The Impact of Elite Fraction and Population Size on Evolved Iterated Prisoner’s Dilemma Agents

Daniel Ashlock and Eun-Youn Kim

Abstract—The iterated prisoner’s dilemma is a simultaneous two-player game widely used in studies on cooperation and conflict. Past work has shown that the choice of representation or available resources such as the number of states or neurons of evolving agents has a large impact on the behavior of evolved agents. This study revisits three qualities of the agent training algorithm for finite state agents to examine their impact on agent behavior: population size, elite fraction, and the number of states the agent is permitted. All three of these algorithm parameters are shown to have an impact on the character of evolved agents. Assessment of agent behavior is performed using three tools. The first is play profiles which bin the ranges of score space. The total score assessment, a global characterization of the type of play that occurs over the course of evolution, is the second assessment used. For the third assessment, an analysis of the ability of agents with different numbers of states to compete with one another is performed. High elite fractions in the training algorithm are found to encourage cooperation. Larger populations increase cooperation for populations of agents with small and intermediate numbers of states but have little effect for agents with large numbers of states. As in past studies with fixed population size and elite fraction, agents with large numbers of states are found to have more diverse and less cooperative behavior. Having more states is also found to grant a competitive advantage.

I. INTRODUCTION

The prisoner’s dilemma [17], [16] is a classic model in game theory. Two agents each decide, without communication, whether to cooperate (C) or defect (D). The agents receive individual payoffs depending on the actions taken. The payoff for mutual cooperation C is the cooperation payoff. The payoff for mutual defection D is the defection payoff. The two asymmetric action payoffs S and T, are the sucker and temptation payoffs, respectively. In order for a two-player simultaneous game to be considered prisoner’s dilemma, it must obey the following four inequalities:

\[ S \leq D \leq C \leq T \]

and

\[ (S + T) \leq 2C. \]

In the iterated prisoner’s dilemma (IPD) the agents play many rounds of the prisoner’s dilemma. IPD is widely used to model emergent cooperative behaviors in populations of selfishly acting agents and is often used to model systems in biology [30], sociology [21], psychology [28], and economics [20]. Many researchers have investigated the evolution of prisoner’s dilemma playing agents [26], [29], [19], [22], [18], [27] with a focus on understanding the evolution of cooperation, particularly in changing environments.

In many studies, the choice of agent representation is made without much concern and sensitivity to the choice of agent representation is ignored. Initial work on this issue appears in [9] which demonstrated that the choice of agent representation has a dominant effect on the probability agents will learn to cooperate. This work was extended to include an additional representation in [1]. A theoretical treatment of why some representations are more likely to yield cooperative behavior appears in [13]. This line of research was extended by examining the impact of simply varying the resources allocated to agents within a representation [7]. Changing the number of states in a finite state representation, the number of probability levels available to a Markov chain, the number of neurons in an artificial neural net, and the time depth of a lookup table were all found to change the behavior of agents substantially.

Taken together, the studies in this research program demonstrate that our understanding of the interaction of the way agents are encoded and the process of training those agents with evolution is so incomplete as to make research that uses evolved game-playing agents unreliable. The goal of using IPD-playing agents as models for cooperation and conflict both for humans and animal communities requires that we complete the characterization of evolutionary training of agents for playing the IPD. Comments from reviewers on a paper currently under revision motivated the current study were we will vary parameters of the training algorithm. We will perform this examination using the direct finite state agent representation because it is the most studied representation and we lack space to thoroughly examine more than one representation.

We retain the agent-training algorithm from earlier studies but modify parameters held constant in other studies. The standard algorithm operates on a fixed size population with thirty or thirty-six agents. One-third of the agents are replaced in each population updating. This is called having an elite fraction of $2/3$ - the elite are those agents not replaced. These factors were held constant while others, such as agent representation or resources made available to an agent were changed. In this study we permit population size and elite fraction to vary, examining nine new pairs of values for these factors. This study uses a finite state representation for
agents, a representation that has exhibited diverse behavior in the past [2]. The experiments look at agents with a small, standard, and large number of states. Each of these three factors is examined to see what impact it has on agent play behaviors.

An earlier study [15] has examined larger population sizes, but its primary focus was the characterization of new types of strategies that emerge after many generations of evolution and it did not juxtapose population size with other factors. The study in [14] examined large population sizes, but with a substantially different agent training algorithm.

Earlier studies that examined representation and agent resources painted in relatively broad stokes. While this study retains variation of a resource as part of its experimental design, the number of states available to an agent, its primary purpose is to narrow the brush strokes to the algorithm parameters elite size and population size in a manner that can be compared with other studies in this series. As we will see, even changing the simple, basic algorithm parameters changes the type of agents that arise under evolution. This study may complete the identification of factors other than the problem modelled that affect the outcome of experiments using evolved IPD agents. A present published studies are likely to fail to state important experimental design parameters because those factors were not yet clearly identified.

