Sequential Task Problem Solving using Cultural Learning in Populations of Neural Networks

Dara Curran and Colm O'Riordan
Department of Information Technology,
NUI, Galway

Abstract. Population learning can be described as the iterative Darwinian process of fitness–based selection and genetic transfer of information leading to populations of higher fitness. Life–time learning describes the process of learning undertaken by individuals in a population during their lifetime. These two learning models are often simulated using genetic algorithms and neural networks, respectively, and can be used to evolve efficient neural networks for a variety of tasks.

Cultural learning describes the process of information transfer between individuals in a population through non–genetic means. Typically this is achieved through communication or the creation of artifacts available to all members of a population. Cultural learning has been simulated by combining genetic algorithms and neural networks using a teacher/pupil scenario where highly fit individuals are selected as teachers and instruct the next generation.

This paper explores the effect of a cultural learning approach to the development of solutions for three test–case sequential decision tasks: connect–four, tic–tac–toe and blackjack. Our cultural learning model allows individuals to vertically impart knowledge acquired during their lifetime. Experiments are conducted with populations employing population learning alone and populations combining population and cultural learning.

Keywords: Cultural Learning, Neural Networks, Sequential Decision Tasks, Games, Artificial Life

1. Introduction

Some research has been performed with regard to the combination of both population and life–time learning approaches(Nolfi and Parisi, 1995; Floreano and Mondada, 1989; S. Nolfi, 1994; Sasaki and Tokoro, 1997; Pereira and Costa, 2001; Watson and Wiles, 2002; Curran and O'Riordan, 2003b; Curran and O'Riordan, 2003a), thus combining the global search power of the underlying genetic algorithm and the finer local search capabilities of gradient descent techniques. Empirically, the combined approach proves to be successful, as the populations tend to converge faster towards a global optimum.

Cultural learning is an alternative model which combines population learning with a modified version of life–time learning that allows populations to pass on knowledge to the next generation through non–genetic means through a process of communication or artifact creation, often achieved through imitation. Much research has been conducted in the field of imitation, particularly with respect to robotics and symbol grounding in animals and artifacts(Billard and Hayes, 1997; Billard and Dautenhahn, 1999; Dautenhahn and Nehaniv, 2002; Hayes and Demiris, 1994; Demiris and Hayes, 1996) and a number of models have been developed to examine the interaction of culture and evolution(Cavalli-Sforza, 1981; Cavalli-Sforza, 1982; Boyd and Richerson, 1985).

In addition, the simulation of culture in populations of artificial organisms has been the focus of much research(Yanco and Stein, 1993; De Jong, 1999; Batali, 1998; Denaro and Parisi, 1996a; Spector and Luke, 1996; Oliphant and Batali, 1997; Kirby and Hurford, 1997a; Saunders and Pollack, 1994; Kirby and Hurford, 1997b; Best, 1999; de Boer and Vogt, 1999; Cangelosi, 1999; Brighton and Kirby, 2001; Borenstein and Ruppin, 2003).

One method of simulating cultural learning is the teacher/pupil approach(Billard and Hayes, 1997; Denaro and Parisi, 1996a; Cangelosi and Parisi, 1996). Highly fit individuals from the population (teachers) are allowed to instruct the next generation (pupils). In this way, important information gained through the life–time of the previous generation is not lost completely and the fitness of the entire population can be improved.

In this paper, we select three sequential decision bench–mark problems (tic–tac–toe, blackjack and connect–four) to examine the effect of cultural learning in populations of neural networks. While previous work has attempted to evolve game–playing agents using a variety of games(Moriarty and Miikkulainen, 1995; Weaver and Bossomaier, 1996; Chellapilla and Fogel, 1999; Richards et al., 1997), none have explicitly employed game–playing as a test–bed for cultural learning experiments.

The aim of the work is to determine whether cultural learning is capable of enhancing the performance of the population in a similar manner to life–time learning. In particular, we are interested in observing the effect of cultural learning on the innate fitness of individuals within the population.

In this sense, this work is similar to recent work by Borenstein & Ruppin who explored the effect of cultural learning on the innate fitness of populations for a variety of problem solving tasks(Borenstein and Ruppin, 2003). However, while their work allowed individuals to impart only innate knowledge, our framework allows individuals to transmit information acquired culturally.

A series of experiments were conducted using populations employing population–learning alone and populations employing both population and cultural learning. While the effects of cultural learning have been examined before, the results presented diverge from previous work and shed more light on the subject.

The remainder of this paper is arranged as follows: Section 2 summarises related research and background material. Section 3 presents the artificial life simulator employed to conduct the experiments. Section 4 describes the experiments carried out: tic–tac–toe (Section 4.1), blackjack (Section 4.2) and connect–four (Section 4.3). Section 5 provides a discussion of these experiments and Section 6 presents conclusions.

2. Related work

The following section outlines some background material including learning models and sequential decision tasks.

2.1. Learning Models

A number of learning models can be identified from observation in nature. These can roughly be classified into population, life–time and cultural learning.

2.1.1. *Population Learning*

Population learning refers to the process whereby a population of organisms evolves, or learns, by genetic means through a Darwinian process of iterated selection and reproduction of fit individuals. In this model, the learning process is strictly confined to each organisms genetic material: the organism itself does not contribute to its survival through any learning or adaptation process.

2.1.2. *Life–time Learning*

There exist species in nature that are capable of learning, or adapting to environmental changes and novel situations at an individual level. Such learning, known as life–time learning is often coupled with population–based learning, further enhancing the populations fitness through its adaptability and resistance to change.

Another phenomenon related to life–time learning, first reported by Baldwin(Baldwin, 1896), occurs when certain behaviour discovered through life–time learning becomes imprinted onto an individuals genetic material through the evolutionary processes of crossover and mutation. To quote Hinton and Nowlan whose model ((Hinton and

Nowlan, 1987)) was the first to demonstrate this effect through simulation, "learning can provide an easy evolutionary path towards co–adapted alleles in environments that have no good evolutionary path for non–learning organisms".

Subsequent work has further explored the interactions between evolution and learning and shown that the addition of individual lifetime learning can improve a population's fitness(Nolfi and Parisi, 1995; Floreano and Mondada, 1989; S. Nolfi, 1994; Sasaki and Tokoro, 1997; Pereira and Costa, 2001; Watson and Wiles, 2002; Curran and O'Riordan, 2003b; Curran and O'Riordan, 2003a).

2.1.3. *Cultural Learning*

Culture can be succinctly described as a process of information transfer within a population that occurs without the use of genetic material. Culture can take many forms such as language, signals or artifactual materials. Such information exchange occurs during the lifetime of individuals in a population and can greatly enhance the behaviour of such species. Because these exchanges occur during an individuals lifetime, cultural learning can be considered a subset of lifetime learning.

A number of approaches have been implemented to simulate cultural learning including fixed lexicons(Yanco and Stein, 1993; Cangelosi and Parisi, 1996), indexed memory(Spector, 1994), cultural artifacts (Hutchins and Hazlehurst, 1991; Cangelosi, 1999) and signal–situation tables(MacLennan and Burghardt, 1993).

The approach chosen was inspired by the teacher/pupil scenario (Billard and Hayes, 1997; Denaro and Parisi, 1996b; Cangelosi and Parisi, 1996) where a number of highly fit agents are selected from the population to act as teachers for the next generation.

Experiments conducted by Hutchins and Hazlehurst(Hutchins and Hazlehurst, 1995) simulate cultural evolution through the use of a hidden layer within an individual neural network in the population. The networks respond to environmental features through their output units but in addition, the outputs of the hidden layer are employed in teaching. Pupils learn from teachers by observing the teacher's hidden layer output and attempting to mimic it through error back–propagation.

In previous work by Parisi *et al.*(Denaro and Parisi, 1996b), it was suggested that the addition of noise to a teacher's output could enhance a population's ability to retain culturally acquired information. Experiment conducted in our previous work(Curran and O'Riordan, 2004) confirmed that small levels of noise introduced to the communication process improved agent performance.

