered in literature review section of this paper.

The paper is organised as follows: Section 2 covers the literature survey, Section 3 lists the search space we have used to evolve game rules, Section 4 explains the chromosome encoding of the rules, Section 5 explains the entertainment metrics and the fitness function based on them, Section 6 is about the rule based controller to play the game, Section 7 contains the experiments and their results, Section 8 covers the user survey and Section 9 concludes our work.

2 Literature review

The concept of measuring entertainment and automatic generation of games and/or its contents is fresh and quite a limited amount of literature is available on the topic. In Togelius and Schmidhuber (2008) an effort has been made to evolve rules of the game. The evolution of games in Togelius and Schmidhuber (2008) is guided by a fitness function based on 'learning ability'. It gives low scores to games that do not require any skill to play and also to those which are hard and impossible where as it assigns high fitness to games which can be learnt quickly. Although there are games being created automatically but they are not being measured against their entertainment value present in the game due to its rules and contents. They employ theory of artificial curiosity based fitness function introduced in Lucas and Kendall (2006) which focuses on the predictability of the game environment.

There have been many other studies on entertainment in computer games (Malone, 1980; McQuiggan et al., 2006; Roberts et al., 2007; Yannakakis and Hallam, 2004, 2007b). One of the interesting facts mentioned in Lucas and Kendall (2006) states that the majority of human users dislike the evolutionary computation and artificial intelligence used in games to be hard enough to beat. People like more sort of entertaining games rather those difficult games to play. The work done in Rani (2006) uses challenge as a tool the make computer games more entertaining for the purpose of gaining human player's attention. The game difficulty level is altered based upon a biofeedback from the human user. A model to measure player's entertainment named gameflow is proposed in Sweetser and Wyeth (2005). The proposed model is composed of eight factors which include:

1 concentration
2 challenge
3 skills
4 control
5 clear goals
6 feedback
7 immersion
8 social interaction.

An entertainment metrics created based upon a survey for the car racing games is presented in Togelius et al. (2006). The same metrics is also used to create racing tracks at runtime for the human players based upon player's playing patterns. Work presented in Yannakakis and Hallam (2004, 2007a, 2007b) and Yannakakis et al. (2008) propose a generic criteria for measuring entertainment value of predator/prey genre of games based upon three criterions which include:

1 appropriate level of challenge
2 behaviour diversity
3 spatial diversity.

In Iida et al. (2003) the authors present their work on entertainment metrics for board based games.

Based upon the discussion above on previous work done in the context of measuring entertainment in computer games it is evident that the value of entertainment depends on the genre of game, among other things, as different genre of games have different environment, players, number of players and other features. In order to identify different genre of games, following are the four genres of computer games classified at a higher level of abstraction:

a classical perfect information games, which include chess and checkers
b classical imperfect information games for example monopoly
c video games, like car racing games
d real world games, DEFCON (2009) is an example of this genre of games (Lucas and Kendall, 2006).

A recently added genre is that of physically activating games (Yannakakis et al., 2008). There are many games under each of the above mentioned categories. Figure 1 lists some of the most popular games under each category. All this makes it difficult to make a single quantitative criterion to measure entertainment that addresses all genres of computer games. The work done so far is either limited to measuring the entertainment and evolving the components in the games based upon the measured entertainment as in Togelius and Schmidhuber (2008) and Yannakakis and Hallam (2005) or the proposed metric consists of so many criterions that it is not appropriate to be applied to every genre of game. Based upon this survey we intend to first propose and entertainment metric for the predator/prey genre of games and then generate the games itself rather than the contents of the game or manipulating the entities in the game. The aim is to generate games that are more entertaining for the human players.

3 Search space for evolving game rules

For using an evolutionary algorithm to search for a set of acceptable game rules we first need to define the space in which the search has to take place. For our current experiments, we have used the search space and the corresponding framework presented in Togelius and Schmidhuber (2008) with some modifications. The framework for the game is as follows:

