# When Innovation Matters: An Analysis of Innovation in a Social Learning Game.

**Ryan Carr, Eric Raboin, Austin Parker, Dana Nau**
Department of Computer Science, University of Maryland
College Park, MD, 20742, USA
{carr2,eraboin,austinjp,nau}@cs.umd.edu

### Abstract

This paper examines the value of innovation within a culture by looking at "innovate" moves in the Cultaptation Project's social learning game (Boyd et al. 2008). We produce a mathematical model of a simplified version of this game, and produce analytic methods for determining optimal innovation behavior in this game. In particular, we provide a formula for determining when one should stop innovating and start exploiting one's accumulated knowledge. We create an agent for playing the social learning game based on these results, and in our experiments, the agent exhibited near-optimal behavior.

## Intro

The development of innovations is important to a society, yet the exploitation of those same innovations is clearly also important and many times, a tradeoff must be made between the two. This paper examines the utility of a simple kind of innovation in the Cultaptation Project's social learning game, and shows how we can solve the problem of deciding when it is time to stop innovating and exploit the accumulated knowledge.

The European Commission's Cultaptation Project was created to address the evolution of human cultures, and the project's researchers have created a game to examine the relative merits of social learning strategies (Boyd et al. 2008). Each player in the game has three kinds of possible moves: *innovate*, *observe*, and *exploit*. The game's authors devised these moves to be simple analogs of the following three kinds of activities: innovating (spending time and resources learning something new), observing (learning something from another player), and exploiting (using the learned knowledge). At each step of the game, each player must choose one of these three activities, and the algorithm or rules that a player uses for making this choice are the player's "social learning strategy."

Finding a good social learning strategy is clearly a hard question, particularly because the game is defined in such a way that the utility functions are *a priori* unknown. However, hidden within the hard question of determining social strategies is *another* hard question of determining the value of innovation: if one decides to use innovation instead of observation, what is the best way to accomplish this? This second question is the focus of our study.

While the Cultaptation social learning game is quite hard to model mathematically, we have been able to construct a mathematical model for a simplified version in which only "innovate" and "exploit" moves are allowed. We further create an agent we call the *smart-innovator* who performs nearly optimal in our experiments. The contributions of this paper are:

- A mathematical model for two simplified versions of the Cultaptation game called the SRI and VRI games.
- A formula for determining optimal strategies in the SRI game and another formula for determining optimal strategies in the VRI game. These formula require access to the underlying probability distribution.
- A *smart-innovator* agent, which makes a model of the underlying probability distribution and uses it to play the SRI and VRI games according to the formulas.
- Experiments detailing the performance of the *smart-innovator* agent, showing it to behave near optimally with wide varieties of underlying probability distribution functions.

## Definitions

### Social Learning Game

In this section, we lay out the Cultaptation Project's social learning game from (Boyd et al. 2008). We assume a probability distribution $\pi$ over the integers. There are $n_b$ exploit moves, with each move's utility value being drawn from $\pi$. We further assume a *change probability* of $c$. On each round of game play, with probability $c$, each move's value will be replaced with another drawn from $\pi$. Let $v_{i,j}$ be the value of move $i$ in round $j$. When a player makes an exploit move $i$ on round $j$ they receive the utility $v_{i,j}$. A player's total utility will be the sum of its utility on every round.

There will be $n$ agents in this environment. In the general game, each agent $a_i$ can make two other moves apart from the $n_b$ exploit moves: innovate (I) and observe (O). Upon making an I move in round $r$, the value $v_{i,r}$ of a randomly chosen move $i$ gets added to the agent's repertoire as the value of move $i$. The agent receives no utility on rounds

| round # | 1 | 2 | 3 | 4 | 5 | ... | $k$ |
|---|---|---|---|---|---|---|---|
| I1's move | I | 1 | 1 | 1 | 1 | ... | 1 |
| I1's Utility | 0 | 3 | 6 | 9 | 12 | ... | $3 \cdot (k-1)$ |
| I2O's move | I | I | O | 3 | 3 | ... | 3 |
| I2O's Utility | 0 | 0 | 0 | 8 | 16 | ... | $8 \cdot (k-3)$ |

Table 1: The sequence of moves from Example 1 and their payoffs.

where she makes an I move. Upon making an O move, an agent will get to observe the value received by some other agent who made any exploit move on the last round. Agents receive no utility for O moves. When a move is observed, the observed move's value on the last move ($v_{i,r-1}$) is added to the agent's repertoire. If no other agent made an exploit move last round, the observing agent receives no information.