2.2. Sequential Decision Tasks

Sequential decision tasks are a complex class of problem that require agents to make iterative decisions at many steps throughout the task. Each decision has a direct effect on the agents environment and in turn affects its subsequent decisions. Our selection of a number of games was driven by two main factors: games are good examples of sequential decision tasks and many artificial intelligence implementations exist for ready comparison and analysis.

The games we chose as a test–bed for cultural learning are ordered by perceived difficulty, beginning with tic–tac–toe followed by the game of blackjack and concluding with the game of connect–four.

3. Simulator Architecture

The simulator implements population, lifetime and cultural learning. Population learning is simulated using a genetic algorithm which generates successive generations using three operators: selection, crossover and mutation. The algorithm employs an encoding scheme (described in Section 3.4) to convert genetic codes to neural network structures.

Lifetime learning, the acquisition of knowledge occurring during the lifetime of each individual, is simulated using neural networks. Each agent in the population is equipped with a neural network responsible for its perception and response to the environment. The neural network structure is derived from an individual's gene code at birth.

Cultural learning is implemented using a *vertical* cultural transmission model (Boyd and Richerson, 1985; Belew, 1990) inspired by Hutchins and Hazlehurst's model. The approach employs the last hidden layer of each agent's neural network as teaching apparatus. As an agent encounters stimuli in its environment, it responds both behaviourially (emitting a signal through its output nodes) and 'verbally' (emitting a signal through its teaching nodes). Unlike Hutchins and Hazlehurst's model, the model employed for this work allows the number of teaching nodes to evolve along with the network structure. Thus, no limitations are imposed on the communication complexity available to the population.

At the end of each generation, a percentage of the population's fittest networks are selected and are allowed to become teachers for the next generation. The teaching process takes place as follows: a teacher is stochastically assigned n pupils from the population where $n = \frac{N_{pop}}{N_{teachers}}$, N_{pop} being population size and $N_{teachers}$ the number of teachers.

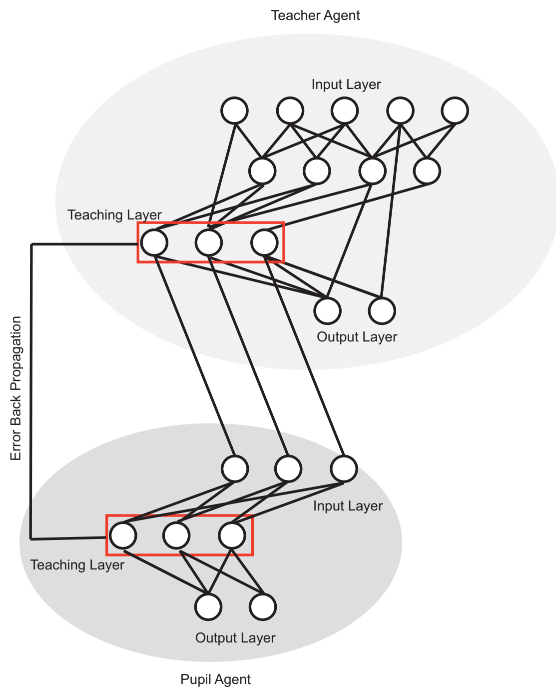

Figure 1. Agent Communication Architecture

Each pupil follows the teacher in its environment and observes the teacher's teaching output as it interacts with its environment. Both teacher and pupil receive environmental stimuli and respond with teaching signals. A teaching cycle occurs when the pupil's output is corrected to more closely resemble the teacher's using error back–propagation. Once the number of required teaching cycles is completed, the teachers die and the pupils are released into their environment. At the end of their lifetime, the fittest pupils are themselves selected to become teachers for the next generation and can impart the knowledge acquired through previous cultural exchanges. Thus, information is passed down through successive generations.

3.1. Encoding Scheme

One of the most crucial aspects of the simulator is the translation of genetic codes to neural network structures. Many encoding schemes were considered in preparation of the simulator, prioritising flexibility, scalability, difficulty and efficiency. These included Connectionist Encoding(Belew et al., 1992), Node Based Encoding (White, 1994), Graph Based Encoding(Pujol and Poli, 1998), Layer Based Encoding(Mandischer, 1993), Marker Based Encoding(Moriarty and Miikkulainen, 1995), Matrix Re–writing(Kitano, 1990; Miller et al., 1989), Cellular Encoding(Gruau, 1994), Weight–based encoding(Sutton, 1986; Kolen and Pollack, 1991) and Architecture encoding (Koza and Rice, 1991).

The scheme chosen is inspired by Marker Based Encoding which allows any number of nodes and interconnecting links for each network giving a large number of possible neural network architecture permutations. In addition, the scheme allows any number of hidden layers to be evolved.

Marker based encoding represents neural network elements (nodes and links) in a sequential list. Each element is separated by a marker to allow the decoding mechanism to distinguish between the different types of element and therefore deduce interconnections(Moriarty and Miikkulainen, 1995).

	Start Marker	Node Label	Threshold	Link to Node	Link Weight	Link to Node	Link Weight	End Marker	
...	SM	5	0.8	4	0.83	3	-0.51	EM	...

Figure 2. Marker Based Encoding

In this implementation, a marker is given for every node in a network. Following the node marker, the node's details are stored in sequential order on the bit string. This includes the node's label and its threshold value. Immediately following the node's details, is another marker which indicates the start of one or more node–weight pairs. Each of these pairs indicate a back connection from the node to other nodes in the network along with the connection's weight value. Once the last connection has been encoded, the scheme places an end marker to indicate the end of the node's encoding.

A consequence of the encoding scheme is that the size of each agent's chromosome is directly related to the size of its neural network (the number of nodes and links contained within it). In addition, there is no requirement for genotypes to be the same size across the population.

Two parents possessing gene codes of different lengths will produce two offspring whose gene code lengths are equal to that of their parents'. Thus, if parent A has a large genome of length a and parent B has a smaller genome of length b, one of their offspring will have a genome of length a while the other will have one of length b.

Depending on the problem domain, smaller or larger neural network architectures may be beneficial to the population, and the encoding allows the evolutionary process the freedom to determine gene code sizes by selecting appropriate parents.

3.2. Crossover

As a result of the chosen encoding scheme, crossover may not operate at the bit level as this could result in the generation of invalid gene codes. Therefore, the crossover points are restricted to specific intervals – only whole node or link values may be crossed over.

Two–point crossover is employed in this implementation. Once crossover points are selected, the gene portions are swapped. The connections within each portion remain intact, but it is necessary to adjust the connections on either side of the portion to successfully integrate it into the existing gene code. This is achieved by using node labels for each node in the network. These labels are used to identify individual nodes and to indicate the location of interconnections.

Once the portion is inserted, all interconnecting links within the whole gene code are examined. If any links are now pointing to non–existing nodes, they are modified to point to the nearest labeled node. In effect, this link realignment is a form of link mutation, as the re–attachment of crossed over network segments generated variation.

3.3. Mutation

The mutation operator introduces additional noise into the genetic algorithm process thereby allowing potentially useful and unexplored regions of problem space to be probed. The mutation operator usually functions by making alterations on the gene code itself, most typically by altering specific values randomly selected from the entire gene code. In this implementation, weight mutation is employed. The operator modifies a weight according to a random percentage value chosen randomly from the range -200% to +200%.

4. Experiments

The games represent the environment in which the agents live. To be successful, an agent must become sufficiently skilled to play each game adequately.

4.1. Tic Tac Toe

Tic–tac–toe, or three in a row is a very simple two player game played on a 3x3 board. Each player is assigned either the X or O symbol and takes turns placing one symbol onto the board at a time. Each player attempts to place three of his/her pieces in a horizontal, vertical or diagonal line of three.

Agents play games against a modified minimax player, whose first move is randomized, allowing agents to play games of some variety. Fitness is assigned according to the length of the game. In other words, agents are rewarded for bringing the game to as close to a draw as possible, as it is very unlikely that an agent will beat the modified minimax player.