The remainder of this study is structured as follows. Section II gives the design of experiments including representations, resource types, and evolutionary algorithm design. Section III presents and discusses the results, comparing agent populations evolved with different resource levels and algorithm parameters. Section IV draws conclusions and outlines next steps.

II. DESIGN OF EXPERIMENTS

The finite state machines used in this study use the Mealy architecture with responses encoded on the transitions (the Moore architecture encodes responses on the states). An example of a Mealy machine appear in Figure 1. Transitions in the finite state machine used in this study are driven by the opponent’s last action and so an initial action is also required. Machines are stored as a string of states with a state incorporating both transitions and responses associated with the state. The initial action is stored with the first state. Crossover operates on the string of states; this study uses two point crossover. The mutation operator operates by changing the initial state or action 5% of the time (each), the destination of a state transition 40% of the time, and a response to an opponent’s action 40% of the time. This mutation operator is retained from previous studies for consistency.

We examine 27 sets of algorithm parameters. These experiments form a full factorial over the values population size \{72, 108, 144\}, elite fraction \{1/3, 1/2, 2/3\}, and numbers of states \{4, 16, 128\}. For each set of parameters, 100 sets of evolutionary runs are performed for 3200 generations. In each generation the members of the population play a round-robin tournament and the average fitness against all opponents is computed. This fitness is used for fitness-proportional selection of pairs of parents from the elite to replace the non-elite portion of the population. Pairs of parents are copied, the copies undergo two-point crossover of their string of states, and then the copies are mutated to generate the children which are placed into the population.

This design of experiments, together with the analysis techniques described subsequently, will permit us to assess the way that the three factors, population size, elite fraction, and number of states, interact to affect the types of agents that arise.

A. Analysis Techniques

We retain two analysis techniques from [7], play profiles and the total score statistic. The ability of agents trained using different algorithm to compete in one-on-one contests is also assessed. Sampled over a large number of pairs of agents, the Bernoulli variable “who won” can be used to estimate the probability of one agent type over another.

1) Play Profiles: On of the primary assessments of an evolutionary system for training prisoner’s dilemma agents is the probability that a given population is cooperative. The past studies [12], [11], [3], [6], [8], it was established that when 150 rounds of iterated prisoner’s dilemma with the

![Fig. 1. An example 4-state finite state machine for playing IPD. Transitions are labeled as input/output and the sourceless arrow in the upper right is labeled with the machine’s initial action.](image)

![Fig. 2. An example play profile for 3200 generations of evolution, partitioning the possible population average scores into ten bins. Colors are explained in the text.](image)
0, 1, 3, 5 payoff scheme are used in fitness evaluation that an average score of 2.8 signifies that FSMs are in a cycle of sustained cooperation. In earlier studies this was the working definition of cooperation.

Play profiles extend this fitness measure to binning all possible scores. The average fitness of a population is a sum of pairs of payoffs with one of three values: (1,1), (0,5), or (3,3). This places the mean value in the range $1 \leq \mu \leq 3$. This region is divided into ten equal intervals, the top one corresponding to the definition of cooperation given above, $2.8 \leq \mu \leq 3.0$. The scores are binned in each of seven epochs: 50, 100, 200, 400, 800, 1600, and 3200 generations. The number of populations with population average scores in each of the ten intervals is counted. The resulting $10 \times 7$ table is the play profile for an experiment. An example play profile appears in Figure 2. Width of a bar is proportional to the number of populations in the corresponding bin.

The colors used in some of the bins are adapted from a coloring scheme used in earlier studies. The green shade is used in the box that contains the score for mutual defection; red is used in the box that includes the score for playing randomly; blue appears in the being containing mutual cooperation. Other bars are colored so that lighter shades of grey correspond to higher levels of cooperation. These colors aid in interpreting play profiles.

2) Total Score: For each experiment, the average fitness in each generation of each replicate was saved in each generation. Figure 3 shows an example of such a curve and the area underneath that curve. This area is a good surrogate for the degree of cooperativeness of the population of agents. This measure incorporates information from all generations. We perform the natural affine normalization to make the minimum area zero (by treating a average score of one as minimum area zero of states, given that the agents did not tie, was computed using the normal approximation to the binomial. If $v$ is the number of agents with more states that won and $n$ is the number of samples this confidence interval is given by:

$$\frac{v}{n} \pm 2.12 \times \sqrt{\frac{v/n \cdot (1-v/n)}{n}} \quad (3)$$

Two sets of agent trained in different ways are presumed to have significantly different competitive ability if the confidence interval does not include $p = 0.5$.

