

The Effect and Evolvability of Observational Learning in A-Life

Eugene Chen, Satwinder Kaur, Cody Rosevear

Department of Computing Science
University of Alberta
Edmonton, Alberta, T6G 2E8, Canada
{echen|skaur4|rosevear}@ualberta.ca

Abstract

Observational Learning (OL), the form of learning that results from observing others rather than through direct trial and error, has been shown in previous work to improve machine learning along a number of dimensions (Hussein et al. 2017), (Simard et al. 2017). However, it remains understudied within the context of artificial life (A-life) programs. The following work uses a simple A-life scenario as a way to explore to what degree observational learning can be considered beneficial to an agent, such that the tendency for OL can be evolved rather than explicitly programmed within the agents. Results suggest that OL does have an effect, and that these differences confer enough of an advantage to allow the tendency to observe to out evolve a non-behavioural control gene.

1 Introduction

Many recent advances in the field of artificial intelligence (AI) are the result of significant improvements in the effectiveness and practicality of machine learning algorithms. Indeed, as the growth of data and computing power has steadily continued throughout the last half-century, many impressive feats of learning by machines that were once too computationally onerous to be considered feasible have become tractable, and the paradigms of deep and reinforcement learning (Goodfellow, Bengio, and Courville 2016), (Sutton and Barto 2018) have been profitably combined with advances in heuristic search (Chaslot et al. 2008) to yield breakthroughs across a number of problem domains (Moravčík et al. 2017), (Mnih et al. 2013), (Silver et al. 2017).

Contrarily, there have been no similarly high-profile breakthroughs which use learning algorithms that take advantage of the knowledge of other agent's actions via imitation or observation, or through explicit teaching. A-life is one experimental context where observational learning has the potential to significantly improve the performance of populations of agents along a number of metrics, including their longevity, fecundity, and survival rate.

The following work therefore explores what effects various genetic propensities to engage in OL have on population trends within such an environment, and investigates whether OL capacities confer enough of a benefit so as to be evolvable within the context of a simple A-life program.

2 Problem Formulation

The objective is to explore the effects of (OL) on sheep fecundity, longevity, and survival, and see to what extent these effects promote the evolution of OL. To do this, at the end of a run, we measure the average value of the gene across the entire sheep population, denoted by the variable $OLGene_{avg,t}$, at time $t = 0$ and $t = 1000$, expecting that the average value of the gene will have increased over time, such that relation 2.1 yields a value of True.

The value of $OLGene_{avg,t}$ at time $t = x$ is defined by equation 2.2, where n_t is the total number of living sheep at time t , and $OLGene_i$ is the current probability of observing for sheep i .

Finally, we consider a control gene, denoted by the variable $Control_{avg}$, which is computed in the same fashion as the $OLGene_{avg}$, but has no link to any type of sheep behaviour. We interpret the difference between the two genes, as described by 2.3, as a crude measure of the utility, and hence evolvability, of the observational gene.

$$OLGene_{avg,0} < OLGene_{avg,1000} \quad (2.1)$$

$$OLGene_{avg,t} = (\sum_{i=1}^{n_t} OLGene_i) / n_t \quad (2.2)$$

$$OLGene_{avg,1000} - Control_{avg,1000} \quad (2.3)$$

3 Related Work

Although there has been some prior work on the evolution of learning agents within A-life environments more generally (Ackley and Littman 1991), these

agents did not possess OL abilities of any kind, and learned only with respect to their own actions. More recent work has also explored the feasibility of evolving novel behaviours (Soares et al. 2018), (Bulitko et al. 2017), or agent brain architectures (Stanley and Miikkulainen 2002) rather than hand engineering either, but neither of these approaches are concerned with achieving OL per se. The former is a proof of concept that is concerned with evolving novel behaviours and detecting them efficiently, rather than novel learning capacities specifically, while the latter is focused on evolving architectures that optimize traditional non-OL capacities.

Insofar as true OL has been considered within an A-life context, it has been limited to the case of simple imitation, and the capacity to learn was not itself evolved; rather, it was hard-coded into the agents from the start (Le, O'Neill, and Brabazon 2018). The following work therefore seeks to both extend the type of learning used, such that it is no longer simply imitative, as well as evolve the very propensity to learn observationally, rather than hard-coding this propensity within the agents.