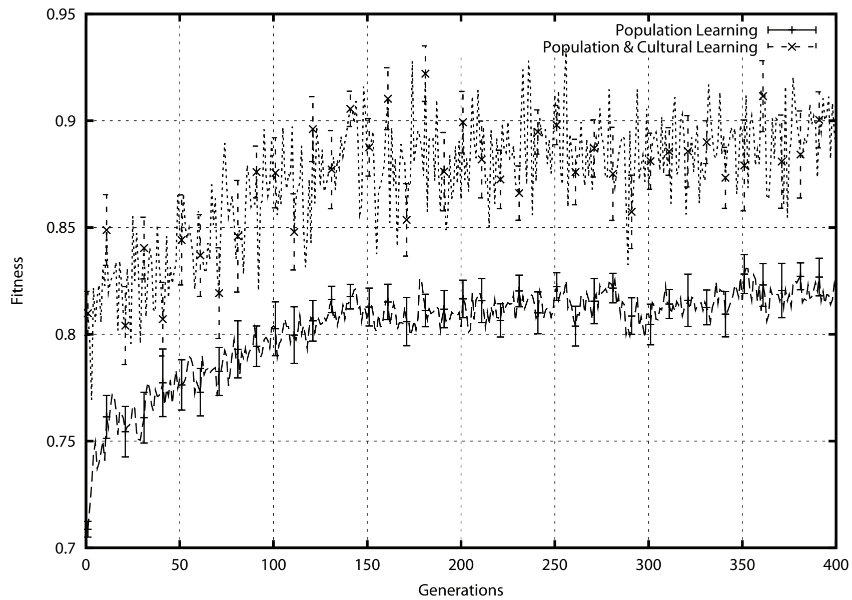

Figure 3. Tic–Tac–Toe Population Fitness

Each agent's neural network structure contains 18 input nodes, 2 for each board position where 01 is X, 10 is O and 11 is an empty square. Nine output nodes corresponding to each board position are used to indicate the agent's desired move. The node with the strongest response corresponding to a valid move is taken as the agent's choice.

The simulator allows agents to evolve any number of hidden layers each with an unrestricted number of nodes, giving maximum flexibility to the evolutionary process. During the teaching process, a teacher agent plays alongside the pupil. At each move, both the pupil and teacher emit some teaching output in response to the current board position. At every teaching cycle, the pupils teaching output is corrected with respect to the teachers using error back–propagation.

Populations of 100 agents were generated for these experiments and allowed to evolve for 250 generations. Crossover was set at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was set at 0.05. These parameters were determined empirically to provide good performance.

4.1.1. *Experimental Results*

Figure 3 shows the average fitness of two populations throughout the experiment run along with error bars showing fitness variance within populations. It is clear from these results that the population additionally employing cultural learning out–performs the population employing population learning alone from the start of the experiment. While the population learning population stabilises at around 0.825, the population employing cultural learning achieves fitness values of 0.9 and above.

Interestingly, it appears that the population employing population learning alone is less diverse in its fitness variance than that of the population employing both cultural and population learning. Cultural learning appears to be producing individuals with fitness ranges larger than that of population learning alone. On average, even the worst individual in the cultural learning population is performing better than the best individual in the population learning population. Increased diversity, particularly in early generations, is generally seen as a benefit as the search space is more thoroughly examined, leading to superior solutions.

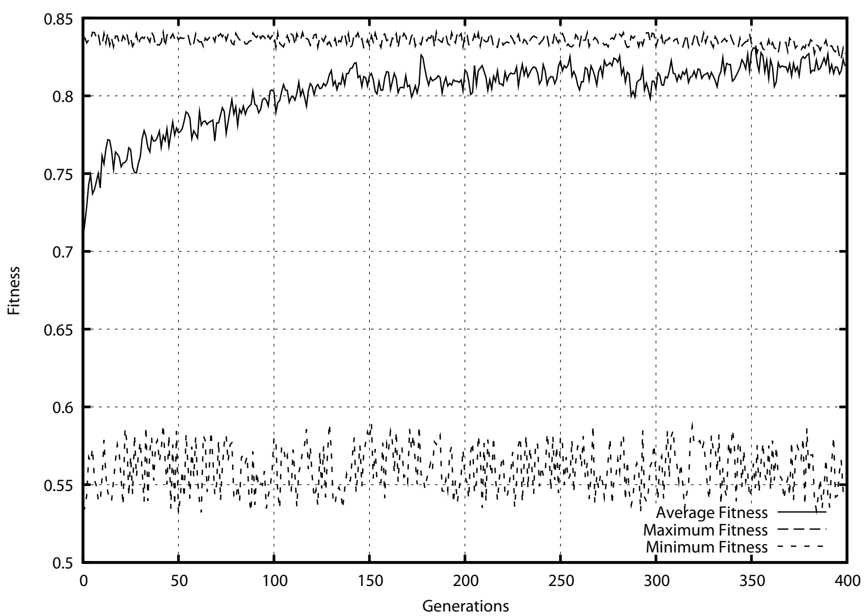

Figure 4. Tic–Tac–Toe Average, Maximum and Minimum Fitness for Non–Teaching Population

Table I. Tic–Tac–Toe Average Fitness

Population	**Avg. Fitness**	**Max Fitness**	**Min Fitness**	**S. D.**
Pop. Learning	0.8029272	0.8308780	0.7086666	0.0004040
Cultural Learning	0.8732511	0.9337738	0.7692220	0.0007648

Figure 4 shows the average, maximum and minimum fitness values for the cultural learning population. Minimum and maximum values represent the average best and worst individual in the population. While both maximum and minimum values appear to be stable throughout the experiment run, the populations average fitness tends towards the maximum fitness value.

By the second half of the experiment run, the populations average fitness is virtually indistinguishable from the populations maximum fitness value. Individuals that are incapable of improvement are quickly culled from the population, and the cultural learning mechanism is allowing even genetically mediocre individuals to achieve high fitness levels.

Table I shows values for the average, average maximum, average minimum and standard deviation for both populations. These figures show that the population employing cultural learning is capable of achieving higher averages, maximum, and minimum average values for the whole interval. There is strong evidence (p value < 0.0001, 95% confidence) to suggest that the difference between the two populations is statistically significant.

In order to investigate further the effect of cultural learning on the population, the populations fitness is measured before and after the teaching cycles begin. Thus, the fitness levels of the population are measured before and after teaching to determine an agents innate fitness and its fitness acquired through cultural learning.

Figure 5 shows three fitness values: one for the population employing population learning alone, one for the cultural learning population prior to teaching and the last showing the cultural learning population after teaching is applied.

The population employing cultural learning performs very differently before and after teaching is applied. Prior to teaching, the cultural learning populations fitness is considerably lower than that of the population learning population. Indeed, the populations genotypic fitness (the fitness measured before any cultural influence is applied) is consis-

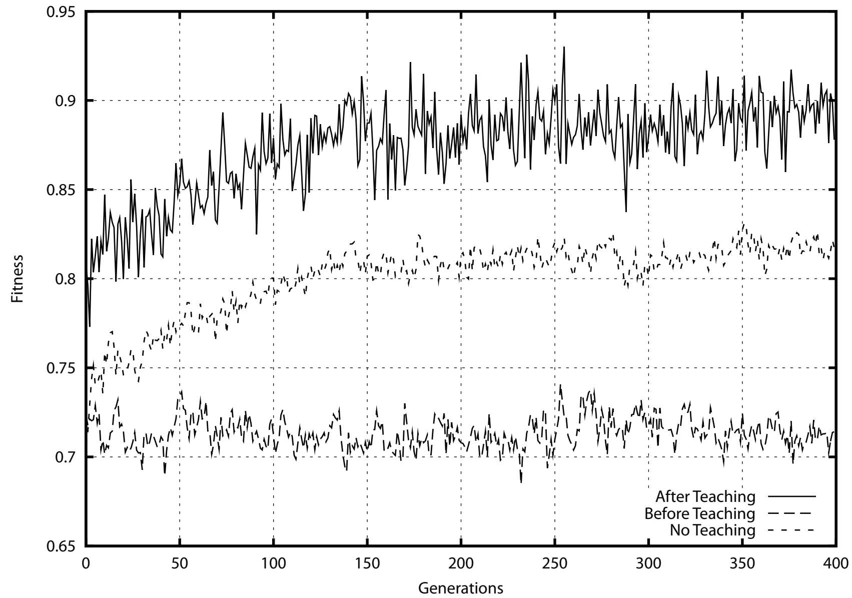

Figure 5. Tic–Tac–Toe Average Fitness for Population Before and After Teaching

tently low and appears to be stable throughout the experiment. This result is discussed further later in the paper.