The agent may only make I or O moves and moves from the repertoire (having been put there through either an I or an O move). Notice that because of the probability of change $c$, the value of $v_{i,r}$ on the current round $r$ may not be the same as what is stored in the agent's repertoire.

**Example 1** *Consider two strategies: the innovate-once strategy (hereafter I1) which innovates exactly once and exploits that innovated move for the rest of the game, and the innovate-twice-observe-once strategy (hereafter I2O) which innovates twice, observes once, and exploits the higher valued move for the rest of the game. For simplicity of exposition, we allow only four exploit moves: 1, 2, 3, and 4; and exactly two agents, one I1 and one I2O. We suppose a uniform distribution over* $[1, 10]$ *(with mean* 5*) and a probability of change of* 0*. Suppose the initial values for the moves are:* $v_{1,0} = 3, v_{2,0} = 5, v_{3,0} = 8, v_{4,0} = 5$*. On the very first move, I1 will make an innovate, which we suppose gives I1 the value of move 1, putting* $v_{1,0}$ *as the value for move 1 in I1's repertoire. On every sequential move, I1 will make move 1, exploiting the initial investment. If the agent dies* $k$ *rounds latter, then the history of moves and payoffs will be that given in Table 1; giving a utility of* $3 \cdot k - 1$*.*

*In contrast, I2O will make an innovate, giving the value for move 3:* $v_{3,1} = 8$*, then makes another innovate giving the value for move 2:* $v_{2,2} = 5$*, and finally observes. On move 2, I1 made move 1, and since these are the only two agents, this was the only exploit move made. Therefore I2O observes that another agent got a value of 3 from move 1 last round. On move 4, I2O's repertoire consists of* $\{v_{1,2} = 3, v_{2,2} = 5, v_{3,1} = 8\}$*. Since the probability of change is* 0*, the obvious best move is move* 3*, which I2O makes for the rest of her life. The average per-round payoff of I2O on round* $k$ *is* $8 \cdot k - 3$*, so for rounds 2 to 4, I2O will actually have a worse than I1, while after round 4, the utility of I2O will actually be higher.*

There are many potential setups for these sorts of games. Generally the *de facto* objective is to acquire more utility than other players, though one can imagine games where the objective is for an individual or group to maximize utility. In the Culptaptaion social learning competition, an evolutionary setup is used. Each game starts with 100 agents. Each round each agent has a 2% chance of dying. On death, an agent is replaced, and if there is no mutation, the strategy used by this newborn agent is chosen from the agents currently alive according to their average lifetime payoff (the new agent is naïve and shares nothing except social learning strategy with its parent). In this game, the strategy with the highest average payoff per round is the most likely to propagate. Mutation happens 2% of the time, and if there is a mutation, then one of the competing strategies is chosen at random and added to the pool (without regard for any strategy's performance). Through mutation, new strategies can be introduced into otherwise homogeneous populations.

In the social learning competition, there are two sorts of games into which an agent may be placed. First is a pairwise contest, where one contestant strategy starts with a dominant population of agents, and an invader strategy is introduced only through mutation. The second is the melee contest, which starts with a dominant strategy of simple asocial learners, and is invaded by several contestant strategies through mutation. We call any instance of either an invasion or a melee game a *Culptaptaion game*.

### Simplified Games

In this paper, we will examine two simplified forms of the Cultaptation social learning game.

The first form is the finite-round innovation game. In this game, we take the original social learning game and eliminate observe moves, the probability of change, and the possibility of death. We further set the number of rounds to some positive integer $l$. We call this the set-round-innovation game (SRI game). In the SRI game there is only ever one agent. The goal of the SRI game is to achieve maximal utility over the entire game.

The second simplified form is identical to the first, except that agents die with probability 0.02 each round, instead of living a fixed number of rounds. We will call this the variable-round-innovation game (VRI game).

## Problem Description

The motivation for this work comes from the following result, which implies that innovation moves must be performed even in the version of the social learning game on which the social learning competition will be based.

**Proposition 1** *If no agent ever innovates in any form of the social learning game, then the utility of all agents will always be zero.*

**Proof:** All utility comes from exploiting moves available in an agent's repertoire. Any move in the repertoire due to an observation requires there to have been an earlier exploit. Any move in a repertoire was therefore originally discovered by an innovate move. Thus any non-zero utility implies that at least one agent made at least one innovate at one point.

This implies that innovation is necessary in all social learning games, including the Cultaptation game. Thus even

in the full Cultaptation game there is motivation to determine how many times one must innovate in order to optimize one's utility.