4 Proposed Approach

4.1 Overview

The experiment will be run in three phases. The first and second involve no evolution of the OL gene itself. Instead, the probability to observe will be hard-coded, so as to investigate its effect on sheep behaviour under the assumption that it is present across the entire population. Phase three will then remove the hard-coded value and initialize the gene for the first generation of sheep, allowing it evolve thereafter.

Environment: The environment consists of sheep prey agents, wolf predator agents, and grass that, once eaten, regrows at a constant rate. It utilizes a map that contains no objects apart from patches of ground with grass (coloured green), patches without grass (coloured brown), and surrounding cliffs (coloured black), which immediately kill any sheep that steps onto it. See figure 1 for a graphical depiction of the environment.

Agents: The sheep agents possess both an age and energy attribute, and reproduce via asexual reproduction once they meet a user-selected minimum age and energy level, losing half of their current energy in the process. The offspring's genes will be determined by sampling from a Gaussian distribution with a mean equal to the current value of each of their parent's genes, and a fixed standard deviation determined by the user-selected mutation rate parameter. The probability to observe gene is limited to the range between 0.00 and 1.00, excluding the latter boundary case.

Wolves and sheep take energy to move, and their energy levels are replenished by consuming sheep

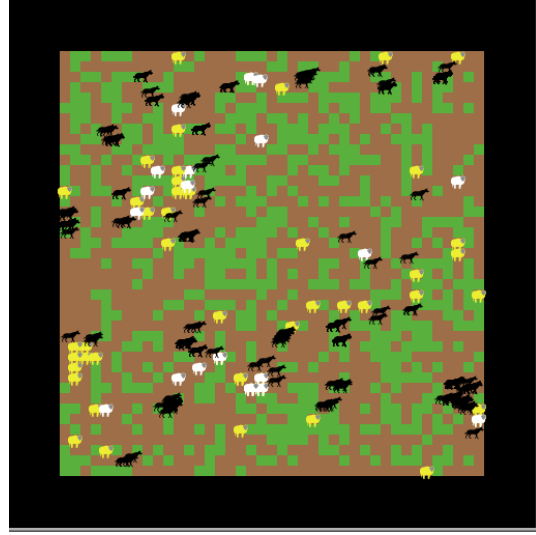


Figure 1: Map with cliffs environment

and grass, respectively. When an agent's energy levels drop to 0 or below, it dies, leaving behind no corpse. Wolves move by taking a fixed step size in the direction of the nearest sheep, while sheep move based on the output of a utility function that is computed by performing a two step search process on all patches within its immediate sight radius. The sheep looks at all possible patches that it can move to, and all patches within a user-set travel radius parameter from these potential target patches, and then computes the sum of the affinity function for wolves in each of the patches neighbouring the target patch, and adds this to a linear combination expressing the utility of the target patch itself.

The sheep thus seeks to find $\max_p u_p$ for all patches p within its sight radius, where u_p is the utility of the target patch, $AG(p_{grass})$, $AW(p_{wolves})$, $AW(p_{cliffs})$ are the products of the amount of grass, wolves, and cliffs with their respective affinities for the current sheep, and np are all of the neighbouring patches of the target patch within the sheep's travel radius. See equation 4.1 for detail. Sheep movement occurs directly by their jumping to a target patch once they have decided upon it using their utility function.

Agent's have both genes and attributes for affinities. The former are determined at birth, and thus not modifiable within an agent's lifetime, while the latter can be modified via learning. See Table 1 for a summary of the genes and attributes for each sheep.

$$u_p = AGp_{grass} + AWp_{wolves} + ACp_{cliffs} + \sum_{np} (AWnp_{wolves}) \quad (4.1)$$

Variable	Type
Affinity to Grass	Attribute
Affinity to Wolves	Attribute
Affinity to Cliffs	Attribute
Age	Attribute
Energy	Attribute
Sight Radius	Attribute
Probability To Observe	Gene
Initial Affinity To Grass	Gene
Initial Affinity To Wolves	Gene
Initial Affinity To Cliffs	Gene

