Introduction

Since the introduction of the von Neumann-Morgenstern utility theorem, the use of a utility function was the dominant explanation for risk aversion. However, Niv et al. showed in 2002 that risk aversion can be developed in simple models through reinforcement learning.[1] We study how risk aversion is affected by the introduction of knowledge sharing and resource sharing in a group of learning agents, as well as how these factors affect the death rate of the population.

Experimental Design

Our simulation consists of a population of 100 agents and a length of 1000 time steps. With the goal of maximizing the reward obtained, each agent is presented with two possible choices, *A* and *B*, which represent a "safe" choice and a "risky" choice, respectively. Each agent is initialized at $t=0$ with expected rewards for each option of 100+X, with $X \sim U([-.5,.5])$.[2] Centering the expected rewards on the same value ensures that, on average, each option has an equal probability of being chosen during the first time step. Each time step consists of four basic substeps: first, each agent makes a choice between options *A* and *B* and, based on this choice, receives a reward which is either placed into the population's resource pool (if the population is sharing resources) or added to the agent's fitness (if the population is not sharing resources); agents then share knowledge, if applicable, and update their expected rewards for each option based on the knowledge obtained in the previous substep; each agent then loses fitness, either drawn from the population's resources pool or from the individual agent's fitness, to simulate a

[1] Yael Niv et al., "Evolution of Reinforcement Learning in Uncertain Environments", *Adaptive Behavior* 10, no. 5 (2002), doi: 10.1177/10597123020101001

[2] Future versions of the experiment will initialize expected rewards to a constant value across all agents since the improvement was made so that the choice-making algorithm no longer requires that the values be different.

"cost of living" for this time step; finally, any agent which has a fitness of less than zero is removed from the population to simulate the death of that individual.

During the first substep, each agent chooses between options *A* and *B* by allotting probabilities of choosing these options based on a Boltzmann distribution then selecting based on a randomly generated value between 0 and 1. The underlying computation is as follows: compute the partition function of the distribution, *Z*, based on the expected rewards for all options, such that

$$Z=\text{Sum}[\text{Exp}[R_i/T]]$$

and then set the probability of choosing the *i*th option to

$$P_i= \text{Exp}[R_i/T]/Z$$

where R_i is the expected reward for the *i*th choice and *T* is an additional parameter which describes the "temperature" of the system, determining the likelihood of an agent to choose an option which does not have the highest expected reward. A value of *T* which approaches zero causes the agent to choose the "best" option with a probability which approaches one. On the other hand, as *T* approaches infinity all options become equally likely to be chosen, regardless of their expected reward. For this experiment, we chose a value of T=20.

After choosing an action to take, the agent is rewarded 100 fitness if action *A* was chosen or either 0 or 200 fitness, with equal probability, if action *B* was chosen. Thus, the actual expected value of the reward for each action is 100 fitness.

After all choices have been made and rewards obtained, each agent updates its expected reward for each option. If the agents are not sharing knowledge, the expected reward for the choice they made is updated as

$$R_i \leftarrow (1-\eta)R_i+\eta r$$

where η is the learning rate parameter and r is the reward most recently obtained. The expected rewards for options that were not chosen are not updated. If the agents are sharing knowledge, the expected reward for the chosen action is updated as

$$R_i \leftarrow (1-\eta)R_i + \eta W r + \eta(1-W)/(N_i-1)\text{Sum}[r_j,\{j,1,N_i-1\}]$$

where W is the parameter which determines how much weight an agent gives its own experiences compared to those of other agents, N_i is the number of agents who chose this action and r_j is the reward that each of these agents obtained. N_i-1 is used so that the updating agent does not count itself twice. The expected reward for actions which this agent did not choose are updated as

$$R_i \leftarrow (1-\eta)R_i + \eta W R_i + \eta(1-W)/N\text{Sum}[r_j,\{j,1,N_i\}].$$

Once all of the previous substeps have been completed, each agent loses 100 fitness. If the population is not sharing resources, this is taken directly from the agent's fitness. If this leaves the agent with less than zero fitness, that agent is removed from the population. If the population is sharing resources, the 100 fitness for each agent comes from the resource pool. If this leaves the pool with less than zero fitness, the entire population dies.[3]

We ran this simulation thirty times each for four cases:[4] the population exhibits neither knowledge nor resource sharing; the population exhibits knowledge sharing only; the population exhibits resource sharing only; and the population exhibits both knowledge and resource sharing. At each time step we recorded three dependent variables: the proportion of the population which favored option *A* over option *B*; the average difference between the expected reward for *A* and *B*; and the proportion of the original population which still remained.