A quick analysis of the SRI game determines that only certain kinds of innovations are useful. Consider any innovate move made after an exploit move. It is possible that that innovation discovers a move with higher utility than the previous exploit. In this case, we would want to have innovated before exploiting – in fact, in the general case, all innovates should come before any exploits. Therefore we need only consider strategies which innovate a given number of times.

The question remains, however, how many innovates is best? This is our problem.

**Definition 1 (Optimal Innovation Problem)** *Given a probability distribution $\pi$, what is the optimal number of innovates in the SRI game? the VRI game?*

## Analytic Results

### SRI Game

In this section, we introduce and prove a formula which allows for the computation of the optimal number of innovates for a given distribution in the SRI game.

To present these results, we will represent the utilities of each of the $n$ actions as a set $V = v_1, v_2, \ldots, v_n$ and we will assume, without loss of generality, that $v_1 \leq v_2 \leq \cdots \leq v_n$.

**Possible Strategies** First, we examine the strategies that players may adopt for agents in this version of the game. Players must choose either I or E for each of $l$ rounds. Thus, there are $2^l$ possible strategies. Note, however, that there are far fewer intelligent strategies. For instance, I is the only move that makes sense for the first round, since the agent does not know any actions and, thus, cannot choose one to exploit. Similarly, E is the only move that makes sense for the last round, since the agent would not have the opportunity to use any action it learned by choosing I.

Finally, note that, since action utilities do not change, it never makes sense to choose a strategy with an I move following an E move, since this strategy would be guaranteed to do at least as well by swapping the two moves, since the total number of times the agent exploits remains the same, and the utility of the action it chooses to exploit is at least as high. Thus, the only strategies worth considering are those that begin by innovating $k$ consecutive times, where $0 < k < l$, and then exploit $l - k$ consecutive times. For the rest of the analysis, we will refer to the strategy that begins by innovating $k$ times as $S_k$.

**Expected Utility of a Strategy** Since all of the strategies we are concerned with contain some number of I moves followed by E moves, we can obtain the expected utility of a strategy by multiplying the utility of the best action found with $k$ innovates by $l - k$, the number of times the strategy exploits:

**Proposition 2** *Let $F(k, V)$ be the expected utility of the best action found with $k$ random draws from $V$. Then the expected utility of $S_k$ is:*

$$E(S_k) = (l - k)F(k, V)$$

Now, however, we need to derive $F(k, V)$. Since $F$ is the expected maximum value of $k$ draws from $V$, we can obtain it by summing the maximum value of every possible sequence of $k$ innovates, and dividing by the number of possible sequences. Our assumption that $v_1 < v_2 < \cdots < v_n$ will help here; we note that, if on a given sequence of innovates the maximum utility discovered is $v_i$, then the other utilities discovered is some permutation of $k-1$ values from $v_1, v_2, \ldots, v_{i-1}$. Since there are $P(i-1, k-1) = \frac{(i-1)!}{(i-k)!}$ of these permutations and the maximum value can be found on any one of the $k$ innovates, there are $k\frac{(i-1)!}{(i-k)!}$ ways to discover a maximum value of $v_i$. Since there are $P(n, k) = \frac{n!}{(n-k)!}$ possible sequences of discoveries, we know that

$$F(k, V) = \frac{\sum_{i=k}^{n} \left( k\frac{(i-1)!}{(i-k)!} v_i \right)}{\frac{n!}{(n-k)!}}. \quad (1)$$

Now that we have $F(k, V)$, and we know that the only strategies worth considering are ones that innovate for $k$ moves and then exploit for $l-k$ moves, we can easily calculate the optimal strategy by finding the value of $k$ that maximizes the strategy's expected value.

**Theorem 1** *Given a distribution $V$ and lifetime $l$, the optimal strategy for the SRI game is to innovate a number of times equal to*

$$argmax_k \left( (l - k)F(k, V) \right) \quad (2)$$

*and exploit thereafter.*

This theorem follows from the discussion above.

**When to Stop Innovating?** Now that we can find the expected utility of strategies, we can tell if adding more innovate moves to a strategy will give us a higher expected utility. This will be helpful if, for example, we are building an agent that must decide each turn whether to keep innovating or not. If we want to get the maximum expected utility, we are interested in the value of $k$ such that performing $k + 1$ innovates does not improve out expected utility. In other words,

$$(l - k)F(k, V) \geq (l - k - 1)F(k + 1, V). \quad (3)$$

### Experimental Results for SRI Game

To test the effectiveness of Formula 1 and to ensure that no mistakes were made in its calculation, we compared the expected value predicted by the formula to actual expected values achieved in simulation experiments. To do this, we chose two distributions for $\pi$: normal and exponential distributions, each with mean 50 and variance 850. For each distribution, we considered games which lasted exactly 100 rounds, and computed the expected average per round utility of an agent which performs $i$ innovations by averaging that agent's utility over a total of 500 thousand simulations. This value is reported as "Measured EV". We then used Formula 1 to compute another expected utility for innovating after $i$ rounds. These computed values are reported in the figure as "Predicted EV". Figure 1 shows the result of this experiment both distributions.