Table 1: The variables (attributes and genes) that define a sheep

4.2 Event-Driven Observational Learning

For the purposes of the present experiment, observational learning occurs when a sheep detects within its observation radius one of a number of finite events: another sheep eating grass, being bitten by a wolf, or stepping onto a cliff. The sheep is provided with the ability to perceive the change in energy that occurs to each other sheep within its radius, and uses this to update its affinities according to formula 4.2, where A_x is the relevant affinity, which is mapped to the event in question, δ_{E_i} is the change in energy to the observed sheep due to the event, E_i is the observed sheep’s original energy, and obs is the set of all events observed by the learning sheep at that moment.

$$A_x = (\sum_{obs} \delta_{E_i}/E_i)/|obs| \quad (4.2)$$

4.3 Hypothesis

There are two scenarios under consideration; namely, the difference in performance between the sheep populations in phase one and phase two, where the OL is hard-coded to have a value of 0.00 and 1.00, respectively, and the penultimate third phase, where the OL gene is initialized and then allowed to evolve over time.

Concerning the former scenario, it is predicted that the sheep population for phase 2, which has their observational learning gene set to 1.00, will significantly outperform those sheep in phase 1, with respect to all naive measures of success: total amount of sheep born (fecundity), the mean age of the sheep, and their relative survival rate. This is a plausible scenario for a number of reasons: the presence of cliffs, which immediately kill sheep that step onto them, should provide those sheep which learn observationally with a competitive advantage that cannot be replicated by sheep with no observational learning capacity, as it will allow for sheep to learn from their mistakes of others, which they could not otherwise do without

Parameter	Value
Amount of land before cliffs	20
Grass Regrowth Rate	50
Max Grass Amount	1
Num Sheep Start	20
Sheep Grass Energy Gain	4
Sheep Min Breed Age	30
Sheep Min Breed Energy	30
Wolf Reproduction Rate	0.02
Num Wolf Start	10
Sight Radius Upper Bound	40
Initial Travel Radius	3
Affinity Gene Minimum Initial Val	-1
Affinity Gene Max Initial Val	1
Mutation Rate	1
Wolf Move Energy Loss	0.05
Sheep Move Energy Loss	1

Table 2: Experiment parameters for all runs

OL. Moreover, the capacity to observe the respective losses or gains in energy from being bitten by wolves or eating grass should allow them to survive longer and breed more within a single lifetime than sheep who ‘learn’ only via evolution, as in the evolution only case the ‘learning’ must necessarily come about as a result of agents with a poor genetic fit to the environment dying off. This in turn should decrease the rate of growth and mean age of such sheep, allowing those with OL to surpass them and survive better overall due to their capacity to adapt to the situation more quickly; without significant portions of the population having to die.

With respect to the latter scenario, where the OL gene is allowed to evolve after its initialization, it is expected that the probability to observe will increase its relative frequency in the population over time, with those who have a higher probability to observe coming to dominate the gene pool; thus increasing the average value of the probability to observe over time. Moreover, the expectation is that it will outperform a control gene given to each sheep, which is initialized to the same value as the OL gene but which has no effect on behaviour, and therefore should drift entirely at random via the same mutation strategy as used by the OL gene.

5 Empirical Evaluation

5.1 Testbed

All of the experiments were implemented using the NetLogo software package (Wilensky 1999), and based on modifications of previously published models (Soares et al. 2018), (Bulitko et al. 2017), (Wilensky 1997). The experiment was run for 100

independent trials, for 1000 time ticks per run, for each phase, using a range of random seeds for each run, ranging from 101 to 200, to ensure reproducible results and fair comparisons across different phases of the experiment. For all phases the initial non-OL genes of the sheep population are generated by sampling from a uniform distribution, with minimum/-maximum values of $-/+1$, respectively. For phase one and two, the probability to observe gene was set to values of 0.0 and 1.0, respectively, and not allowed to evolve thereafter. For phase three, the initial value for the gene was set to 0.01, but then allowed to evolve thereafter. See Table 2 for a summary of all of the parameters and their values as used in the experiment. Results for all three phases are presented below.