- Play area consists of 14 × 14 grid cells, which are empty except for a couple of walls at fixed positions and of size seven cells each (Figure 1).
- There is one player called agent. There are N other artefacts of M types. The range of N can be from 0–20 and for M it is 0–3. For visual display and for description purposes each type is represented by a different colour: red, green and blue.
- The duration of the game is fixed at 100 game steps. However, the game may finish earlier in the event of agent's death.
- Each type of artefact can move around the play area according to one of the following five schemes. No movement; keep on moving straight until reaching a wall from where turn right and then again keep on moving straight; keep on moving straight until reaching a wall from where turn left and then again keep on moving straight; randomly pick one adjoining cell and try to move into that cell; randomly pick one direction and take n steps in that direction (n is also a randomly picked number between 1 to 10) and afterwards again pick a random direction and steps. It may be noted that an artefact can take zero or one step at a time. Taking n steps would require n time steps. The artefacts cannot move into a cell occupied by a wall.
- At any time step, the agent must move in one of the four adjoining cells (up, down, right, left). However, it cannot move into a wall occupied by a wall.
- The artefacts may collide with one another and the agent. The collision has one of the following possible effects on each colliding entity: no effect; random relocation; death.
- The collisions between entities cause a change in the score. The effect of the collision can be an increment of +1 or a decrement of –1 or no effect. The provision of score change due to two artefacts colliding together, without the intervention of the agent, introduces an element of uncertainty (good or bad luck) in the game.

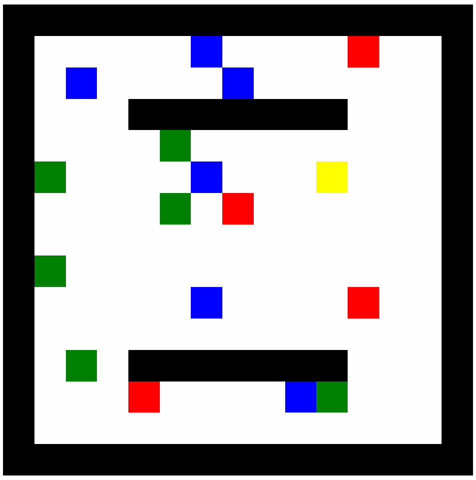
The space in which rules can evolve consists of the following dimensions or axes.

- types of artefacts: 0, 1, 2 or 3
- number of artefacts of each type: possible values are from 0 to 20
- movement logic for each artefacts: possible values are from one to five (one of the five movement strategies described above: no movement, right turn, etc.)

- collision effect possibilities: one possibility from the three described before (no effect, teleport, death)
- score effect: one possibility from +1, 0, and –1.

Figure 1 displays a typical environment of the game.

Figure 1 The play area of the predator-prey game (see online version for colours)



4 Chromosome encoding

Each individual chromosome in the evolutionary algorithm's population represents one complete set of rules for the game. Each gene of the chromosome represents one rule of the game. In our case we have 30 genes in a chromosome. The rule of the game they represent and their possible values are as follows:

1 One gene each for representing the number of red, green and blue artefacts (total three genes). The possible values for these genes are 0 to 20.

2 One gene each for representing the movement logic of red, green and blue artefacts (total three genes). The possible values for these genes are one to five (still, clockwise, counterclockwise, random short, random long).

3 A total of 15 genes for representing collision logic between two artefacts or between an artefact and an agent. Since there are four entities (three artefacts and an agent), hence the possible affects of collision between any two entities can be represented by a 4 × 4 matrix. Collisions between an agent and another agent are not possible because there is only one agent. Hence one element of the 4 × 4 matrix is empty and need not be represented as a gene in the chromosome. The possible values of the collision logic genes are 0 (nothing happens), 1 (the entity dies and is removed from the game) and 2 (the artefact is teleported to some randomly chosen location).

4 A total of nine genes for representing the score addition or depletion on the collision of any two entities. The possible values of theses genes are –1, 0 and 1.

The nine genes represent score effect for collision between: agent and red artefact, agent and green artefact, agent and blue artefact, red and red artefact, red and green artefact, red and blue artefact, green and green artefact, green and blue artefact, and blue and blue artefact. The chromosome encoding is shown in Table 1.

Table 1 Chromosome encoding: the interpretation of 30 genes and their possible values

Number of predators	Red	0–20	*Collision logic*	Blue-green	0–2
	Green	0–20		Blue-blue	0–2
	Blue	0–20		Blue-agent	0–2
Movement logic	Red	0–4		Agent-red	0–2
	Green	0–4		Agent-green	0–2
	Blue	0–4		Agent-blue	0–2
Collision logic	Red-red	0–2	*Score logic*	Red-red	–1,0,+1
	Red-green	0–2		Green-green	–1,0,+1
	Red-blue	0–2		Blue-blue	–1,0,+1
	Red-agent	0–2		Agent-red	–1,0,+1
	Green-red	0–2		Agent-green	–1,0,+1
	Green-green	0–2		Agent-blue	–1,0,+1
	Green-blue	0–2		Green-red	–1,0,+1
	Green-agent	0–2		Blue-red	–1,0,+1
	Blue-red	0–2		Blue-green	–1,0,+1

5 Fitness function

The aim of the evolutionary process is to evolve a game which is entertaining for the player. We have developed a fitness function which can guide the evolution towards interesting games by considering the following four aspects.