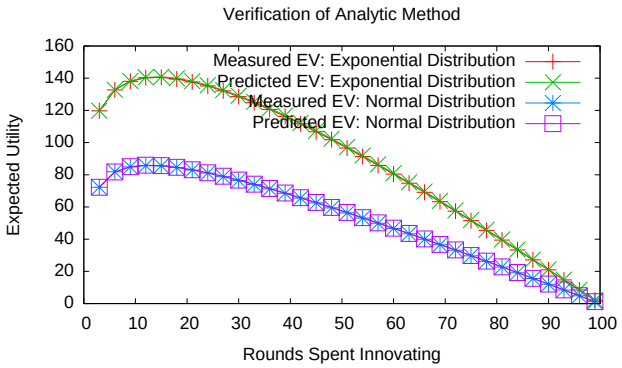


Figure 1: Comparing the utility predicted by equation 1 with the utility achieved in simulations when the underlying distribution is both normal and exponential. Notice that the measured values are almost exactly the same as the predicted values, such that the lines showing the measured expected value are overlaid by the predicted values.

We notice that in the figure, the predicted values match up exactly with the meausred values for both distributions. This we take as evidence that Formula 1 is correctly computed.

## Estimating an Optimal $k$

Thus far, we have only tried computing the optimal number of innovate moves when the set of actions $V$ is known in advance. For the Cultaptation game described earlier in this paper, as well as in many real world situations, this is an unreasonable assumption. Instead, an agent must estimate the optimal number of innovations given only what it has learned from previous moves in the game.

The *smart-innovator* strategy estimates $V$ based on the actions it has learned from previous innovate moves, and bases its decisions on that. Let $L = l_1, l_2, \ldots, l_k$ be the set of actions that the agent has already innovated by round $k$.

We could just assume that $V = L$, and decide whether to innovate further by computing $F(k+1, L)$. However, since $L$ has only $k$ values, innovating more than $k$ times would always appear pointless. It is not alway pointless to innovate more than $k$ times, so this would be suboptimal. What we can do instead is use $L$ to estimate $\pi$, the underlying distribution, and then generate an approximation (or several) of $V$ from the estimate of $\pi$.

For a given assumed family of underlying probability distribution (i.e. normal, uniform, exponential) over $V$, we can estimate the distribution parameters by measuring the mean and standard deviation of $L$. By sampling from this new distribution, we can create a new set of actions $V'$ that models the true distribution of $V$. We can then compute the inequality $(l-k)F(k, V') \geq (l-k-1)F(k+1, V')$ (Formula 3), and decide whether to continue innovating based on that.

No single $V'$ is guaranteed not to yield the same optimal number of innovates as $V$, but by generating multiple $V'$s and averaging the results we hope to reach a close approximation. Algorithm 1 shows how this procedure works under the assumption that $\pi$ is some normal distribution.

For situations when the family of the underlying distribution is not known *a priori*, a $\chi^2$ goodness-of-fit test can be used to compare several distributions (normal, uniform, etc), and determine which is best. From this estimation one can form a new distribution that may be representative of $V$, thereby eliminating the need to assume one particular family of probability distributions.

**Algorithm 1** Determine an estimate of the optimal number of innovates with observed values $L = \{l_1, \ldots, l_k\}$, and number of rounds $l$, under the assumption that $\pi$ is a normal distribution.

Smart Innovator($L$,$l$)

Let $m$ be the mean of $L$.
Let $\sigma$ be the standard deviation of $L$
Let $normal(m, \sigma)$ be a normal PDF with mean $m$ and standard deviation $\sigma$.
Let $LHS = 0$ {Left hand side of inequality}
Let $RHS = 0$ {Right hand side of inequality}
**for** 100 trials **do**
  Create $V' = \{v_1, \ldots, v_{n_b}\}$ where each $v_i$ is a random draw from $normal(m, \sigma)$.
  Let $LHS = LHS + (l-k)F(k, V')$
  Let $RHS = RHS + (l-k-1)F(k+1, V')$
**end for**
If $LHS < RHS$ innovate this round.
Otherwise exploit.