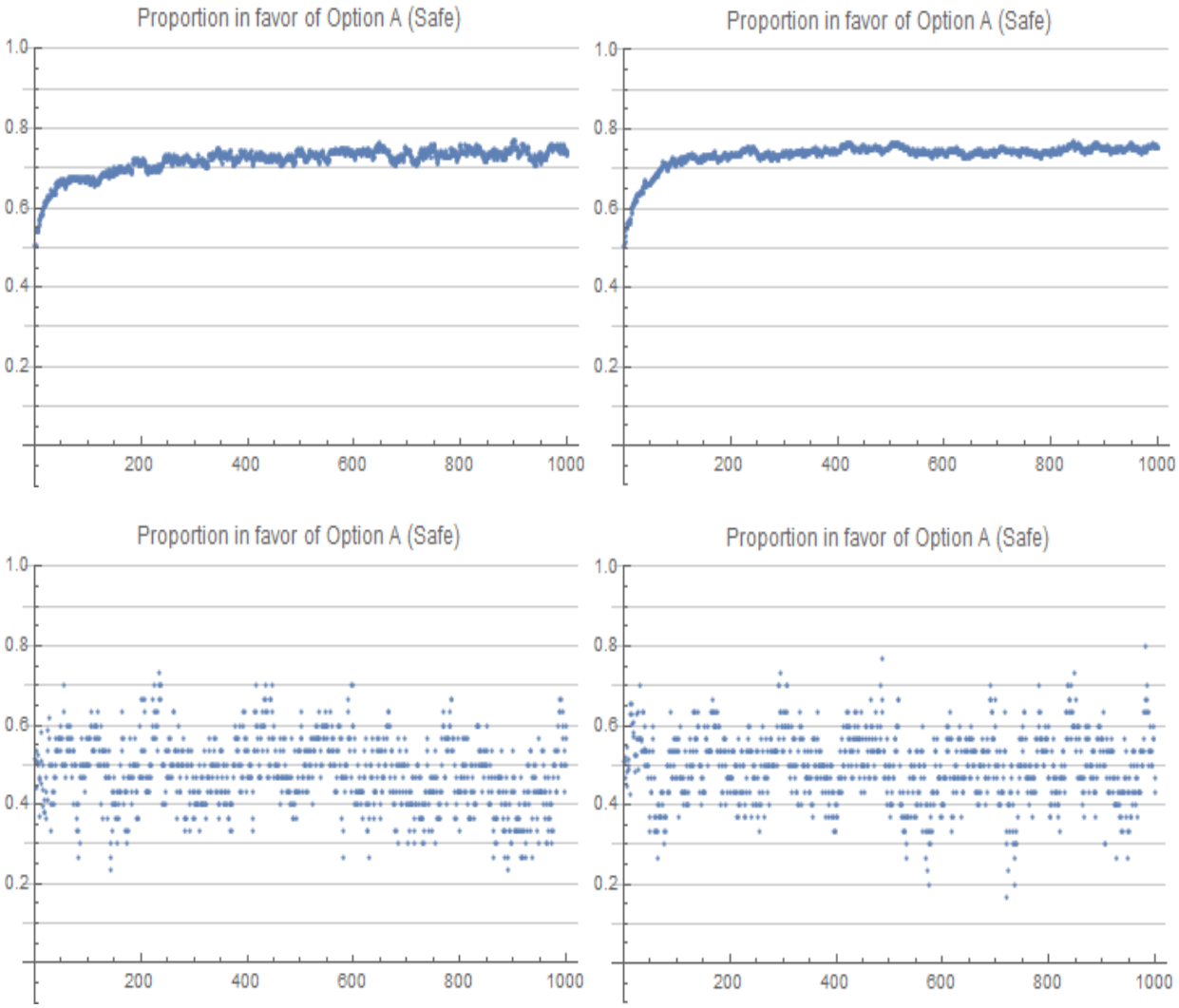
---

[3] In a population with partial resource sharing, only a portion of rewards are added to the pool and only a portion of the cost at each time step is removed from the pool. If the pool is depleted in this case, then each agent still has the opportunity to pay the rest of the cost for as long as it has individual fitness to spare.

[4] We only used full or no resource sharing. Similarly, we only used W=1 for no knowledge sharing and W=1/N_i for full knowledge sharing.

