# Leveraging Eco-Evolutionary Dynamics to Evolve Backgammon Agents

Project Proposal for CSE 841

Austin J. Ferguson — 10/21/2019

**Motivation**

It is impossible to ignore the progress in artificial intelligence and machine learning over the past several years. Both fields are now ubiquitous in academia, industry, and even the real world. However, focusing on the popular area of reinforcement learning, a central question still persists: *how can we best reward AIs so that they improve at solving their specified task?* Here we focus on the complex task of game playing, which often requires the use of multiple strategies throughout a single match to successfully defeat an opponent. In these complex task domains, what constitutes a "good" move, and how and when do we reward AI for these "good" moves?

These questions lie at the core of reinforcement learning, and many different approaches have arisen over the years. For instance, one can simply reward an AI only after it wins a game (*e.g.*, [Tesauro, 1994]). This method can create too distant of a target, where AIs struggle to reach the minimum threshold of winning *any* games. AIs can also be rewarded for intermediate steps that we, as the designer, think are beneficial. This approach is also imperfect, as we have no guarantee that these intermediate steps are beneficial to winning the game (*e.g.*, rewarding resource collection could steer AIs away from other potential strategies). These methods both have benefits and drawbacks, and which to use depends on the problem at hand.

In this work I propose using an evolutionary algorithm (EA) that leverages ecological dynamics found in nature to evolve complex AI agents. While ecology is mostly used in the EA community to prevent premature convergence at sub-optimal solutions, the artificial life community has shown that ecology can promote not only diversity, but also novelty and complexity [Taylor et al., 2016]. By evolving solutions this way we expect the evolution of building blocks that assist in the evolution of more successful AIs. If this is true, this method can be used beyond games to evolve a large variety of complex AI tasks with a wide range of real-world uses.

**High-Level Overview**

To evolve successful game playing AIs, we focus on two actions that will be rewarded: prediction and successful play. Like in [Tesauro, 1994], we will reward AIs that are capable of winning games. To help AIs reach that point, we also reward AIs that successfully *predict* pre-recorded human moves, which is an example of imitation learning / behavior cloning [Sammut, 2010]. Given enough data, AIs capable of predicting a human's moves are expected to be adept at one or more strategies for the game. An AI capable of predicting what a human would do can then use these predictions to play games on its own. However, at best a given AI can only be expected to mimic the strategies it has trained on. Many games do not have a single dominant strategy, and therefore skilled opponents will still be able to defeat AIs trained on a small subset of strategies. This sets the stage for the rewards given for winning games.

In ecology, diversity comes when organisms fill different niches in the environment. By rewarding predictions we are creating niches for the different strategies demonstrated by the humans. With this setup, AIs spanning multiple niches (and therefore multiple strategies) are capable of coexisting at one time. It is likely the case that these more general strategies will appear only after AIs first discover the basic building blocks necessary to employ a strategy. While an AI might initially specialize in a particular strategy, the building blocks it learns may be useful in learning additional strategies. Then the rewards given to AIs for winning games will apply selective pressure for the AIs to branch out, covering multiple strategies in order to win as often as possible. With enough generations, these AIs will approach a point where they span *all* viable strategies, and therefore are highly successful at the game against any opponent.

**Methodology**

This work hinges on a game that is complex enough to be interesting (*i.e.*, has multiple strategies that can be used), while also being tractable in the project's time frame (at most four weeks). *Backgammon* has been selected to fill this role. The classic game has a relatively tractable state space (30 pieces that can each be in one of 26 locations), yet has enough strategic depth to warrant study (*e.g.*, [Keeler and Spencer, 1975]). Backgammon has also been the focus of several artificial intelligence projects in the past. A similar

imitation learning approach (only using purely supervised learning techniques) provided positive results [Tesauro, 1989]. The long-running gold standard of Backgammon AI came from neural networks that were awarded for winning in matches against each other [Tesauro, 1994]. Other relevant studies include co-evolved [Pollack et al., 1997] and genetic program driven Backgammon AIs [Azaria and Sipper, 2005].

For this project, Eco-EA will be used to evolve AIs to play Backgammon [Goings et al., 2012]. Eco-EA is different than most evolutionary algorithms, as it assigns "resources" to the tasks to be completed in an effort to replicate ecological dynamics found in nature. When the first AI completes a certain task, that AI receives a substantial reward for its novelty. Over time, if that task continues to be solved then the resource reward sharply decreases. This still benefits the AI while also promoting exploration into different tasks. In theory, this should promote the AIs to predict the human moves and eventually switch focus to winning games so to maximize resources.