4.2. The Game of BlackJack

Blackjack or twenty–one begins with the dealer dealing two cards face–up to each player and two to his/herself, with one card visible (the up–card) and the other face down. Cards are valued by their face value (10 for all picture cards) except for the ace which can be counted either as 11 or 1. The object of the game is to obtain a higher score (the sum of all card values) than that of the dealers without exceeding 21.

Each player can draw additional cards until they either stand or exceed 21 and go bust. Once all players have obtained their cards, the dealer turns over the hidden card and draws or stands as appropriate. Should the dealers hand bust, all players win.

The dealer is at considerable advantage because he/she only enters the game once all players have fully completed their play. Thus, it is probable that some players will have bust even before the dealer reveals the hidden card. In addition, the fact that only one of the dealers cards is visible means that players must make judgements based on incomplete information. As a rule, the dealer follows a fixed strategy, typically standing on a score of 17 or more and drawing otherwise.

All aspects related to betting such as doubling down and splitting have been removed from this implementation in order to facilitate comparison with previous work which employs a similar approach.

In a casino setting, between 3 and 6 six full decks of cards are shuffled at the start of the first hand and the game is played until the cards run out. Up to six players and one dealer may play at a blackjack table.

Again for simplicity, this implementation considers only a single player playing against the dealer using a single deck of cards which is shuffled at the start of each hand.

4.2.1. *Bench–marking*

In order to assess the performance of any evolved strategy, a set of bench–marks must be obtained for comparison purposes. While there have been many attempts to calculate the performance of blackjack strategies using simulation and probabilistic techniques(Thorp, 1984; R. R. Baldwin and McDermott, 1956; Thorp, 1963), the values produced tend to vary by a rather large margin. For instance, the success of a player employing the standard dealer strategy is reported at between 39% and 44% wins.

As a result of these discrepancies, it was felt that it may be more meaningful to calculate the values for various strategies using our own simulation. These values will be more readily comparable to the performance of evolved strategies, since a large proportion of the blackjack simulator will also be used by the evolving populations to play games.

The blackjack simulator consists of a dealer, who employs the traditional dealer strategy of standing on 17 or greater, and a single player whose strategy can be set at the beginning of the simulation. As in previous work, both dealer and player hand values are calculated by adding card values where each ace is counted as 11 unless it would cause a bust.

Several strategies were considered:

– Dealer's (Stand on 17 or more, Draw on less)

– Random

– Always stand

– Hoyle's (based on the dealer's up card and the possession of an ace)

```
if (dealer card < 6)
    if (ace is held)
        stand on 15
```

Table II. Blackjack Benchmarking

Strategy	Percentage Wins	Standard Deviation
Hoyle	43.70	1.587
Dealer	41.55	1.576
Uribe et al	38.76	1.505
Always Stand	37.91	1.531
Random	30.41	1.511

```
    else
        stand on 13
else
    stand on 17
```

– Uribe Evolved Strategy (taken from the work of Uribe and Sanchez(Perez-Uribe and Sanchez, 1998))

```
if (score> 9) or [(score> 13)and(score< 19) and
(an ACE is held)]
    stand
else
    hit with 50% probability
```

In order to produce statistically meaningful results, we performed 1000 runs of 1000 games for each strategy. The results presented in Table 4.2.1 are average wins for each strategy. We can see from these results that most strategies perform poorly against the dealer and that Hoyles strategy performs best.

4.2.2. *Experiments*

Each experiment allows 100 agents to evolve over 400 generations. At each generation, agents play 100 games against the dealer strategy and an agents fitness is determined by the percentage of wins obtained scaled to [0.0,1.0]. Crossover was set at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was also added with probability 0.05.

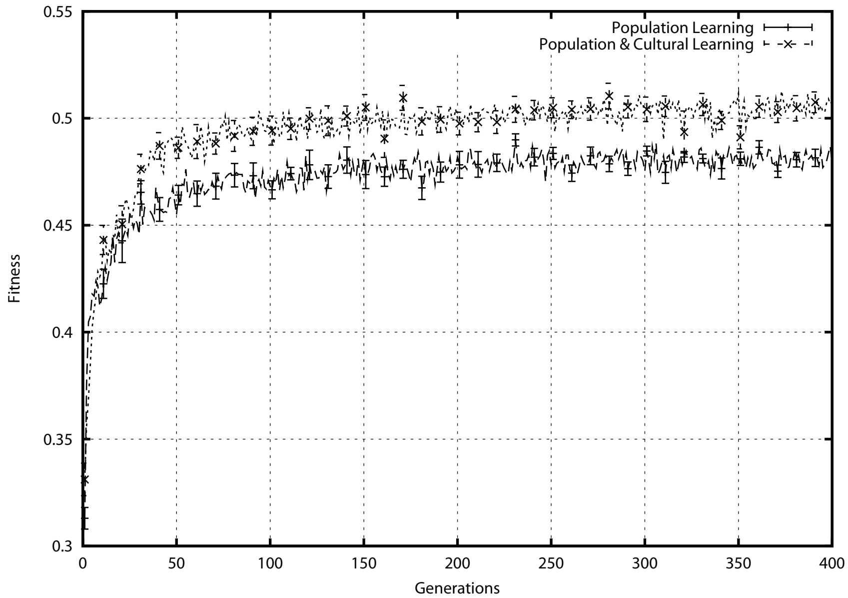

Figure 6. Blackjack Population Fitness

An agents fitness is determined by the number of hands won, normalised to the range [0,1]. This set of experiments does not allow agents to develop card–counting strategies and therefore the fitness values attained must be put into context with the probabilities of success for non–card–counting strategies. It is exceedingly difficult to beat a blackjack dealer consistently more than 50result approaching this figure should be considered optimal.

Figure 6 shows the average error for both populations over the experiment run, along with error bars showing fitness variance within populations. Both populations show similar trends, stabilising within 100 generations to their relative maximum values. The population employing both population and cultural learning is clearly achieving a higher average fitness than the population employing population learning alone.

Unlike the previous experiment, the fitness variance within the populations is similar. This is most likely due to the low probability of any particular individual performing particularly well. However, on average, the worst individuals of the cultural learning population are performing better than the best individuals of the population learning population.

Figures 7 and 8 show the average, maximum and minimum fitness values for both populations. The minimum and maximum fitness values are those of the average best and worst individuals of each generation. Both minimum and maximum values are slightly higher for the popu-

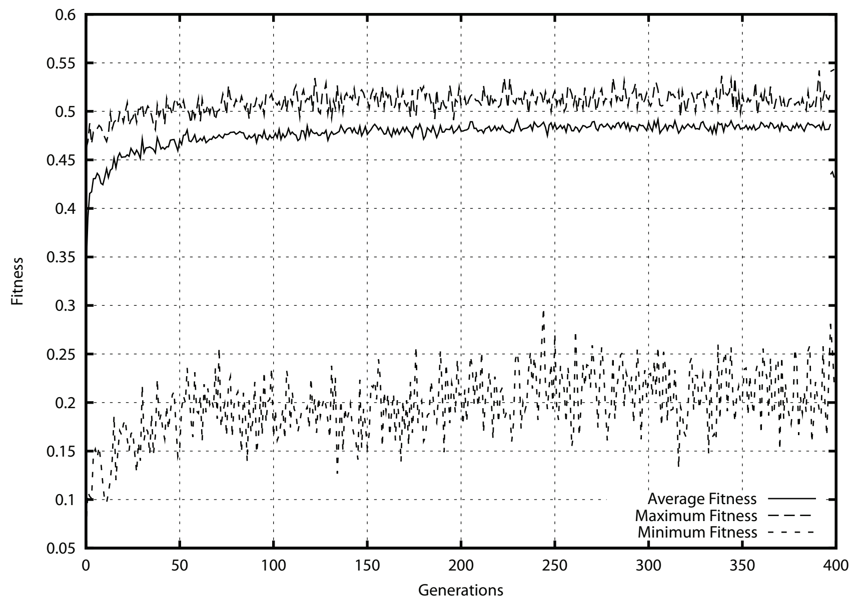

Figure 7. Blackjack Average, Maximum and Minimum Fitness for Non–Teaching Population

Table III. Blackjack Average Fitness

Population	**Avg. Fitness**	**Max Fitness**	**Min Fitness**	**S. D.**
Pop. Learning	0.4726593	0.4898939	0.3129627	0.0002623
Cultural Learning	0.4941667	0.5128222	0.3311286	0.0004356

lation employing cultural learning. The cultural learning mechanism is not only improving the worst individuals in the population, but is also generating novel, high performing individuals.