Results and Analysis

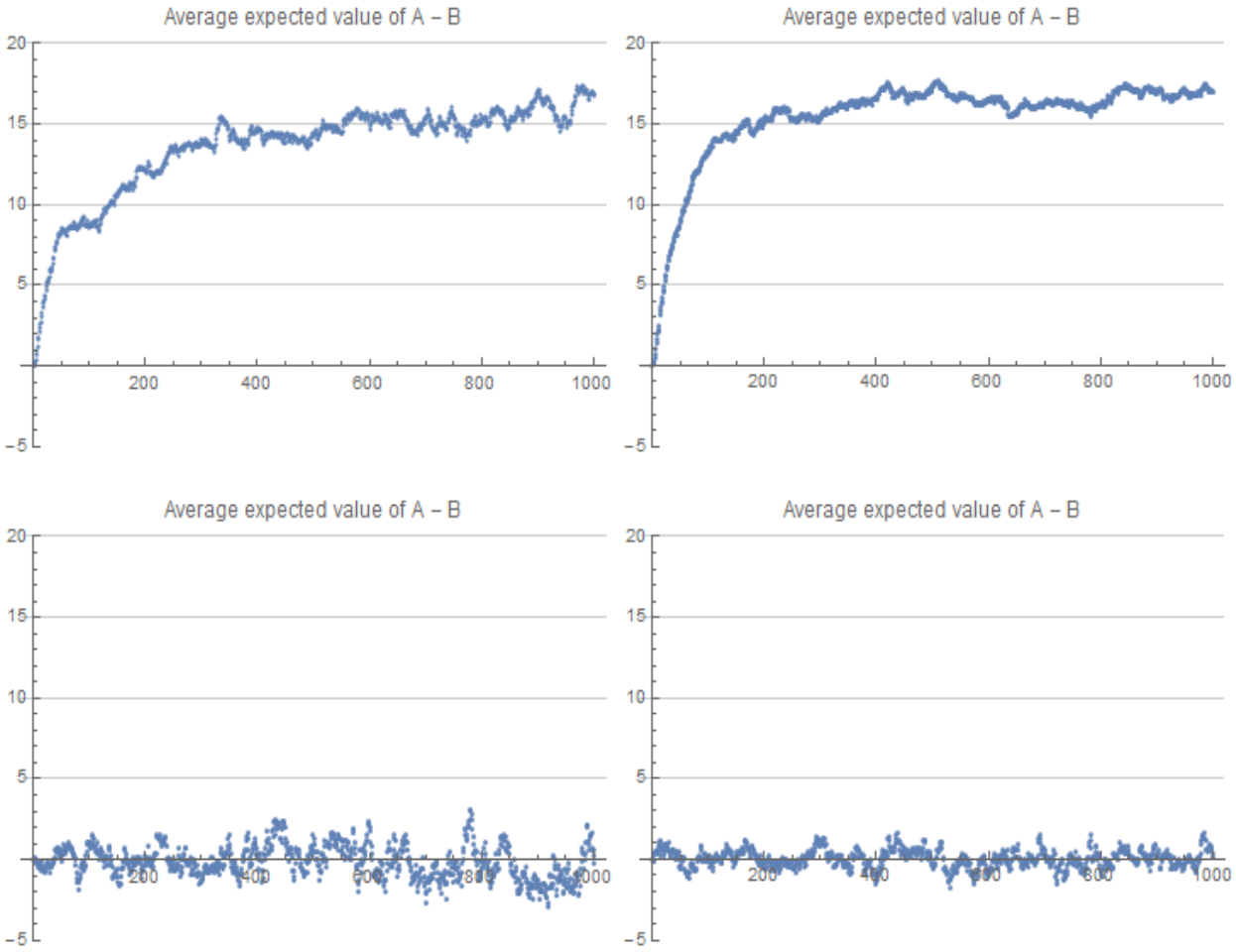
We grouped the thirty runs for each case together and averaged them to produce the plots below.



The top row shows the two cases without knowledge sharing. The left column shows the two cases without resource sharing. The x-axis is the time step.