5.1 Duration of the game

In general, a game should not be too short or too long, as both are boring. For example, if chess is played between a novice and a master, the game is bound to be short and uninteresting for both. Similarly, if a game continues for a long duration, then its players might lose interest. For the game described in this paper, the maximum steps are fixed at 100 (takes three to five minutes if played with arrow keys). However, there is a possibility that one or more evolving rules may allow the death of the agent, thus prematurely terminating the game. The death possibility of the agent should not be very high because in that case the resulting games would be very short and frustrating for the player. The fitness function should be such that it discourages such a possibility.

The duration of play (D) is calculated as the average life span of the agent over N games played according the rules encoded in a chromosome. Since there are many probabilistic factors in the game, hence we will not get the same value of D if the game is played multiple times. For this purpose the game is played N times (N = 10 in our experiments) for a chromosome and an average is taken.

5.2 Appropriate level of challenge

The game is interesting if the rules of the game are such that the player has to display intelligence and dexterity to obtain a reasonable score. In such a case, higher a player scores more is his interest and motivation to play the game again. Too high a score, achieved easily is not challenging enough and similarly too low a score even after an intelligent game play is discouraging. There has to be an appropriate level of challenge provided by the rules.

There is a provision in the game that score may increment or decrement due to collision of artefacts without any involvement of the agent. This causes the introduction of good or bad luck in the game (just like snake and ladders game) and can enhance the enjoyment level, provided that this uncertainty factor is not too high.

The challenge (C) is converted into a fitness function by the formula:

$$C = \exp[-\mathrm{abs}(\text{max score} - AS)/\text{max score}] \tag{1}$$

where AS is average score of the agent calculated by playing ten games according the rules encoded in a chromosome. Since the value of AS can be negative also, hence we use the following processing:

$$\begin{aligned} AS &= AS, \text{ if } AS \geq 0 \\ AS &= \mathrm{abs}(AS) + 20, \text{ otherwise} \end{aligned} \tag{2}$$

The value of 'maximum score' is set to 20 according to the following reasoning. Since there are 100 steps possible and since any score increasing event can increment the score by a fixed value of 1, hence a maximum score of 20 would mean that an average of five steps are needed to score 1 point (assuming that the player manages to avoid any decrements in the score). This calculation also provides for discouraging the excessive increment or decrements of scores due to artefacts colliding together independent of the agent. The fitness function returns a high value in the vicinity of a score of 20 and gradually decreases for higher and lesser scores.

5.3 C. Diversity in the artefact's behaviour

The behaviour of the moving artefacts of the game should be sufficiently diverse so that it cannot be easily predicted. Artefacts with complex movement logic including teleportation would be more entertaining than static or simply moving ones. The diversity is captured by the following fitness function:

$$Div = \sum_{i=1,\ldots N} (\text{Number of cell changes made by artefact i during a game}) \tag{3}$$

where N is the total number of artefacts specified in a chromosome. The diversity is averaged by calculating it for ten games for the same chromosome. This fitness function has a higher value if the artefacts are dynamic rather than static.

5.4 Usability of the play area

It is interesting to have the play area maximally utilised during the game. If most of the moving artefacts remain in a certain region of the play area then the resulting game may seem strange. The usability is captured by the formula:

$$U = \sum_{\text{all cells}} (\text{usability counter value for a cell}) \tag{4}$$

A usability counter is set up for each cell which increments when an artefact arrives in the cell. The usability U is averaged by playing ten different games for a chromosome. The total cells for our current experimentation are 14 × 14 minus the 14 cells used by walls. A cell which is never visited during a game will have a counter value of zero, thus contributing nothing to the usability formula. Furthermore, a cell which has a few visits would contribute less than a cell having large number of visits. This formula also captures the diversity aspect, described above.