Table III show the average, average maximum and average minimum fitness values for both populations taken over the entire experiment run. It is clear from these results that the population employing cultural learning is superior in its development of strategies than the population employing population learning alone. There is strong evidence (p value < 0.0001, confidence interval 95%) that the difference in fitness levels of the two populations is statistically significant.

In order to investigate the relative worth of the evolved blackjack strategy compared to the bench–marked strategies described above, the evolved strategy was extracted from the cultural learning population.

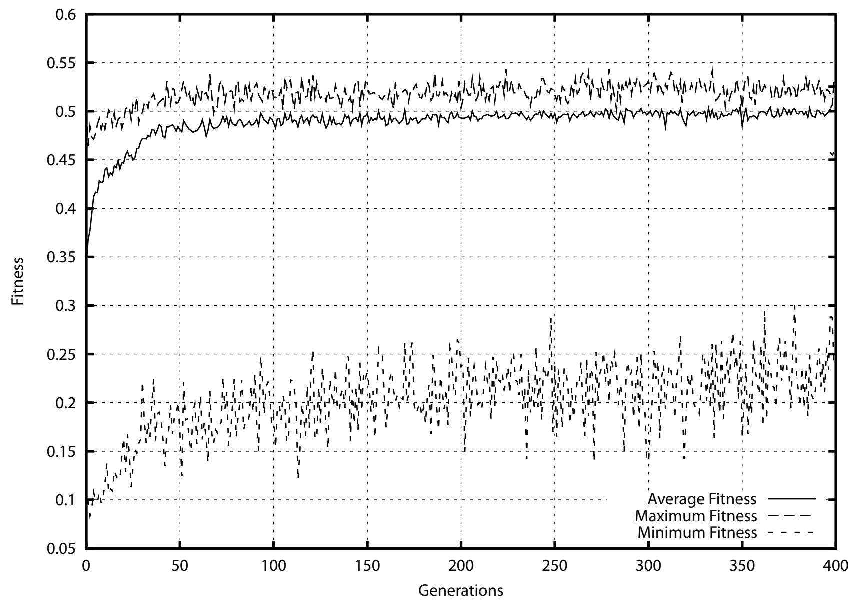

Figure 8. Blackjack Average, Maximum and Minimum Fitness for Teaching Population

This was done by presenting the population with every possible card combination and examining the collective decision of the population. The following resulting strategy was extracted:

```
if (an Ace is held)
{
    if (dealer has a 6 or higher)
        stand on 16
    else
        stand on 17
}
else
{
    if (dealer has a 7 or higher)
        stand on 17
    else
        stand on 13
}
```

The strategy was hard–coded into the blackjack simulator and 1000 runs of 1000 games were played. The averaged results are displayed in Table IV.

Table IV. Final Blackjack Benchmarking Results

Strategy	**% Wins**	**Standard Deviation**
Hoyle	43.69	1.573
Evolved	**43.67**	**1.582**
Dealer	41.52	1.571
Uribe et al	38.43	1.495
Always Stand	38.00	1.529
Random	30.67	1.507

There is strong evidence (p value <0.001, 95% confidence) to support the claim that the evolved strategy and Hoyles strategy are equivalent in terms of performance, suggesting that the population has evolved an optimum strategy given the information available. It is likely that in order to out–perform Hoyles strategy it is necessary to keep track of cards that have been played during a game.

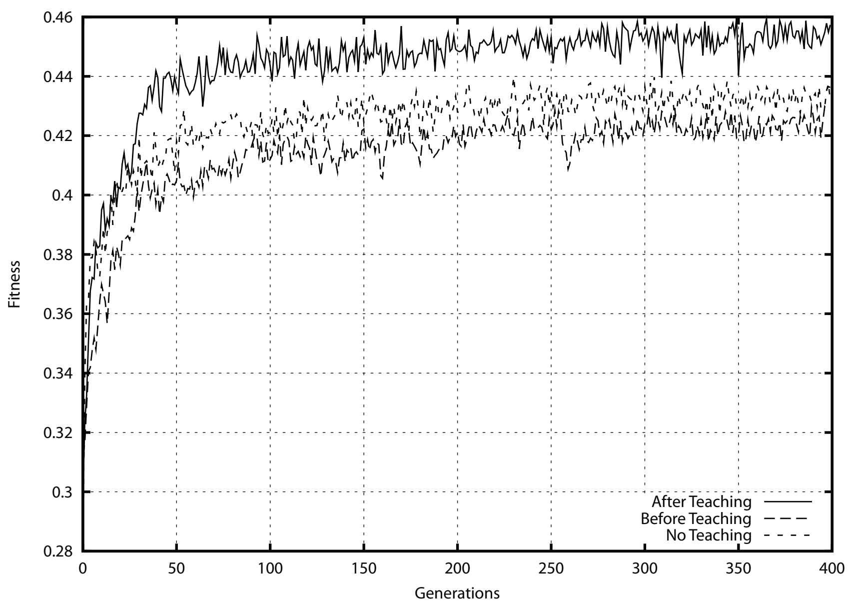

Figure 9. Blackjack Average Fitness for Population Before and After Teaching

As a final investigation into the effect of cultural learning on the population, the cultural learning populations fitness was examined before and after the teaching cycle is applied. The results for this are presented in Figure 9. As in the previous experiment, the application of cultural learning is clearly producing higher fitness levels than population learning alone. However, the cultural learning populations fitness levels prior to teaching are slightly poorer than those of the population learning population.

4.3. Connect Four

The game of connect–four is a two–player game played on a vertical board of 7x6 positions into which pieces are slotted in one of seven available slots. Each player is given a number of coloured pieces (one colour per player) and must attempt to create horizontal, vertical or diagonal piece–lines of length four. Players place one piece per turn into one of the seven slots. The piece then falls onto a free position in the chosen column, creating piles, or towers, of pieces. If a column is full, the player must select an available slot.

Some research has been undertaken in the evolution of connect–four players employing a library of existing games to train the neural networks by back–propagation(Schneider and J, 2002) as well as reinforcement learning methods(Sommerlund, 1996). Our approach allows agents to compete against each other and against a modified minimax player.

Agents play games against a minimax player, whose first move is randomized, allowing agents to play games of some variety. Fitness is assigned according to the length of the game. In other words, agents are rewarded for bringing the game to as close to a draw as possible, as it is very difficult for an agent to beat the modified minimax player.

At each move, the current board position is taken and the agents pieces are added iteratively into each slot. At each iteration, the network is shown the board position through its 84 input nodes. In other words, the network is shown the resulting board position arising from each of it's possible moves. Each board position produces a response from the neural network's output node and the strongest output response is deemed to be the agents preferred board position.