Since there are contests where agents compete to play the IPD [23], [17], the algorithm parameters that yield more effective competitors are of interest beyond the application in this study of showing that different algorithm parameters yield different agent behaviors.

III. RESULTS AND DISCUSSION

The play profiles for all experiments are shown in Figure 4 while the total fitness statistics appear in Figure 5. The assessment of competitive ability is shown in Figure 6. None of the three parameters varied display a consistent trend for any of the assessments, demonstrating a clear non-linearity in the system.

In addition, all three assessments demonstrate that there are statistically significant differences between at least some pairs of agents trained in different ways. This gives an affirmative answer to the hypothesis that experiments using evolved IPD agents must control for population size and elite fraction - and also the number of a available states, though this factor has been investigated in earlier studies.

A. Impact of Elite Fraction

So far as we know, the impact of the elite fraction in the context of training IPD-playing agents has not been previously studied. Since replacing a lower fraction of agents reduces the rate of change of agent behaviors the expectation is that higher elite fractions would yield more cooperative agents. Figure 4 supports this hypothesis for the 4 and 16 state agents, but not for the 128 state agents. For the smallest population size, 72, the populations with the lowest elite fraction become more cooperative.

This study confirms results in earlier studies with population size 36 that agent populations of 128-state FSMs become cooperative more slowly than agents with fewer states. If we view low elite fractions as increasing the rate of evolution - plausible as a low elite size increases the rate of introduction of new types - then these results are consistent with the following explanation. The lower elite size increases the rate of evolution, arriving at a more cooperative state sooner.
The phenomenon of higher elite fractions encouraging cooperation are most pronounced for the populations of 4-state agents. Its relative subtlety for the 16-state agents suggests that the two effects, encouraging cooperation by having a more predictable agent population and speeding the emergence of cooperation as observed clearly in the 128-state agents, are in a state of near balance for the sixteen state agents.

The total fitness statistics also show a more pronounced increase in cooperation for the 4-state agents than for the 16-state agents; the sixteen state agents have better total fitness statistics overall, perhaps masking the improvement. To total fitness statistics for the 128-state machines show lower values, but also less of a trend associated with elite size, supporting the results from the play profiles.

Examining Figure 6, we see that, while agents with less states are uniformly out-competed, the margin of victory for the machines with more states is smaller when the agents with fewer states have a lower elite fraction that the agents with more states. If evolution tends to grant more competitive ability, something that was found to be true in [10], then this observation also supports the hypothesis that smaller elite fractions permit more evolution.

**B. Impact of Population Size**

For the 4- and 16-state agents, larger populations clearly encourage cooperation. This is visible in both the play profiles shown in Figure 4 and the total fitness statistic in Figure 5. These figures also provide support for this trend in the elite-fraction 1/3 populations of 128-state agents, but they exhibit a weaker increase in cooperation. The two higher elite fraction show no such trend for the 128-state agents.

The major impact of having 128 states available to the agents is to astronomically increase the number of available strategies. In particular this increases the potential for **hand-shaking** strategies [15]. These strategies engage is brief, low-fitness play as a means of performing kin-recognition.

These results are in contrast to a study that used a large, geographically structured population [14]. In this study large populations rapidly evolved to a highly cooperative state. This study differed both in using a geographically structured scheme for agent breeding and a sampled, as opposed to round-robin fitness evaluation scheme. This suggests that both factors, geographic structure and sampled fitness evaluation merit additional investigation.

Population size had a modest impact on competitive ability. When there is an observable effect, the winning agent population does better when the loosing agent population is larger. Since larger populations evolve more slowly when learning IPD, this is more support for the hypothesis that more evolution is better for competitive ability. The jump-start one might expect from have a more diverse initial population is apparently absent - but this is not too surprising as game playing agents typically have an early bottleneck in which most agent types die out.

**C. Impact of Available States**

Of the three factors examined in this study, the number of states available to agents is the one for which the most work has been done recently after noticing the anomalous results of changing the number of states in [2]. The strongest result in this study is that changing the number of states modifies the impact of both the other factors varied. Having a large number of states, as noted above, reveses the impact of the elite fraction.

All three numbers of states examined exhibited different behaviors. In past studies, 16 states was used as a default value, dating back to very early studies such as [25]. This study detected that 16 states is within a balance point where the impact of elite size and available states more-or-less canel out one anothers impact. This is ironic and re-enforces the notion that there is a need to more-fully characterize the behavior of systems of evolving game-playing agents before using these tools with confidence as models of cooperation and conflict.