5.5 Combined fitness function

The above four metrics are combined in the following manner. All chromosomes in a population are evaluated separately according to each of the four fitness functions. Then the population is sorted on 'duration of play' and a rank based fitness is assigned to each chromosome. The best chromosome of the sorted population is assigned the highest fitness (in our case it is 20 because we have ten parents and ten offspring), the second best chromosome is assigned the second best fitness (in our case 19), and so on. The population is again sorted on the basis of 'challenge' and a rank based fitness is assigned to each chromosome. Similarly, rank based fitness is assigned after sorting on 'diversity' and 'usability'. The four rank based fitness values obtained for each chromosome are multiplied by corresponding weights and then added to get its final fitness.

$$FF = aD + bC + cDiv + dU \tag{5}$$

where a, b, c, and d are constants. In our experiments we keep the value of these constants fixed at 1. The multiplication with a corresponding weight allows us to control the relative influence of an aspect.

The calculation of rank based fitness gets rid of the problem of one factor having higher possible values than another factor.

6 The controller

We need to determine the fitness of, at least, several hundred chromosomes during the evolutionary process. It is impossible for a human player to play that many games. Hence an automatic controller is required to play as the agent. Such a controller can be implemented as an artificial neural network (ANN). Since the training of the ANN for each different game (chromosome) requires an extra effort and since the type of controller is not important for our fitness function, we have implemented the controller as a human supplied rule set. The same controller is used for playing all games (chromosomes) during the entire evolutionary process. Our rule based agent controller is composed of rules formulated to implement the following policy.

According to the game rules, at each simulation step the agent must take exactly one step. The agent looks up, down, left and right. It notes the nearest artefact (if any) in each of the four directions, and then it simply moves one step towards the nearest score increasing artefact. If there are no score increasing artefacts present it determines its step according to the following priority list:

a move in the direction which is completely empty (there is only the wall at the end). If more than one direction are empty move towards the farthest wall (in the hope that subsequent position changes would show it a score increasing artefact)

b move in the direction which contains a score neutral artefact. The farthest, the better

c move in the direction which contains a score decreasing artefact. The farthest, the better

d move in the direction which contains a death causing artefact. The farthest, the better.

Going into walls is not allowed, and if there is a wall present in the adjoining cell, the possibility of going in that direction is automatically curtailed.

The above mentioned controller rules encourage the agent to maximise its score by trying to collide with the artefacts which increase its score and at the same time try to avoid collision with the rest.

7 Experimentation and results

We use a population of ten chromosomes, randomly initialised. In each generation one offspring is created for each chromosome by duplicating it and then mutating any one of its gene. All genes have equal probability of being selected. The mutation is done by replacing the existing value with some other permissible value. All permissible values have equal probability of being selected. The parents and offspring form a pool of 20 chromosomes from which ten best are selected for the next generation. This evolutionary process is continued for 100 generations.

The fitness of a chromosome, in our case, is based on data obtained by playing the game according to the rules encoded in the chromosome. Since there are several probabilistic components in the game, hence the data obtained from playing the same game twice is never the same. To mitigate the effect of this noisiness in the fitness, we take the average of ten games for every fitness evaluation. To analyse the effect of each of the four proposed metrics of entertainment individually we have evolved four different populations. Each of these populations is guided by one of the proposed fitness function. We have also evolved a population using the combined fitness function.

7.1 Duration of the game

Considering the duration of the game play as determined by the average life span of the agent we rapidly evolve a population of chromosome in which there are very less possibilities for the agent to die in the allotted 100 steps. Table 2 shows one such evolved chromosome. Note that the agent must die if it collides with a red artefact, but the number of red artefacts is zero.

7.2 Appropriate level of challenge

Considering this metric alone, we were able to evolve chromosomes which favoured a score of 10 to 20. An example is shown in Table 3. According to this chromosome the score is incremented if the agent collides with green or blue artefacts. There are no blue artefacts (column 3) and 17 green artefacts (column 2). The green artefacts remain still and the agent can move towards the nearest one, because there are no impeding artefacts. There are five red artefacts, but they are dynamic and their collision logic causes their death upon collision with the green artefacts. There is no effect on the green artefacts when the red ones collide with them. Since there are 17 green artefacts, hence the red ones are cleared away from the game at an early stage.