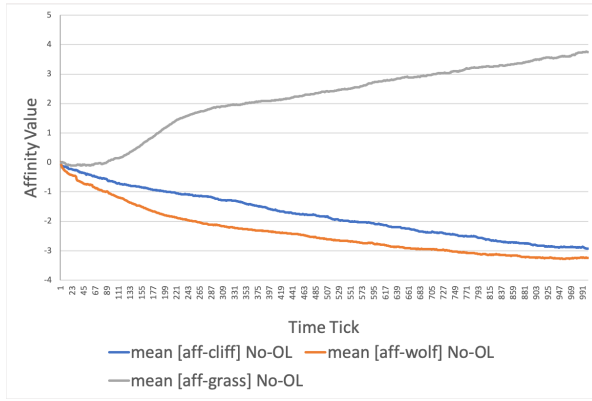


Figure 2: Mean Affinity Attributes Over Time: No Observational Learning

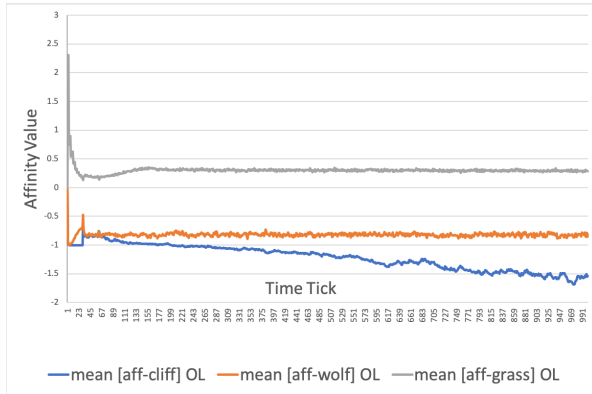


Figure 3: Mean Affinity Attributes Over Time: Observational Learning

6 Discussion

The affinity genes (omitted) and attributes for phase 1 (no OL) are, as expected, identical to one another

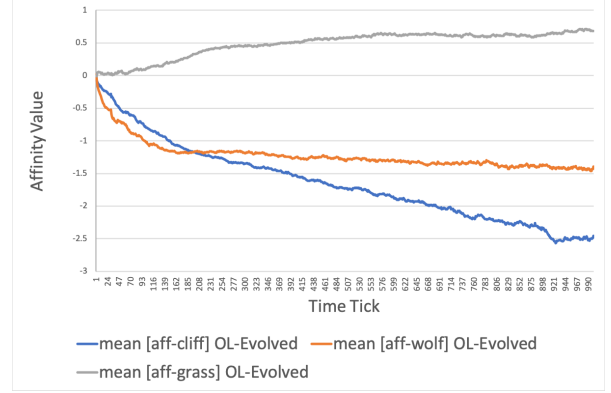


Figure 4: Mean Affinity Attributes Over Time: Evolution of Observational Learning

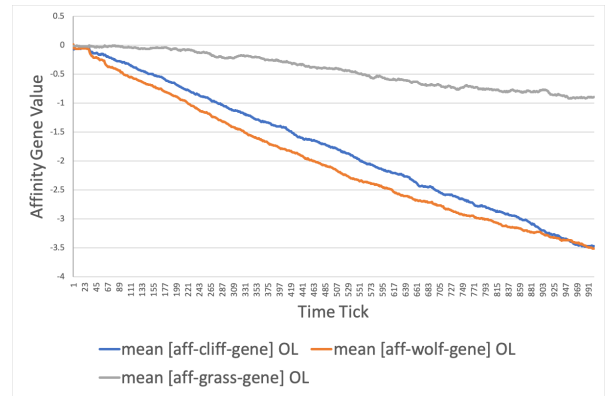


Figure 5: Mean Affinity Genes Over Time: Observational Learning

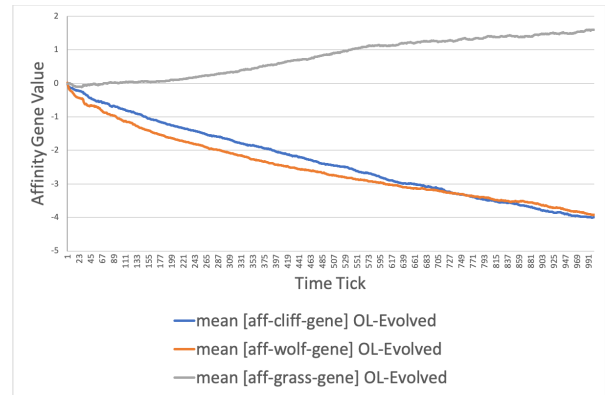


Figure 6: Mean Affinity Genes Over Time: Evolution of Observational Learning

(See Figure 2), with positive values for grass only and negative values for both cliffs and wolves. This reflects that the only way sheep can 'learn' is via natural