The competitiveness results demonstrated that agents with more states have a clear competitive advantage. In 81 comparisons between 4-state and 128-state agents, the lowest probability of victory by 128-state was \(0.7725 \pm 0.0444\), for 2/3 elite and population size of 72 for the 128-state agents and 1/3 elite with a population size of 72 for the 4-state agents. Figure 6 shows that there were no pairs of experiments where the confidence interval on probability of victory by agents with more states contained \(p = 0.5\).
Fig. 4. Shown are play profiles for all twenty-seven experiments performed.
This study clearly demonstrated an interaction between elite fraction, population size, and number of states made available. Two impacts of elite-fraction were detected that pull in opposite directions.

- When agents are small enough (few states) to evolve efficiently, higher elite fractions encourage cooperation by making the rate at which the population changes slower.
- For large agent sizes, smaller elite fractions permit more rapid evolution of agents. These larger agents were found, in previous studies, to evolve to a cooperative state more slowly. This, in turn, caused lower elite fractions to encourage cooperation.

The study also showed that for agents with few states, larger population sizes encourage cooperation; this effect was not visible for agents with larger numbers of states. The study also confirmed the results of earlier studies on the relatively slow emergence of cooperation in agents with more states.

One clear outcome of the study is that elite fraction, or its equivalent in other sorts of training algorithms, as well as population size, need to be added to the list of parameters that influence the outcome of training game playing agents with an evolutionary algorithm.

A. A Protocol for Reporting Evolution of Game Playing Agents

This study is one in a fairly large collection that examine uncontrolled sources of variation in the evolution of game playing agents. An early priority for future research is to compile a careful list of factors that matter and draft a proposed set of reporting criteria for this sort of agent based study. Such a proposal must be a draft and would require input from a broad variety of researchers to become acceptable for adoption by the research community.

B. Evolutionary Time and Lineage

Another interesting effect found in this study is the impact of both number of states and time spent evolving on an agents competitive ability. The quantity “time spent evolving” needs to be examined with greater care. The number of generations a population is permitted to evolve is a coarse measure of the amount of evolution that has taken place. A better measure might be average lineage length.

**Definition 1:** A **lineage** is a chain of descent with nodes that are members of an evolving population and links that join a parent and child. The **lineage length** of an individual is the longest lineage for which that individual is the endpoint. Notice that a lineage may be part of a larger lineage - a population member that appears in the second generation is the end of a lineage of length one, even if it goes on to have its own children with longer lineages that descend through it.

We propose lineage length as an assessment for future work because it seems a more accurate means of measuring the amount of evolution that takes place. If, for example, an elite population remained the same for many generations, producing inferior children, then it would have an average lineage length shorter than its time of evolution as measured in generations. Comparing lineage length with level of cooperation and competitive ability may yield cleaner results than measuring evolutionary time in generations.

Note that the lineage length of an individual is simply one plus the larger of its parent’s lineage lengths; this means that the entire pattern of descent need not be saved making this proposed assessment relatively inexpensive in both space and...
computing power.

C. Deep Time Studies

In [15] it was found that new strategies, not seen before, arose in the gap between generations 32768 and 65536 (these were the last to sampling epochs) in populations of 20 state finite state agents playing the iterated prisoner’s dilemma. Given the apparently slower rate of adaption in agents with 128 states, this suggests that even deeper time studies may yield new, complex strategies in agents with large numbers of states.

D. Other Representations and Algorithms

This study demonstrates that directly encoded finite state agents exhibit a sensitivity to the choice of population size and elite fraction. As noted in the introduction there are many other representations that can be used to evolve agents to play IPD. There are also many games beside IPD that it is potentially interesting to train agents for. This suggests a vast cone of future research projects checking other representations and games.

This study has also stuck with a particular agent training algorithm used in several other studies in the name of...
controlling sources of variation. The use of a round robin tournament to assess fitness, the use of an elite for reproduction, and the absence of a hall of fame or other source of long-term genetic memory other than direct inheritance are all features that still have not been investigated.

E. Geography

Two earlier studies [10], [14] examined the evolution of finite state agents for the IPD using very different geographic structures. In both studies the diversity of agent types were found to have been enhanced. In the latter study, it was found that the particular type of geographic structure with local breeding but global fitness evaluation, encourages a robust form of cooperation. The agents not only cooperated with one another, but exhibited an enhanced resistance to invasion. These studies suggest that it may be worth structuring an investigation around the following questions:

- What is the impact of having and varying a geographic population structure?
- What is the impact of round robin, geographically local, or geographically random choice of opponents for fitness evaluation?
- Do disturbances, disruptions that remove players from parts of the geography, have an impact on the types of agents that evolve?

This latter question dovetails with ecological investigations on the importance of disruptions to ecology and biodiversity [31], [24].

References