4.3.1. *Experiments*

Populations of 50 agents to evolve over 400 generations. At each generation, agents play games against a the minimax player, once going first and the second time second. Fitness is measured according to how close the game comes to a draw and is scaled to [0,1]. Crossover was set

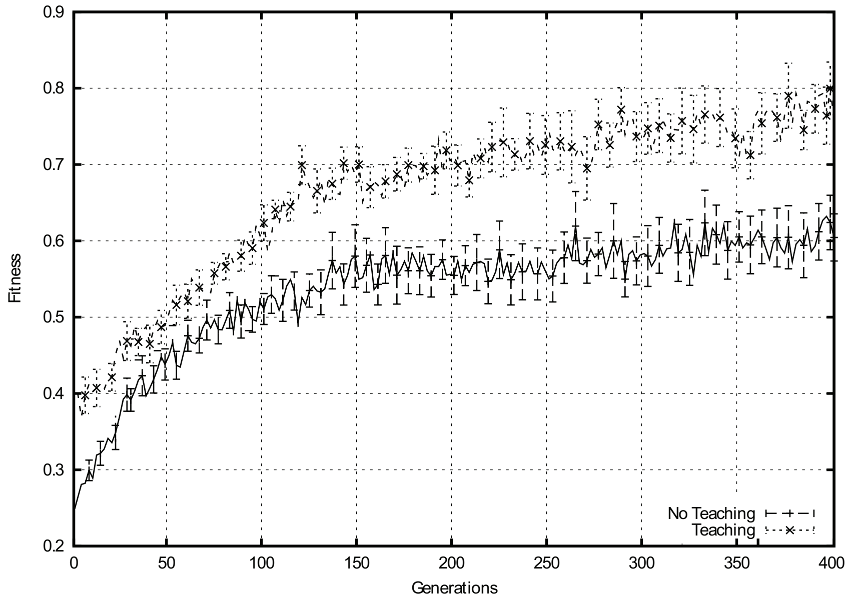

Figure 10. Connect–Four Population Fitness

at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was also added with probability 0.05.

The results illustrated in Figure 10 clearly show the effect of cultural learning on the population. The population employing population learning alone achieves fitness levels of around 0.6, indicating that the population is at least capable of adequately competing against the minimax player in more than half of games played.

However, when cultural learning is applied to the population, the performance improvement is evident. The population achieves fitness levels of close to 0.8, significantly higher than population learning alone, indicating that the population is capable of performing well against a minimax opponent.

Figures 11 and 12 show the average, maximum and minimum fitness values for the two populations. A number of clear distinctions can be observed from these results: firstly, average and maximum fitness values are higher for the population employing cultural learning while minimum values are not altered significantly. This implies that while the cultural learning process is producing high performing individuals, there are still elements in the population that are incapable of successfully interacting with their environment.

Secondly, the average fitness value is higher in the population employing cultural learning, as we have seen from Figure 4.3.1. However,

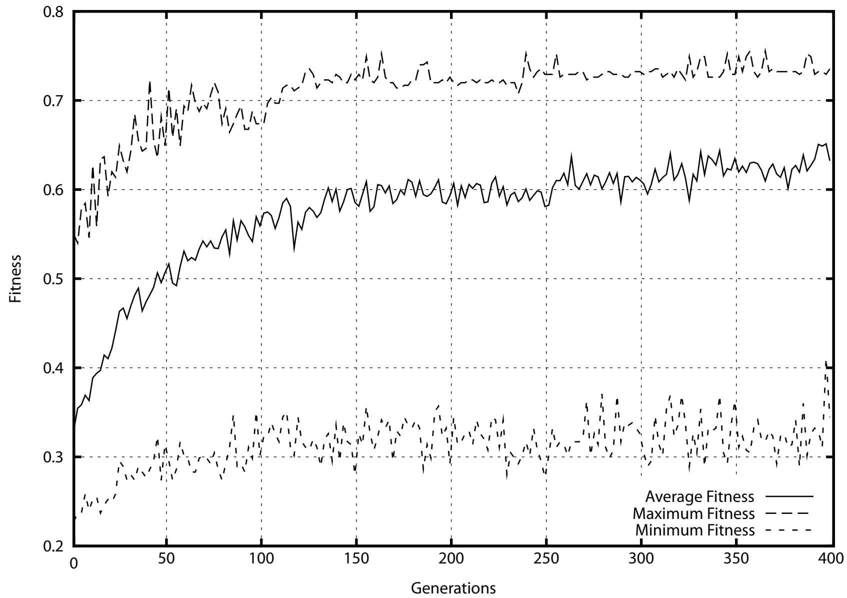

Figure 11. Connect–Four Average, Maximum and Minimum Fitness for Non–Teaching Population

the difference between the average and maximum fitness values is significantly reduced in the population employing cultural learning. Clearly, the cultural learning process is not only generating novel, high performing individuals, but it is also causing the entire population to more closely resemble those individuals that are best adapted to their environment.

Table V shows the average, average maximum and average minimum fitness values for both population taken over the entire experiment run. It is clear from these figures that cultural learning is producing individuals of higher average fitness, but is also capable of producing novel high performing individuals evidenced by the large differences between the maximum values of both populations. Furthermore, there is strong evidence (p value < 0.0001) that the performance differences between the two populations are statistically significant.

Figure 13 shows the fitness values for the population employing population learning alone, the cultural learning population prior to teaching and the cultural learning population after teaching takes place.

Once again, as in previous experiments, cultural learning appears to be selecting individuals for their genetic ability to learn, rather than for their innate ability to solve a particular task. This is illustrated by the fact that the fitness values for the population employing cultural

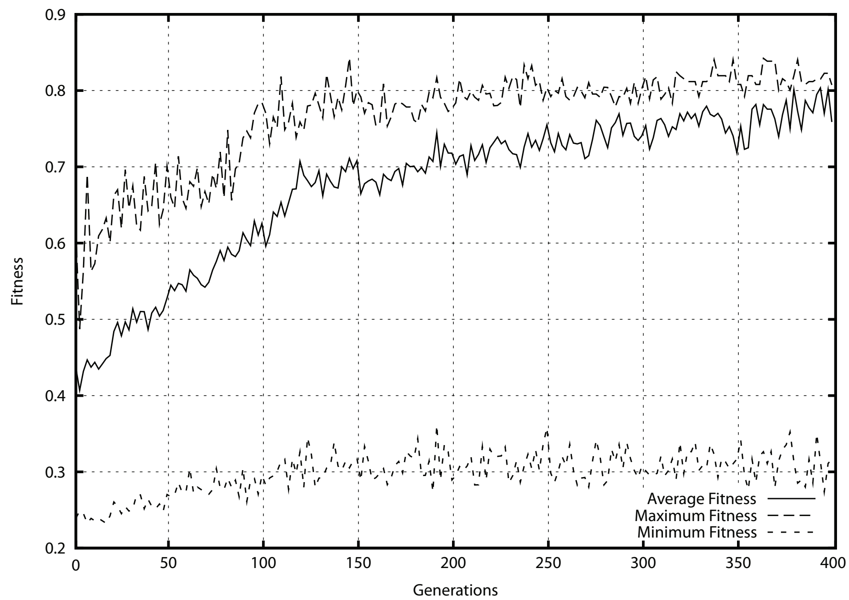

Figure 12. Connect–Four Average, Maximum and Minimum Fitness for Teaching Population

Table V. Connect–Four Average Fitness

Population	**Avg. Fitness**	**Max Fitness**	**Min Fitness**	**S. D.**
Pop. Learning	0.5369646	0.6320001	0.2542307	0.0059148
Cultural Learning	0.6648816	0.7993531	0.3689372	0.0109289

learning are considerably lower prior to teaching than those of the population employing population learning alone.

Significantly, once teaching is applied to the cultural learning population, the fitness level rises and considerably exceeds that of the population employing cultural learning alone. Thus, the cultural learning process is generating individuals with a genetic predisposition toward learning. If teaching is not applied such individuals perform poorly, but as once teaching commences, the innate potential of such individuals is realised in full.