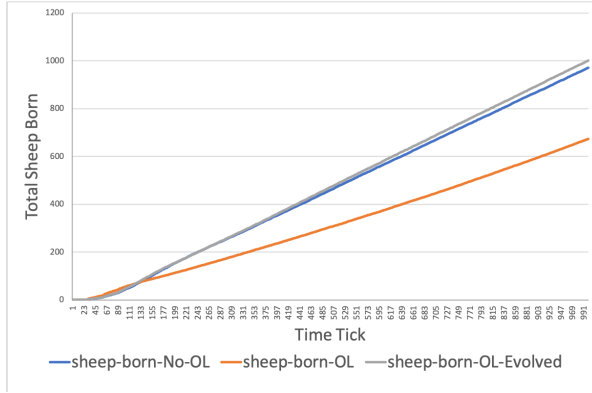


Figure 7: Total Number of Sheep Born Over Time: All Phases

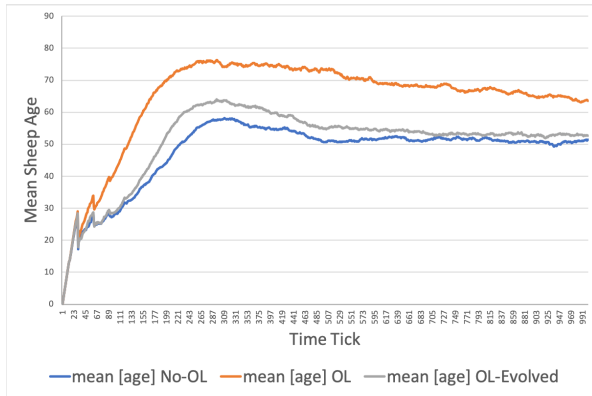


Figure 8: Mean Sheep Age Over Time: All Phases

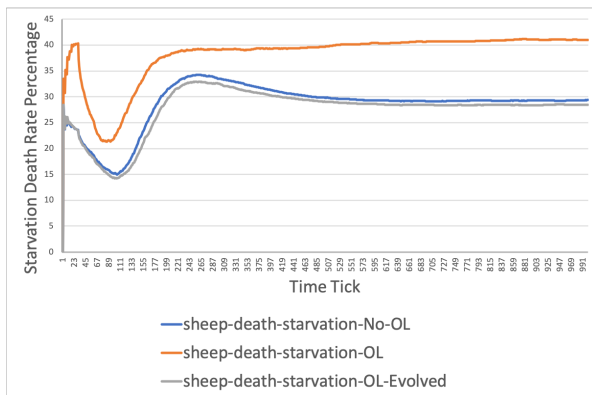


Figure 9: Sheep Death by Starvation Over Time: All Phases

selection with respect to the affinity genes. Contrarily, in phase 2 (OL) there is a divergence between the affinity genes and affinity attributes (See Figure 6 and 3). The former all trend negatively as time goes on,