### Experimental Results for Smart-Innovator

To confirm that the *smart-innovator* strategy produces close to optimal results, we compared the average number of rounds spent innovating by agents performing this strategy, versus the known optimal number of innovate moves computed by the formula earlier in this paper. For simplicity, we assume the agent is given the family of distribution (normal, laplace, etc.), but not its parameters. For each of these distributions, the agent was required to innovate for at least 5 moves before estimating $V'$.

The results in Table 2 show the average performance for 50 trials of a 100 round SRI game. Notice that the *smart-innovator* strategy is able produce close to optimal results for a range of standard deviation values, even though it was not given those values in advance.

For a single run of the SRI game, the *smart-innovator* strategy may produce a higher utility than expected for the optimal number of innovates. This is because the expected value is a prediction of the average case, not a prediction of the best case. Given that actions are still drawn randomly from the distribution, it is entirely possible for a single run of *smart-innovator* or any other strategy to draw higher or lower utility than expected. This is exhibited in the standard deviation values column of Table 2.

### VRI Game

In this section we introduce and prove an algorithm for finding optimal strategies in the VRI game.

| Distribution | Mean | Stdev | Optimal $I_o$ Innovates | Average SI Innovates | Stdev of SI Innovates | Expected $I_o$ Utility | Average SI Utility | Stdev of SI Utility | Error Percentage |
|---|---|---|---|---|---|---|---|---|---|
| Uniform | 50 | 40 | 11 | 10 | 1.15 | 96.07 | 95.76 | 8.18 | 0.32% |
| | | 80 | 12 | 12 | 1.34 | 147.53 | 147.73 | 16.04 | 0.14% |
| | | 160 | 13 | 13 | 1.28 | 250.8 | 246.4 | 28.3 | 1.76% |
| Normal | 50 | 40 | 15 | 14 | 1.46 | 101.59 | 103.52 | 17.85 | 1.90% |
| | | 80 | 17 | 16 | 1.48 | 160.65 | 165.01 | 37.83 | 2.71% |
| | | 160 | 18 | 18 | 1.95 | 279.75 | 277.16 | 76.41 | 0.93% |
| Laplace | 50 | 40 | 19 | 17 | 1.93 | 106.5 | 111.17 | 44.26 | 4.39% |
| | | 80 | 21 | 21 | 1.59 | 172.7 | 161.86 | 45.98 | 6.28% |
| | | 160 | 23 | 23 | 1.36 | 306.1 | 304.94 | 161.61 | 0.38% |
| Exponential | 50 | 40 | 20 | 20 | 0.99 | 120.1 | 119.46 | 24.66 | 0.53% |
| | | 80 | 22 | 23 | 0.81 | 200.41 | 180.24 | 56.37 | 10.07% |
| | | 160 | 24 | 24 | 1.28 | 361.81 | 373.47 | 150.03 | 3.22% |

Table 2: Table showing the optimal number of innovates in the SRI game ($I_o$), compared with the average number of innovates performed by the *smart-innovator* strategy (SI). Note that regardless of the distribution or its parameters, the SI strategy either correctly estimates, or comes very close to correctly estimating the optimal number of innovate moves.

**Strategies to Consider** In the SRI game, it was immediately obvious that performing an exploit action before any innovate action was suboptimal. In the VRI game, however, one can imagine that an agent might want to insert an exploit action into the middle of a sequence of innovates to make sure that, in case it dies before it finishes innovating, it still obtains at least some utility. Thus we must consider a much wider variety of strategies in this version of the game.

Nevertheless, we can make one observation about strategies in this environment. Since the number of actions that can be learned by innovating is still fixed, and action utilities do not change, any optimal strategy must contain a finite number of innovates. We also know that after this strategy performs its last innovate, it exploits its best action on all subsequent moves.

**Which Strategy is Best?** Given a strategy, we know that it performs $k$ total innovates in its first $m$ turns ($m \geq k$), and afterwards exploits its best action forever. We can calculate the expected utility of the final series of exploits as follows:

$$\sum_{i=m+1}^{\infty} 0.98^i F(k,V) = \sum_{i=0}^{\infty} 0.98^i F(k,V) - \sum_{i=0}^{m} 0.98^i F(k,V)$$

Note that these terms are both geometric series, and simplify as follows:

$$\sum_{i=0}^{\infty} 0.98^i F(k,V) - \sum_{i=0}^{m} 0.98^i F(k,V) =$$
$$\frac{F(k,V)}{1-0.98} - F(k,V)\left(\frac{1-0.98^{m+1}}{1-0.98}\right) =$$
$$\frac{F(k,V)}{(1-0.98)}(1-(1-0.98^{m+1})) =$$
$$50F(k,V)(0.98^{m+1})$$