Table 2 The best chromosome evolved by optimising 'duration of game'

Number of predators	Red	0	*Collision logic*	Blue-green	1
	Green	2		Blue-blue	0
	Blue	4		Blue-agent	0
Movement logic	Red	3		Agent-red	1
	Green	3		Agent-green	0
	Blue	0		Agent-blue	2
Collision logic	Red-red	0	*Score logic*	Red-red	0
	Red-green	2		Green-green	0
	Red-blue	1		Blue-blue	0
	Red-agent	2		Agent-red	0
	Green-red	1		Agent-green	1
	Green-green	0		Agent-blue	1
	Green-blue	1		Green-red	0
	Green-agent	0		Blue-red	0
	Blue-red	0		Blue-green	0

Table 3 The best chromosome evolved by optimising 'challenge'

Number of predators	Red	5	*Collision logic*	Blue-green	0
	Green	17		Blue-blue	0
	Blue	0		Blue-agent	2
Movement logic	Red	4		Agent-red	1
	Green	0		Agent-green	0
	Blue	4		Agent-blue	2
Collision logic	Red-red	0	*Score logic*	Red-red	0
	Red-green	2		Green-green	0
	Red-blue	0		Blue-blue	0
	Red-agent	1		Agent-red	0
	Green-red	0		Agent-green	1
	Green-green	0		Agent-blue	1
	Green-blue	2		Green-red	0
	Green-agent	0		Blue-red	0
	Blue-red	1		Blue-green	0

7.3 Diversity in the artefact's behaviour

When the diversity is considered in isolation, the solutions evolved tend to have somewhat higher number of artefacts. Another tendency is to have one of the dynamic movement logics. Yet another observation was that the death of agent was avoided in most of the better chromosomes so that the game may continue for maximum number of steps. A sample chromosome is shown in Table 4.

Table 4 The best chromosome evolved by optimising 'diversity of behaviour'

Number of predators	Red	0	*Collision logic*	Blue-green	0
	Green	15		Blue-blue	0
	Blue	10		Blue-agent	0
Movement logic	Red	3		Agent-red	1
	Green	4		Agent-green	0
	Blue	3		Agent-blue	2
Collision logic	Red-red	0	*Score logic*	Red-red	–1
	Red-green	2		Green-green	0
	Red-blue	0		Blue-blue	1
	Red-agent	1		Agent-red	0
	Green-red	0		Agent-green	1
	Green-green	0		Agent-blue	1
	Green-blue	2		Green-red	0
	Green-agent	2		Blue-red	-1
	Blue-red	0		Blue-green	0

7.4 Usability of the play area

The chromosomes evolved by using usability of the play area metric seem similar to that for diversity metric. One such chromosome is shown in Table 5.

Table 5 The best chromosome evolved by optimising 'diversity of behaviour'

Number of predators	Red	0	*Collision logic*	Blue-green	0
	Green	19		Blue-blue	0
	Blue	20		Blue-agent	0
Movement logic	Red	0		Agent-red	1
	Green	3		Agent-green	0
	Blue	4		Agent-blue	2
Collision logic	Red-red	0	*Score logic*	Red-red	0
	Red-green	0		Green-green	0
	Red-Blue	1		Blue-blue	0
	Red-agent	2		Agent-red	0
	Green-red	2		Agent-green	1
	Green-green	0		Agent-blue	1
	Green-blue	2		Green-red	0
	Green-agent	2		Blue-red	0
	Blue-red	2		Blue-green	0