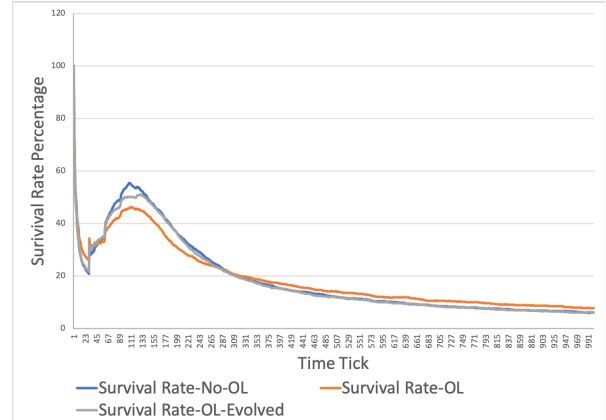


Figure 10: Sheep Survival Rate Over Time: All Phases

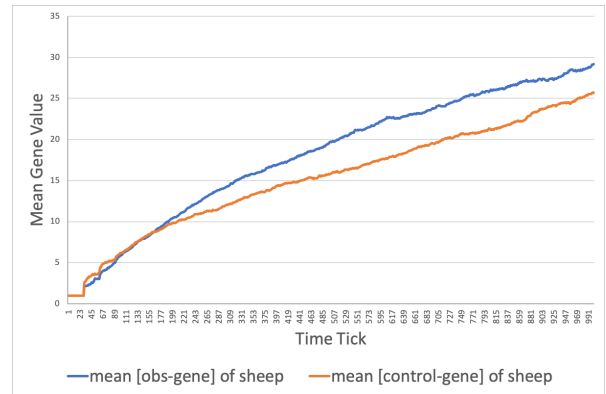


Figure 11: Evolution of Observational Learning: Observational Gene

including grass, with would normally be disastrous for the sheep population; however, the capacity to learn quickly on the basis of observation allows sheep to avoid the negative effects of their bad genes. This is similar to the shielding effect described in (Ackley and Littman 1991). Moreover, the values for all affinities do not trend in either direction as much as in the no-learning case. This is likely due to the fact that the learning formula does not use a learning rate, but simply overwrites the current affinities with the average proportional change to energy observed at a given time tick. This prevents the type of accumulation of higher values as is observed in the evolution only case.

Concerning metrics of utility, such as the fecundity, longevity, and relative survival of sheep, the presence of observational learning appears to be a mixed blessing. On the one hand, observational learning appears to yield less fecund sheep as compared to the case of no learning and evolved learning (See Fig-

ure 7). On the other hand, the sheep with OL both lived longer, on average, 8 and died at the hands of wolves and cliffs at a lesser rate than both of the other cases (See Figure 9). However, the survival rate of all sheep relative to the total number of sheep born for all cases shows no appreciable difference (See Figure 10). This suggests an overall trend wherein although the sheep with observational learning are less fecund, and tend to have less offspring, they appear to be 'smarter' in that they tend to both live longer and die less at the hands of avoidable events such as wolves and cliffs, due to their capacity to learn about the negative consequences of these events quickly, and act on the change made to their affinities.

Finally, despite this apparent trade-off between fecundity and 'smarter' sheep, allowing the OL gene to evolve nevertheless results in the average value of the probability to observe increasing throughout the population as time goes on, to such a degree that it outperforms the control gene (See Figure 4.2). It may be that what is needed is a longer run time in order to allow the OL gene to climb higher and break away more significantly from the control gene, as right now the difference is fairly modest.

7 Future Work

7.1 Short-Term Future Work

Currently, the learning rule right now is an all or nothing rule that simply replaces the current affinities of the sheep with the proportional change in energy observed. It would be more realistic to replace this proportionality with a more fine grained learning rate that lets agents learn while still allowing some degree of their prior experience to still be reflected in their current affinities. Of particular interest would be allowing the learning rate to itself be genetically determined: just as in real life there are differences between slow-learners and fast learners, differences in learning rate could be advantageous or not in different environmental circumstances, and allowing this to be tuned by evolution would extend the question of to what degree is observational learning evolvable, beyond merely the propensity to engage in it.

7.2 Long-Term Future Work

One avenue for long-term future work would be to extend the question of whether agents can be induced to evolve the capacity to explicitly teach one another, rather than simply engaging in imitative behaviour or learning via observation. This is a natural progression of the current question in terms of the complexity of behaviour, and future experiments could focus on exploring the difference in necessary and sufficient conditions for the emergence of imitative and/or observational learning versus explicit machine teaching.

Additionally, the action of explicitly teaching another agent could be defined in such a way so as to carry some type of cost, as it does in the real world, in terms of time and/or energy spent. The simplest case would be to structure the environment in such a way that by taking time to teach other agents, the current agent suffers from an opportunity cost, since their time could be better spent gathering resources for themselves. This adds a game-theoretic component to the question of whether to teach or not to teach at any given point in time, and depending on the environment this can be a non-trivial question, as it is conceivable that there would be some situations where teaching others certain skills would facilitate cooperative strategies that are better than greedy ones which focus only on the agent's current short term benefit.