We will begin deriving optimal strategies by first considering the strategies that simply innovate $k$ times and then exploit forever (note that for these strategies, $m = k$). We will call the strategy that innovates $k$ times and then exploits forever $S_k$. Finding the best of these is as easy as computing

$$argmax_k \left(50F(k,V)(0.98^{k+1})\right)$$

Since $(0.98)^{k+1}$ asymptotically approaches 0 and $F(k,V)$ approaches the utility of the best action, we know that $50F(k,V)(0.98^{k+1})$ will increase until $F(k,V)$ stops growing faster than $(0.98)^{k+1}$, and after this point it will decrease forever, eventually approaching 0. Thus, we know that, if $k$ is an optimal value,

$$50F(k,V)(0.98)^{k+1} \geq$$
$$50F(k-1,V)(0.98)^{k} \geq$$
$$50F(k-2,V)(0.98)^{k-1} \geq \ldots$$

and

$$50F(k,V)(0.98)^{k+1} \geq$$
$$50F(k+1,V)(0.98)^{k+2} \geq$$
$$50F(k+2,V)(0.98)^{k+3} \geq \ldots$$

Simplifying these tells us that, if $i < k$,

$$F(i,V) \leq F(i+1,V)(0.98) \tag{4}$$

and if $i \geq k$,

$$F(i,V) \geq F(i+1,V)(0.98) \tag{5}$$

We will use these inequalities to prove that $S_k$ is the optimal strategy in terms of expected utility.

**Proposition 3** *Given a distribution $V$, let $k = argmax_k(50F(k,V)0.98^{k+1})$. Then $S_k$, the strategy that innovates $k$ times and then exploits forever, has the highest possible expected utility.*

**Proof:** Consider an arbitrary strategy, $X$. We will show that $S_k$ is the optimal strategy by making a series of changes to $X$, each of which makes $X$'s expected utility no worse, and show that the resulting strategy can be no better than $S_k$. As mentioned above, we know nothing about $X$ except that it makes $j$ total innovates, and its last innovate occurs on turn $m$. If $j > k$, let $n$ be the turn on which $X$ makes its $k$-th innovate, otherwise let $n = m$.

We will begin by examining only the first $n$ moves of $X$. We know that $X$ must make some number of innovates $j_1$, where $1 \leq j_1 \leq k$, in this period, and that the other $n - j_1$ moves are exploits. We will show that we can move all of the exploits to the end of this sequence without lowering the expected value of the overall strategy.

Consider the last exploit action in this sequence that is not already at the end (i.e. it is followed by at least one innovate action). The expected utility gained by this exploit is

$$F(a, V)(0.98)^b$$

for some $a < k$ and some $b$. The expected utility gained by swapping this exploit action with the innovate action that follows it is

$$F(a + 1, V)(0.98)^{b+1}$$

Note that, by inequality (4), this second quantity is at least as large as the first. Hence, we can perform the swap without lowering the expected utility of the strategy. Since we are only considering the first $n$ moves, and by the definition of $n$ there are at most $k$ innovates in this sequence, we can always use (4) to guarantee that swapping actions in this way will not lower our expected utility.

Therefore, we can move all of the exploit actions to the end without lowering the expected utility, creating a sequence of $j_1$ innovates followed by $n - j_1$ exploits. Let $X_1$ be the strategy that contains this sequence as its first $n$ moves, and is identical to $X$ thereafter. Note that the expected utility of $X_1$ is at least as high as $X$'s, and its values of $m$ and $j$ are the same.

Now, if $j > k$, we need to consider the moves $X_1$ makes from turn $n + 1$ to turn $m$.[1] We will show that the last innovate in this sequence can always be replaced with an exploit, without lowering the expected utility of the overall strategy. We know that the last innovate takes place on turn $m$, and the expected utility gained from the infinite series of exploits beginning on turn $m + 1$ is

$$50F(j, V)(0.98)^{m+1}$$

Consider replacing the innovate on turn $m$ with an exploit. Then the infinite series of exploits would begin on turn $m$, and its expected utility would be

$$50F(j - 1, V)(0.98)^m$$

Note that since $j > k$, (5) tells us that the second quantity is at least as large. Hence, eliminating the last innovate does not lower the expected utility of the overall strategy. Note that we can eliminate all the innovates that occur after turn $n$ in this way, resulting in a strategy that makes only exploit actions after turn $n$. Let $X_2$ be the strategy that is identical to $X_1$ for the first $n$ moves, and makes only exploit actions thereafter. Note that the expected utility of $X_2$ is at least as high as $X_1$'s.