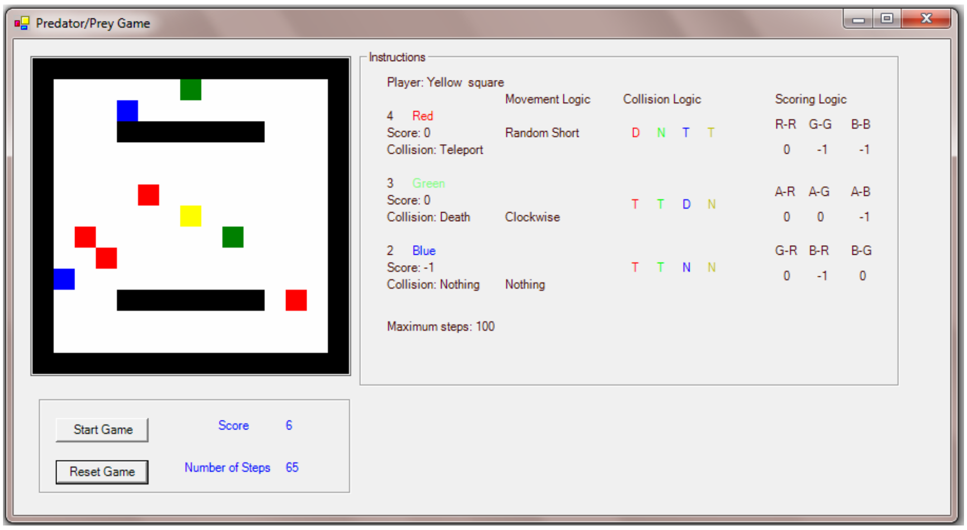
7.5 *The combined fitness function*

The above mentioned combined fitness function was used to evolve a population of chromosome. Most of the resulting evolved chromosomes are playable and seem interesting. They are much better than random games. Also they seem better than some games evolved using individual components of the fitness function. However, an extensive user survey is needed to verify and quantify these observations. A chromosome showing the best evolved chromosome is shown in Table 6.

Table 6 The best chromosome evolved by optimising the combined fitness function

Number of predators	Red	2	*Collision logic*	Blue-green	0
	Green	20		Blue-blue	0
	Blue	0		Blue-agent	1
Movement logic	Red	0		Agent-red	1
	Green	3		Agent-green	0
	Blue	4		Agent-blue	2
Collision logic	Red-red	0	*Score logic*	Red-red	0
	Red-green	1		Green-green	0
	Red-blue	0		Blue-blue	0
	Red-agent	0		Agent-red	0
	Green-red	0		Agent-green	1
	Green-green	0		Agent-blue	1
	Green-blue	2		Green-red	0
	Green-agent	1		Blue-red	0
	Blue-red	1		Blue-green	0

Figure 2 The interface for playing the evolved games (see online version for colours)



Strangely the fact that we had a suite of games formed by the better chromosomes of the evolved population seemed itself to be the best source of entertainment. The provision of a different game delays the boredom from setting in. We suppose that this is due to the simplicity of the basic game rules and changing of not only the environment but also the rules adds a sort of challenge which engages the interest of the player. The interface for playing the game is shown in Figure 2.

8 User survey

To verify the results obtained above and counter check them with the entertainment values assigned by the human players we conducted a user survey. The GA was run for 100 iterations and the best chromosome in the archive was given to the human users (20) to assess the game. The criterion was binary 1 means that the game was entertaining and 0 means that the game was not entertaining. Table 7 summarises the data collected from the user survey. The survey suggests that 70% of the users found the evolved game based on the combined fitness function interesting (column titled FF). We also made a human survey of games evolved using the individual components of fitness in isolation. These results are also presented in Table 7. Only 10% of the users found the games entertaining evolved on duration of the game metrics (column title D), 35% said evolved game were entertaining when evolved against appropriate level of challenge (column titled C), 20% and 15% games were found entertaining using diversity in artefact's behaviour and usability of the play area as metrics respectively (columns titled Div and U).

Table 7 Result of the user survey

	Human user survey				
	D	*C*	*Div*	*U*	*FF*
User 1	1	0	1	0	1
User 2	0	1	0	0	1
User 3	0	0	0	0	0
User 4	0	0	1	1	1
User 5	0	0	0	0	0
User 6	0	0	1	0	1
User 7	0	0	0	0	1
User 8	0	0	0	0	1
User 9	0	1	0	0	1
User 10	0	0	0	0	0
User 11	0	0	0	1	1
User 12	0	1	0	0	0
User 13	0	0	0	0	1
User 14	1	0	0	0	1
User 15	0	1	0	1	1
User 16	0	1	1	0	1
User 17	0	0	0	0	0
User 18	0	1	0	0	1
User 19	0	1	0	0	1
User 20	0	0	0	0	0
Percentage	10	35	20	15	70

9 Conclusions

The idea of evolution of game rules to produce new games seems to be very promising. In this work we have presented an experiment for evolution of a simple computer game. In the process we have presented some metrics for entertainment which are based on duration of play, level of challenge, diversity in artefact's behaviour, and usability of play area. These metrics are combined in a fitness function to guide the search for evolving the rules of the game. The result of our experiment is that games can be evolved in this manner. Even though we cannot claim that the best evolved game is the most entertaining of all, but still we are able to evolve a suite of equally interesting games which are much better than randomly generated games.