Finally, with respect to both imitation/observational learning and explicit teaching, one could explore how kinship relations would alter the scenarios and evolutionary outcomes of such kinds of learning. If there is a genetic relationship between two agents, it is, in some sense, within the potential teacher's interest to secure suitable skills for their kin to survive, at least from the gene's point of view (Dawkins 2006). Exploring to what extent genetic markers of kinship alter the evolvability of both the willingness to imitate or teach others is another natural next step.

8 Conclusions

We have proposed combining the of study non-imitation based observational methods with an A-life context, to explore to what extent the former can be made useful enough to evolve rather than require being hard-coded. Experiments suggest that OL does have a positive effect when contrasted with agents who exhibit no observational learning, although it is not uniformly positive, as it seems to limit agent fecundity while simultaneously allowing for longer average lifespans and similar relative survival rates. Nevertheless, the OL gene conferred enough of an advantage to evolve in phase three and to outperform the control gene, on average, over many runs. This suggests that further and more rigorous study of the evolvability of observational learning is warranted.

Acknowledgments

The authors thank Vadim Bulitko for helpful commentary and constructive criticism concerning the present work, and all contributors to the original NetLogo code-base (Bulitko et al. 2017) (Soares et al. 2018) upon which our implementation is based.

References

- Ackley, D., and Littman, M. 1991. Interactions between learning and evolution. *Artificial life II* 10:487–509.
- Bulitko, V.; Carleton, S.; Cormier, D.; Sigurdson, D.; and Simpson, J. 2017. Towards positively surprising non-player characters in video games. In *Proceedings of the Experimental AI in Games (EXAG Workshop) at the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 34–40.
- Chaslot, G.; Bakkes, S.; Szita, I.; and Spronck, P. 2008. Monte-carlo tree search: A new framework for game ai. In *Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*, 216–217.
- Dawkins, R. 2006. The selfish gene: with a new introduction by the author. UK: Oxford University Press. (Originally published in 1976).
- Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Hussein, A.; Gaber, M. M.; Elyan, E.; and Jayne, C. 2017. Imitation learning: A survey of learning methods. *Association for Computing Machinery Computing Surveys* 50(2):21:1–21:35.
- Le, N.; O’Neill, M.; and Brabazon, A. 2018. The baldwin effect reconsidered through the prism of social learning. In *2018 Institute of Electrical and Electronics Engineers (IEEE) Congress on Evolutionary Computation (CEC)*, 1–8. Institute of Electrical and Electronics Engineers.
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- Moravčík, M.; Schmid, M.; Burch, N.; Lisý, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356(6337):508–513.
- Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. 2017. Mastering the game of go without human knowledge. *Nature* 550(7676):354–354.
- Simard, P. Y.; Amershi, S.; Chickering, D. M.; Pelton, A. E.; Ghorashi, S.; Meek, C.; Ramos, G.; Suh, J.; Verwey, J.; Wang, M.; et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742*.
- Soares, E. S.; Bulitko, V.; Doucet, K.; Cselinacz, M.; Soule, T.; Heck, S.; and Wright, L. 2018. Learning to recognize a-life behaviours. *Poster collection: The Annual Conference on Advances in Cognitive Systems (ACS)*.
- Stanley, K. O., and Miikkulainen, R. 2002. Evolving neural networks through augmenting topologies. *Evolutionary computation* 10(2):99–127.
- Sutton, R. S., and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.
- Wilensky, U. 1997. Netlogo wolf sheep predation model. <http://ccl.northwestern.edu/netlogo/models/WolfSheepPredation>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
- Wilensky, U. 1999. Netlogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Appendices

Isolating the Observational Learning Gene

In light of the similar patterns of growth observed between the control gene and the OL gene, in addition to the former’s relatively low ceiling in the original experiment, we performed a few additional exploratory experiments to assess the potential impact of the OL gene under different initialization and evolution schemes for the affinity genes, in an effort to decouple the effect of affinity gene evolution on the OL’s gene’s ceiling.

We ran three additional experiments (see figure 12) using the same testbed as articulated earlier, but modified so that the OL gene evolves under a number of different conditions: with fixed affinity genes that do not evolve, initialized to 0; with fixed affinity genes that do not evolve, initialized randomly; and with affinity genes that are allowed to evolve, as before.