Finally, note that at this point $X_2$ is the strategy that performs some number of innovates $j_1 \leq k$, followed by an infinite number of exploits. Thus, by the definition of $k$, we know that $S_k$ has expected utility at least as high as $X_2$'s, and hence it has expected utility at least as high as $X$'s. Since $X$ was an arbitrary strategy, $S_k$ has the highest possible expected utility.

[1] If $j \leq k$, let $X_2 = X_1$ and skip this step.

## Estimating Optimal $k$ for VRI

Similar to how the *smart-innovator* strategy could be used to estimate the optimal number of innovate moves in the SRI game, we can produce a strategy for the VRI game that accounts for the probability of death. The performance of the new strategy should have the same relation to the optimal number of innovates in the VRI model as the previous strategy did to the SRI model.

The overall structure of the strategy remains intact. We can still estimate $V'$ in the same manner as before. However, the inequality changes to $(0.98^k)F(k, V') \geq (0.98^{k+1})F(k + 1, V')$. An agent in the VRI game that wishes to employ this strategy must estimate a new $V'$ and evaluate the inequality for each step.

### Experimental Results for VRI Games

To confirm that the adapted *smart-innovator* produces close to optimal behavior, we compared the average number of rounds spent innovating by agents performing this strategy, versus the optimal number of innovate moves computed by the formula introduced in the VRI section of this paper, discounted by the probability that the agent would die before reaching that point.

The results in Table 3 show the average performance for 300 trials of a VRI game game with 0.02 probability of death. The utility values here represent the total sum of all exploit moves made by an agent, rather than the utility per round as in the Table 2. Also note that the high standard deviation for the number of innovate moves is influenced by agents dying before they are finished innovating. Again, the results show somewhat clearly that *smart-innovator* strategy is able produce close to optimal results for a range of standard deviation values.

## Related Work

This work addresses a simplified version of the Cultaptation social learning game. The authors are aware of no other work on this particular game, although there are related problems in anthropology, biology and economics, where effective social learning strategies and their origins are of particular interest.

The social learning competition attempts to shed light on an open question in behavioral and cultural evolution. Despite the obvious benefit of learning from someone else's work, several strong arguments have been made for why social learning isn't purely beneficial (Boyd and Richerson

| Distribution | Mean | Stdev | Optimal $I_o$ Innovates | Average SI Innovates | Stdev of SI Innovates | Expected $I_o$ Utility | Average SI Utility | Error Percentage |
|---|---|---|---|---|---|---|---|---|
| Uniform | 50 | 40 | 7 | 7 | 1.81 | 4423.58 | 4420.08 | 0.08 |
| | | 80 | 8 | 8 | 2.15 | 6330.39 | 6703.6 | 5.57 |
| | | 160 | 9 | 9 | 2.65 | 12425.22 | 11296.01 | 10 |
| Normal | 50 | 40 | 9 | 9 | 2.6 | 4397.94 | 4596.98 | 4.33 |
| | | 80 | 11 | 11 | 3.01 | 7562.08 | 7164.6 | 5.55 |
| | | 160 | 12 | 12 | 3.71 | 12548.92 | 12344.42 | 1.66 |
| Laplace | 50 | 40 | 11 | 11 | 3.18 | 4669.78 | 4531.05 | 3.06 |
| | | 80 | 12 | 13 | 4.15 | 6641.93 | 7109.36 | 6.57 |
| | | 160 | 14 | 14 | 4.71 | 11951.48 | 12328.23 | 3.06 |
| Exponential | 50 | 40 | 13 | 12 | 3.93 | 5217.24 | 5253.89 | 0.7 |
| | | 80 | 14 | 14 | 5.15 | 9171.82 | 8671.25 | 5.77 |
| | | 160 | 16 | 15 | 5.63 | 14252.42 | 15560.78 | 8.41 |

Table 3: Table showing the optimal number of innovates in the VRI game ($I_o$), compared with the average number of innovates performed by the *smart-innovator* strategy (SI).

1995; Rogers 1988). The answer to how to best learn in a social environment is seemingly non-trivial. Game theoretical approaches have been used to explore this subject, but there is still ongoing research in improving the models that are used (Henrich and McElreath 2003; Enquist, Ghirlanda, and Eriksson 2007).

(Laland 2004) discusses strategies for this problem in detail, and explores when it is appropriate to innovate or observe in a social learning situation. Indiscriminate observation is not always the best strategy, and there are indeed situations where innovation is appropriate. This is largely influenced by the conclusions of (Barnard and Sibly 1981), which reveals that if a large portion of the population is learning only socially, and there are few information producers, then the utility of social learning goes down.