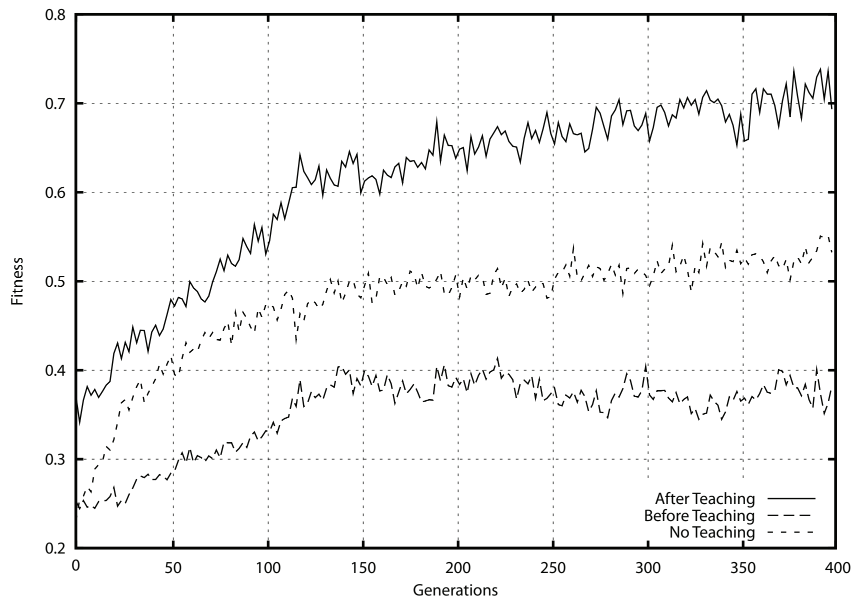

Figure 13. Connect–Four Average Fitness for Population Before and After Teaching

5. Discussion

The results obtained correlate with previous work, showing that the addition of cultural learning is capable of enhancing population fitness. The model of cultural transmission allows individuals to impart information that they themselves have acquired culturally, rather than innate knowledge, leading to some interesting results.

In particular, it is clear that culture is being passed on through generations, as the populations fitness continues to improve despite a significant deterioration in innate fitness. The evolutionary process judges individuals on the basis of their performance after cultural information has been acquired and as a consequence, the genotypic behaviour of individuals becomes less and less important as the cultural exchanges become more successful. Individuals in a cultural learning setting only become competitive once they acquire the populations culture. The innate fitness of such individuals is considerably poorer than that of the population learning population, indicating that most of the knowledge required to survive in the environment is being stored in the culture, not in the genomes.

First generation teachers impart innate knowledge, as they have no teachers to imitate. From then on, pupils acquire knowledge that has itself been acquired by their teachers, cascading back to the first generation.

However, the culture is constantly shaped by the influx and outflux of different teachers and therefore changes in character over time. Such information transmission is much faster than population learning, and allows the cultural learning population to achieve higher fitness levels, despite its genotypic deterioration.

6. Conclusion

This paper presents a model of cultural learning employing vertical transmission of culture in a population of neural network agents presented with a set of sequential decision task problems. Cultural learning gives populations the opportunity to sample acquired information within the population itself. This allows weaker members of the population to gain access to environmental information which would otherwise be impossible to attain without incurring possible fitness losses.

The model allows individuals to culturally impart knowledge they themselves have acquired from previous teachers rather than transmitting only innate knowledge. As in previous work, the results indicate that cultural learning provides improved performance over population learning in each test–case.

However, the results obtained with regard to the populations innate fitness differ somewhat to those previously obtained. While the cultural learning populations fitness continues to improve over time, its innate fitness (the populations fitness prior to acquiring knowledge through teaching) deteriorates significantly. Thus, a large portion of the populations knowledge about its environment is stored in the culture, rather than in its genomes.

Future work will focus on the cultural transmission of knowledge in dynamic environments, investigating whether the increased plasticity of acquired culture (as opposed to genetically acquired knowledge) leads to increased robustness.

References

Baldwin, J.: 1896, 'A new factor in evolution'. *American Naturalist 30* pp. 441–451.

Batali, J.: 1998, 'Computational simulations of the emergence of grammar'. In: S.-K. M. Hurford, J. R. and K. C. (eds.): *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.

Belew, R. K.: 1990, 'Evolution, Learning and Culture: Computational metaphors for adaptive algorithms'. *Complex Systems* **4**, 11–49.

Belew, R. K., J. McInerney, and N. N. Schraudolph: 1992, 'Evolving Networks: Using the Genetic Algorithm with Connectionist Learning'. In: C. G. Langton,

C. Taylor, J. D. Farmer, and S. Rasmussen (eds.): *Artificial Life II*. Redwood City, CA: Addison-Wesley, pp. 511–547.

Best, M. L.: 1999, 'How Culture Can Guide Evolution: An Inquiry into Gene/Meme Enhancement and Opposition'. *Adaptive Behavior* **7**(3/4), 289–306.

Billard, A. and K. Dautenhahn: 1999, 'Experiments in learning by imitation: grounding and use of communication in robotic agents'. *Adaptive Behaviour* **7(3/4)**, 411–434.

Billard, A. and G. Hayes: 1997, 'Learning to Communicate through Imitation in Autonomous robots'. In: *7th International Conference on Artificial Neural Networks*. pp. 763–738.

Borenstein, E. and E. Ruppin: 2003, 'Enhancing Autonomous Agents Evolution with Learning by Imitation'. *Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour* **1**(4), 335–348.

Boyd, R. and P. Richerson: 1985, *Culture and the evolutionary process*. University of Chicago Press.

Brighton, H. and S. Kirby: 2001, 'The Survival of the Smallest: Stability Conditions for the Cultural Evolution of Compositional Language'. In: J. Kelemen and P. Sosk (eds.): *ECAL01*. pp. 592–601, Springer-Verlag.

Cangelosi, A.: 1999, 'Evolution of communication using combination of grounded symbols in populations of neural networks'. In: *Proceedings of IJCNN99 International Joint Conference on Neural Networks (vol. 6)*. Washington, DC, pp. 4365–4368, IEEE Press.

Cangelosi, A. and D. Parisi: 1996, 'The emergence of a language in an evolving population of neural networks'. *Technical Report NSAL–96004, National Research Council, Rome*.

Cavalli-Sforza, L. L.: 1981, *Cultural Transmission and Evolution, A Quantitative Approach*. Princeton University Press.

Cavalli-Sforza, L. L.: 1982, 'Theory and Observation in Cultural Transmission'. *Science* **218**, 19–27.

Chellapilla, K. and D. B. Fogel: 1999, 'Evolving Neural Networks to Play Checkers without Relying on Expert Knowledge'. In: *IEEE Transactions on Neural Networks*, Vol. 10. pp. 1382–1391.

Curran, D. and C. O'Riordan: 2003a, 'Artificial Life Simulation Using Marker Based Encoding'. In: *Proceedings of the 2003 International Conference on Artificial Intelligence (IC-AI 2003)*, Vol. II. Las Vegas, Nevada, USA, pp. 665–668.

Curran, D. and C. O'Riordan: 2003b, 'On the Design of an Artificial Life Simulator'. In: V. Palade, R. J. Howlett, and L. C. Jain (eds.): *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES 2003)*. University of Oxford, United Kingdom, pp. 549–555.

Curran, D. and C. O'Riordan: 2004, 'The Effect of Noise on the Performance of Cultural Evolution'. In: *Proceedings of the 2004 Congress on Evolutionary Computation (CEC 2004)*. Portland, Oregon, USA.

Dautenhahn, K. and C. Nehaniv: 2002, *Imitation in Animals and Artifacts*. MIT Press.

de Boer, B. and P. Vogt: 1999, 'Emergence of Speech Sounds in Changing Population.'. In: D. Floreano, J.-D. Nicoud, and F. Mondada (eds.): *ECAL*, Vol. 1674 of *Lecture Notes in Computer Science*. pp. 664–673, Springer.

De Jong, E. D.: 1999, 'Analyzing the Evolution of Communication from a Dynamical Systems Perspective'. In: *Proceedings of the European Conference on Artificial Life ECAL'99*. pp. 689–693.

Demiris, J. and G. Hayes: 1996, 'Imitative learning mechanisms in robots and humans'. In: *Proceedings of the 5th European Workshop on Learning Robots*. pp. 9–16.

Denaro, D. and D. Parisi: 1996a, 'Cultural evolution in a population of neural networks'. In: *Proceedings of the 8th Italian Workshop on Neural Nets*. pp. 100–111.