Further work needs to be done to make the proposed entertainment metrics more generic so they can automatically cover more types of computer games. The automatic game creation approach can be applied to other genre of games. We are currently conducting experiments for evolution of rules for board games. Some other direction could be to use co-evolution where one population tries to evolve the rules and other tries to evolve the strategy to play on those rules. In the context of entertainment metrics in our case the four fitness criterion are linearly combined, an alternate would be multi objective genetic algorithm.

References

DEFCON, available at http://www.introversion.co.uk/defcon/.

Iida, H., Takeshita, N. and Yoshimura, J. (2003) 'A metric for entertainment of boardgames: its implication for evolution of chess variants', *IWEC2002 Proceedings*.

Lucas, S.M. and Kendall, G. (2006) 'Evolutionary computation and games', *IEEE Computational Intelligence Magazine*, Vol. 1, pp.10–18.

Malone, T.W. (1980) 'What makes things fun to learn? Heuristics for designing instructional computer games', *Proceedings of the 3rd Acm Sigsmall Symposium and the First Sigpc Symposium on Small Systems*, pp.162–169.

McQuiggan, S., Lee S. and Lester, J. (2006) 'Predicting user physiological response for interactive environments: an inductive approach', *Proceedings of the 2nd Artificial Intelligence for Interactive Digital Entertainment Conference*, AAAI Press, pp.60–65.

Olson, C.K., Kutner, L.A., Warner, D.E., Almerigi, J.B., Baer, L., Nicholi Jr., A.M. and Beresin, E.V. (2007) 'Factors correlated with violent video game use by adolescent boys and girls', *Journal of Adolescent Health*, pp.77–83.

Rani, P., Sarkar, N. and Liu, C. (2006) 'Maintaining optimal challenge in computer games through real-time physiological feedback', in Dylan D. Schmorrow (Ed.): *Chapter 4, Task Specific Information Processing in Operational and Virtual Environments, Foundations of Augmented Cognition*, pp.184–192, Lawrence Erlbaum Associates Publishers.

Roberts, D.L., Strong, C.R. and Isbell, C.L. (2007) 'Estimating player satisfaction through the author's eyes', *Proceedings of the AIIDE'07 Workshop on Optimizing Player Satisfaction, AAAI Press Technical Report WS-01-01*, June, AAAI Press, Stanford, USA, pp.31–36.

Schmidhuber, J. (2006) 'Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts', *Connection Science*, Vol. 18, pp.173–187.

Sweetser, P. and Wyeth, P. (2005) 'Gameflow: a model for evaluating player enjoyment in games', *ACM Computers in Entertainment*, Vol. 3.

Takeshita, N. and Yoshimura, J. (2003) 'A metric for entertainment of board games: its implication for evolution of chess variants', *Proc. IWEC2002*, pp.65–72.

Togelius, J. and Schmidhuber, J. (2008) 'An experiment in automatic game design', *Proceedings of IEEE Computational Intelligence and Games*, Australia.

Togelius, J., De Nardi, R. and Lucas, S.M. (2006) 'Making racing fun through player modeling and track evolution', *Proceedings of the SAB'06 Workshop on Adaptive Approaches for Optimizing Player Satisfaction in Computer and Physical Games*.

Yannakakis, G.N. and Hallam, J. (2004) 'Interesting games through stage complexity and topology', *8th International Conference on the Simulation of Adaptive Behavior SAB'04*, July 13–17, Los Angeles, CA, USA.

Yannakakis, G.N. and Hallam, J. (2005) 'A generic approach for obtaining higher entertainment in predator/prey computer games', *Journal of Game Development*, December, Vol. 1, No. 3, pp.23–50.

Yannakakis, G.N. and Hallam, J. (2007a) 'Towards optimizing entertainment in computer games', *Applied Artificial Intelligence*, Vol. 21, pp.933–971.

Yannakakis, G.N. and Hallam, J. (2007b) 'Modeling and augmenting game entertainment through challenge and curiosity', *International Journal on Artificial Intelligence Tools*, December, Vol. 16, No. 6, pp.981–999.

Yannakakis, G.N., Hallam, J. and Lund, H.H. (2008) 'Entertainment capture through heart rate activity in physical interactive playgrounds', *User Modeling and User-Adapted Interaction*, Special Issue on Affective Modeling and Adaptation, February, Vol. 18, Nos. 1–2, pp.207–243.