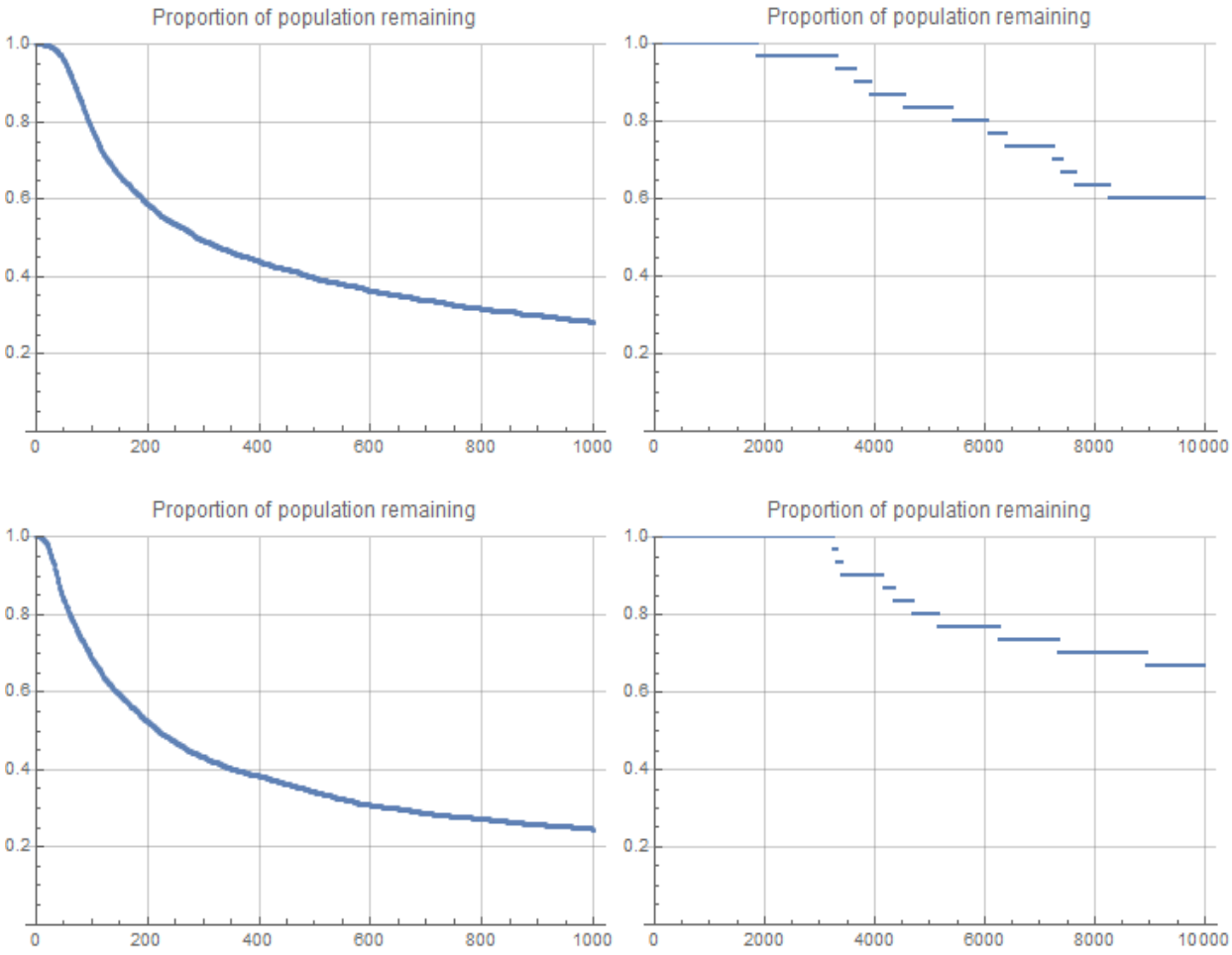
There is a clear difference between the apparent risk aversion in the populations without knowledge sharing and the apparent risk-neutrality in the populations with knowledge sharing. A

randomized analysis of variance[5] of the series in the left column (analysis of all sets is still in progress) shows that the two cases result in different levels of risk aversion with a p-value of 1.



Again, the effect of knowledge sharing on the learned risk aversion is clear in these plots, which measure how much option *A* is preffered over option *B* among the populations over the course of the runs. There also seems to be a noticeable difference between the plots with resource sharing and those without, especially in the first 300 time steps. However, analysis of this effect is yet to be completed, so this is currently just speculation.

[5] Piater, J. H. (1999). *A Randomized ANOVA Procedure for Comparing Performance Curves*. Cambridge, Massachusetts: MIT Press.



In order to see any die-off in the populations with resource sharing, these cases were allowed to run for 10000 time steps. This limits statistical analysis that can be done across columns, but allows us to see a difference (and test to determine if this difference is staistically significant) between the two cases with resource sharing.

The effect of resource sharing on the survival of the population for the first 1000 time steps of its life is obvious. While the two cases without resource sharing had averages of roughly 25% and 22% survival after this time, none of the total of 60 runs with resource sharing had the populations die for nearly twice that length. Analysis of the two cases without resource sharing with an α level of .05 rejected the null hypothesis that knowledge sharing has no effect on the survival of the population (given no resource sharing). This coincides with the hypothesis that a

population which shares knowledge is more likely to take risks—due to the relative risk-neutrality they learn—and die off as a result of more failed risks than a population without knowledge sharing.

Conclusion

Although complete analysis of the results of our experiment has not been finished, preliminary analysis suggests that knowledge sharing has a significant impact on the level to which a population learns to become risk averse, whereas resource sharing affects how long the population lasts before significant die-off occurs.

Future work includes completing the analysis of the results from the conducted experiment. Future extensions of this experiment might include varying the amount of knowledge and resource sharing so that they take on values between 0 and 1, changing the rewards given by options *A* and *B* to determine how large a risk premium risk averse populations are willing to pay, or introducing more options to the agents. Other possible alterations include introducing nonhomogeneous distributions of knowledge and resource sharing parameters among the population to measure other factors such as resource contribution within the population, or even having nonhomogeneous rewards for otherwise homogeneous agents.