Denaro, D. and D. Parisi: 1996b, 'Cultural evolution in a population of neural networks'. In: *M.Marinaro and R.Tagliaferri (eds), Neural Nets Wirn-96.New York: Springer*. pp. 100–111.

Floreano, D. and F. Mondada: 1989, 'Evolution of plastic neurocontrollers for situated agents'. In: *Animals to Animats*, Vol. 4. MIT Press.

Gruau, F.: 1994, 'Neural Network Synthesis using Cellular Encoding and the Genetic Algorithm.'. Ph.D. thesis, Centre d'etude nucleaire de Grenoble, Ecole Normale Superieure de Lyon, France.

Hayes, G. and J. Demiris: 1994, 'A robotic controller using learning by imitation'. In: *Proceedings of the 2nd International Symposium on Intelligent Robotic Systems*.

Hinton, G. E. and S. J. Nowlan: 1987, 'How learning guides evolution'. In: *Complex Systems*, Vol. I. pp. 495–502.

Hutchins, E. and B. Hazlehurst: 1991, 'Learning in the Cultural Process'. In: *Artificial Life II, ed. C. Langton et al.* pp. 689–706, MIT Press.

Hutchins, E. and B. Hazlehurst: 1995, 'How to Invent a Lexicon: The Development of Shared Symbols in Interaction'. In: N. Gilbert and R. Conte (eds.): *Artificial Societies: The Computer Simulation of Social Life*. UCL Press: London, pp. 157–189.

Kirby, S. and J. Hurford: 1997a, 'Learning, culture and evolution in the origin of linguistic constraints'. In: P. Husbands and I. Harvey (eds.): *ECAL97*. pp. 493–502, MIT Press.

Kirby, S. and J. Hurford: 1997b, 'Learning, culture and evolution in the origin of linguistic constraints'. In: *4th European Conference on Artificial Life*. pp. 493–502, MIT Press, Cambridge, MA.

Kitano, H.: 1990, 'Designing Neural Networks using Genetic Algorithm with Graph Generation System'. In: *Complex Systems, 4, 461-476*.

Kolen, J. F. and J. B. Pollack: 1991, 'Back Propagation is Sensitive to Initial Conditions'. In: R. P. Lippmann, J. E. Moody, and D. S. Touretzky (eds.): *Advances in Neural Information Processing Systems*, Vol. 3. pp. 860–867, Morgan Kaufmann Publishers, Inc.

Koza, J. R. and J. P. Rice: 1991, 'Genetic Generation of Both the Weights and Architecture for a Neural Network'. In: *International Joint Conference on Neural Networks, IJCNN-91*, Vol. II. Washington State Convention and Trade Center, Seattle, WA, USA, pp. 397–404, IEEE Computer Society Press.

MacLennan, B. and G. Burghardt: 1993, 'Synthetic Ethology and the Evolution of Cooperative Communication'. In: *Adaptive Behavior 2(2)*. pp. 161–188.

Mandischer, M.: 1993, 'Representation and Evolution of Neural Networks'. In: R. F. Albrecht, C. R. Reeves, and N. C. Steele (eds.): *Artificial Neural Nets and Genetic Algorithms Proceedings of the International Conference at Innsbruck, Austria*. Wien and New York: Springer, pp. 643–649.

Miller, G. F., P. M. Todd, and S. U. Hedge: 1989, 'Designing Neural Networks Using Genetic Algorithms'. In: *Proceedings of the Third International Conference on Genetic Algorithms and Their Applications*. pp. 379–384.

Moriarty, D. E. and R. Miikkulainen: 1995, 'Discovering Complex Othello Strategies Through Evolutionary Neural Networks'. *Connection Science* **7**(3–4), 195–209.

Nolfi, S. and D. Parisi: 1995, 'Learning to adapt to changing environments in evolving neural networks'. In: *Technical Report 95-15, Institute of Psychology, National Research Council, Rome, Italy.*

Oliphant, M. and J. Batali: 1997, 'Learning and the emergence of coordinated communication'. *The newsletter of the Center for Research in Language* **11**(1).

Pereira, F. B. and E. Costa: 2001, 'How learning improves the performance of evolutionary agents: A case study with an information retrieval system for a distributed environment.'. In: *Proceedings of the International Symposium on Adaptive Systems: Evolutionary Computation and Probabilistic Graphical Models (ISAS 2001)*. pp. 19–23.

Perez-Uribe, A. and E. Sanchez: 1998, 'Blackjack as a Test Bed for Learning Strategies in Neural Networks'. In: *International Joint Conference on Neural Networks, IJCNN'98*. pp. 2022–2027.

Pujol, J. C. F. and R. Poli: 1998, 'Efficient Evolution of Asymmetric Recurrent Neural Networks Using a Two-dimensional Representation'. In: *Proceedings of the First European Workshop on Genetic Programming (EUROGP),*. pp. 130–141.

R. R. Baldwin, W. E. Cantey, H. M. and J. P. McDermott: 1956, 'The Optimum Strategy in Blackjack'. In: *Journal of the American Statistical Association*.

Richards, N., D. Moriarty, P. McQuesten, and R. Miikkulainen: 1997, 'Evolving Neural Networks to Play Go'. In: *Proceedings of the 7th International Conference on Genetic Algorithms*. East Lansing, MI.

S. Nolfi, J. L. Elman, D. P.: 1994, 'Learning and Evolution in Neural Networks'. In: *Adaptive Behavior*, Vol. 3. pp. 5–28.

Sasaki, T. and M. Tokoro: 1997, 'Adaptation toward Changing Environments: Why Darwinian in Nature?'. In: P. Husbands and I. Harvey (eds.): *Fourth European Conference on Artificial Life*. Cambridge, MA, pp. 145–153, MIT Press.

Saunders, G. M. and J. B. Pollack: 1994, 'The Evolution of Communication in Adaptive Agents'. Technical report, Department of Computer and Information Science, The Ohio State University.

Schneider, M. O. and L. G. R. J: 2002, 'Neural connect four, aconnectionist approach to the game'. In: *Proceedings of the VIIth Brazilian Symposium on Neural Networks (SBRN'02)*. pp. 236–241, IEEE Computer Press.

Sommerlund, P.: 1996, 'Artificial Neural Networks Applied to Strategic Games'. Unpublished Manuscript, URL: http://www.diku.dk/student/peso/960513.ps.Z. (Last accessed 20/05/2005).

Spector, L.: 1994, 'Genetic programming and AI planning systems'. In: *Proceedings of Twelfth National Conference on Artificial Intelligence*. Seattle, Washington, USA, pp. 1329–1334, AAAI Press/MIT Press.

Spector, L. and S. Luke: 1996, 'Culture Enhances the Evolvability of Cognition'. In: *Cognitive Science (CogSci) 1996 Conference Proceedings*.

Sutton, R. S.: 1986, 'Two problems with backpropagation and other steepest-descent learning procedures for networks'. In: *Proc. of 8th Annual Conf. of the Cognitive Science Society*. pp. 823–831.

Thorp, E. O.: 1963, *Beat the Dealer. A winning strategy for the game of twenty one*. Vintage Books, NY.

Thorp, E. O.: 1984, *The Mathematics of Gambling*. Lyle Stuart.

Watson, J. and J. Wiles: 2002, 'The rise and fall of learning: A neural network model of the genetic assimilation of acquired traits'. In: *Proceedings of the 2002 Congress on Evolutionary Computation (CEC 2002)*. pp. 600–605.

Weaver, L. and T. Bossomaier: 1996, 'Evolution of Neural Networks to Play the Game of Dots-and-Boxes'. In: *Artificial Life V: Poster Presentations*. pp. 43–50.

White, D. W.: 1994, 'GANNet: A genetic algorithm for searching topology and weight spaces in neural network design'. Ph.D. thesis, University of Maryland College Park.

Yanco, H. and L. Stein: 1993, 'An Adaptive Communication Protocol for Cooperating Mobile Robots'. In: *From Animals to Animats 2. Proceedings of the second International Conference on Simulation of Adaptive Behavior*. pp. 478–485, MIT Press, Cambridge Ma.