These results demonstrate a much higher effect for OL in the absence of impactful affinities, which suggests that there is some sort of co-evolution/interaction occurring between the affinity and OL genes that reduces the latter’s utility, to some extent. Insofar as good genes can do the job well enough the impact of OL appears to be limited by these ‘good’ genes. Exploring more environments, and allowing the simulation to run longer may allow the OL gene to evolve further; further disentangling the interaction between the affinities and OL is another topic for future work.

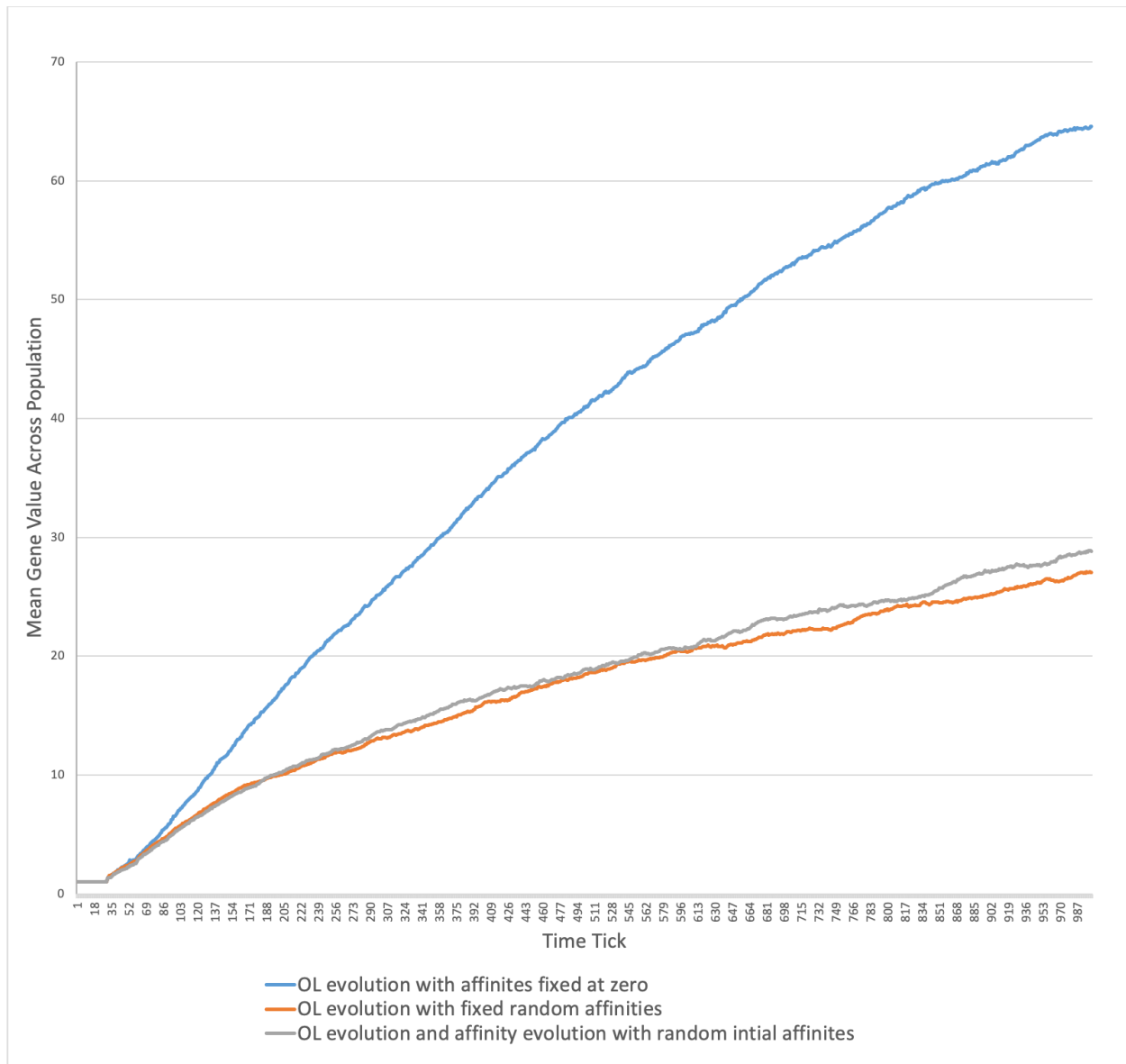


Figure 12: Evolution of Observational Learning:
Across Affinity Initializations