The Modular Agent Based Evolution Framework (MABE) [Bohm and Hintze, 2017] will be used to facilitate this study. Backgammon will be coded as a "world" in MABE, allowing for brain types to effortlessly be swapped and compared. For this study we will focus on artificial neural network and genetic programming brains. Following the precedent set by previous backgammon studies, AIs will be given all possible moves from the current state. The move with the highest score will be selected (either used or taken as a prediction, depending on the task at hand). Eco-EA will be used as implemented in the Empirical library (https://github.com/devosoft/Empirical) for this work.

In order to untangle whether rewarding both predictions and wins is beneficial, two controls will be conducted: one that only rewards predictions, and one that only rewards wins. To benchmark the final "champions" of this study, they will be pitted against the historical PubEval Backgammon program (https://www.bkgm.com/rgb/rgb.cgi?view+610), as is de facto tradition for Backgammon AIs. To further examine the results of this study, data will be recorded on various aspects of the evolutionary run: such as average/min/max fitness each generation, the performance of AIs on each task, and a suite of phylogenetic measurements [Dolson et al., 2018].

**Expected Results**

I expect the dual-reward Eco-EA approach to significantly outperform AIs evolved in both control conditions. Rewarding predictions should help AIs reach the point where they are loosely competitive, at which point the selective pressure to win games will help push them to evolve more advanced strategies. The control that does not reward predictions is expected to come closer to the combination approach, but only in runs where AIs overcome the initial hurdle of winning *any* games.

From data recorded on how each AI performs on each task combined with phylogenetic information, we can examine the history of the evolutionary runs that lead to overall success or failure. Early in a typical run, I expect to see AIs performing better on prediction tasks compared to game-playing tasks. For successful runs, my expectation is for AIs to eventually branch out from predictions and sweep many of the game-playing tasks. Looking at the phylogenetic data, I expect early AIs to diverge and cover different predictive tasks. Eventually I expect dominant AIs to emerge and cover multiple tasks, corresponding to the evolution of more complex agents.

**Broader Impacts**

If results are successful, this work will serve as a generalized framework for evolving AIs to play a wide variety of complex games. Future work can try this approach on games with larger state spaces and/or more strategies than Backgammon to test how well the approach scales. This framework has the potential to work for other complex tasks outside of game playing, evolving AIs capable of doing complex real-world activities. Another benefit of this work is that, as a stepping stone, it can produce AIs that mimic play-styles of individual humans. This benefit is two-fold: it can provide additional agents to play against or train other AIs against, and it can be used as an educational tool. If an AI closely models an individual's approach, this AI can be probed to examine where the individual's strategy fails. This information can then be used to create scenarios that force the individual to confront these shortcomings, creating learning experience tailored specifically to them.

# References


[Azaria and Sipper, 2005] Azaria, Y. and Sipper, M. (2005). GP-gammon: Genetically programming backgammon players. *Genetic Programming and Evolvable Machines*, 6(3):283–300.

[Bohm and Hintze, 2017] Bohm, C. and Hintze, A. (2017). MABE (modular agent based evolver): A framework for digital evolution research. In *Artificial Life Conference Proceedings 14*, pages 76–83. MIT Press.

[Dolson et al., 2018] Dolson, E., Lalejini, A., Jorgensen, S., and Ofria, C. (2018). Quantifying the tape of life: Ancestry-based metrics provide insights and intuition about evolutionary dynamics. In *Artificial Life Conference Proceedings*, pages 75–82. MIT Press.

[Goings et al., 2012] Goings, S., Goldsby, H., Cheng, B. H., and Ofria, C. (2012). An ecology-based evolutionary algorithm to evolve solutions to complex problems. In *Artificial Life Conference Proceedings 12*, pages 171–177. MIT Press.

[Keeler and Spencer, 1975] Keeler, E. B. and Spencer, J. (1975). Optimal doubling in backgammon. *Operations Research*, 23(6):1063–1071.

[Pollack et al., 1997] Pollack, J. B., Blair, A. D., and Land, M. (1997). Coevolution of a backgammon player. In *Artificial Life V: Proc. of the Fifth Int. Workshop on the Synthesis and Simulation of Living Systems*, pages 92–98. Cambridge, MA: The MIT Press.

[Sammut, 2010] Sammut, C. (2010). Behavioral Cloning. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, pages 93–97. Springer US, Boston, MA.

[Taylor et al., 2016] Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., and Ikegami, T. (2016). Open-ended evolution: Perspectives from the OEE workshop in York. *Artificial life*, 22(3):408–423.

[Tesauro, 1989] Tesauro, G. (1989). Neurogammon wins computer olympiad. *Neural Computation*, 1(3):321–323.

[Tesauro, 1994] Tesauro, G. (1994). TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural computation*, 6(2):215–219.