In the social learning competition, an "observe" move introduces new actions into the repertoire of an agent, which the agent may optionally exploit. The relationship between this and social learning in animals is demonstrated by (Galef Jr. 1995), which differentiates social learning from mere imitation. In highly variable environments, socially learned information may not always be the most beneficial, yet animals that learn socially are still able to learn locally adaptive behavior. This is the result of having rewards or punishments associated with expressed behavior, similar to the payoff values in the social learning game, which can guide the actual behavior of an animal.

Discussing the means of social information transmission, (Nettle 2006) outlines the circumstances in which verbal communication is evolutionarily adaptive, and why few species have developed the ability to use language despite its apparent advantages. The model Nettle describes is similar to the Cultapatation game, where the cost of innovation versus observation can vary depending on the parameters of the system. The modeled population reaches an equilibrium at a point that includes both individual and social learning. The point of equilibrium is affected by the quality of observed information, and the rate of change of the environment.

Other work on similar games include (Giraldeau, Valone, and Templeton 2002), which outlines reasons why social information can become unreliable. Both biological factors, and the limitations of observation, can significantly degrade the quality of information learned socially. (Schlag 1998) explores rules that can be applied in a similar social learning environment which will increase the overall expected payoff of a population, by restricting how and when agents act on information learned through observation.

The structure and intentions of the Cultaptation Institute's social learning game are much akin to those laid out by Axelrod in his Prisoner's Dilemma tournament on the evolution of cooperation (Axelrod and Hamilton 1981). The analysis of the Prisoner's Dilemma has had substantial use in many different areas. It has been used as a model for arms races, evolutionary systems, and social systems. The Cultaptation social learning game has similar potential applications, yet currently no analysis. This work hopes to begin to fill that void.

## Conclusion

In this paper, we have provided mathematical models for simplified versions of the Cultaptation game, proven their correctness, and shown their effectiveness in experimentation. In these models the amount of resources one should spend innovating depends primarily on the underlying probability distribution. However, just like in the real world, this probability distribution is rarely known *a priori*. Armed only with knowledge of the family of underlying distributions, we are able to achieve very close to optimal results in experimentation.

As future work, we will examine the value of observation in the Cultaptation game. Initial analysis leads us to believe that one may posit the existence of a probability distribution over the observed move values, and then apply the framework detailed in this paper to "guess" the optimal number of observe moves. By combining this technique with the techniques given in this paper, we hope to given provable guarantees about the full Cultaptation game.

## Acknowledgments

This work was supported in part by AFOSR grants FA95500510298, FA95500610405, and FA95500610295, DARPA's Transfer Learning and Integrated Learning programs, and NSF grant IIS0412812. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the funders.

## References

Axelrod, R., and Hamilton, W. D. 1981. The evolution of cooperation. *Science* 211:1390.

Barnard, C., and Sibly, R. M. 1981. Producers and scroungers: A general model and its application to captive flocks of house sparrows. *Animal Behavior* 29:543–550.

Boyd, R., and Richerson, P. 1995. Why does culture increase human adaptability? *Ethology and Sociobiology* 16(2):125–143.

Boyd, R.; Enquist, M.; Eriksson, K.; Feldman, M.; and Laland, K. 2008. Cultaptation: Social learning tournament. `http://www.intercult.su.se/cultaptation`.

Enquist, M.; Ghirlanda, S.; and Eriksson, K. 2007. Critical social learning: A solution to rogers's paradox of nonadaptive culture. *American Anthropologist* 109(4):727734.

Galef Jr., B. G. 1995. Why behaviour patterns that animals learn socially are locally adaptive. *Animal Behavior* 49:1325–1334.

Giraldeau, L. A.; Valone, T. J.; and Templeton, J. J. 2002. Potential disadvantages of using socially acquired information. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 357(1427):1559–1566.

Henrich, J., and McElreath, R. 2003. The evolution of cultural evolution. *Evolutionary Anthropology* 12:123–135.

Laland, K. 2004. Social learning strategies. *Learning and Behavior* 32:4–14.

Nettle, D. 2006. Language: Costs and benefits of a specialised system for social information transmission. In Wells, J., and et al., eds., *Social Information Transmission and Human Biology*. London: Taylor and Francis. 137–152.

Rogers, A. R. 1988. Does biology constrain culture? *American Anthropologist* 90(4):819–831.

Schlag, K. 1998. Why imitate, and if so, how?, : A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory* 78:130–156.