Social Learning Between Groups: Imitation and the Role of Experience

Karl H. Schlag, Department of Economics, University of Vienna*

October 19, 2011

Abstract

We consider social learning between novices and experienced who face the same environment. Individuals repeatedly choose an action and observe choice and performance of someone belonging to the other group between making choices. While it is known who has more experience the environment itself is unknown. We search for learning rules that make all individuals within the same group better off in each enviornment. These rules involve imitation only. Novices imitate any more successful experienced and sometimes also the less successful. In particular, they should not always (or unconditionally) imitate their peers. Experienced should ignore what they observed about the novices.

1 Introduction

Cultural transmission is recognized as being important for the evolution of culture (Henrich and Gil-White, 2001). Cues help to identify groups who are likely to possess better than average information, in particular when information about success is less available or more noisy. Experience or age is a popular cue (Henrich and Broesch, 2011). To copy or imitate the behavior of those possessing such cues saves cost of learning and helps pass on knowledge acquired by others. Imitation seems the only natural behavior for anyone without information about the environment. Yet typically newcomers also have information of their own. They have their own limited experience in the new environment and possibly have some information about the

*karl.schlag@univie.ac.at

success of those possessing such a cue. We wish to investigate how own information should be combined with knowledge (or belief) that others are more experienced. The importance of "when to copy" is acknowledged in the field both among humans (Apesteguia et al. 2007) and among animals (Galef and Laland, 2005).

We are interested in social learning when individuals with different experience learn from each other. Young generations may be learning from the old, scholars from teachers, outsiders from insiders, and vice versa. Learning by imitation is natural and intuitive at first sight when novices learn from the more experienced. Yet should novices really imitate when they are successful and the experienced have realized a bad outcome? To imitate in such situations means to trust the experienced at the expense of throwing away own information that could be valuable in the comparison of actions. Similarly, what about the experienced? Should they learn from the novices? We use theory to shed light on these issues.

We consider a setting in which individuals have very little a priori information about their environment. They do not choose a learning rule that is best according to some specific prior. Instead they choose a rule that is able to learn which action is better in all environments. The special feature is that individuals observe someone else who has different experience and that individuals know whether the individual they observe is more or less experienced than they are. Outcomes are noisy and individuals wish to choose the action that yields the highest average payoff. Individual learning cannot be effective. As individuals have no memory, no pair of payoffs will reveal which of two actions is better. On the other hand, we show how social learning can be effective by aggregating and storing individual experience.

The case of learning among individuals with equal experience has been investigated by Schlag (1998). The *Proportional Imitation Rule* (PIR) has the following properties. When all use this rule, the frequency of the better action always increases, among all rules with this property the increase under PIR is strongest and PIR never recommends to imitate an individual with lower payoff. In contrast, under the rule "Imitate if Better", or under any rule that does prescribe to imitate the observed whenever switching actions, the frequency of the better action can decrease. The Proportional Imitation Rule is defined as follows. Never imitate someone else who has a lower payoff and imitate the action chosen by someone with a higher payoff with probability proportional to how much higher this payoff is.

In this paper we wish to analyze the case where individuals have different experience and where it is know who has more experience. More experience is identified with having a larger proportion who choose the better action. This can result from having faced the environment for a longer time. In this framework, experienced will be reluctant to imitate inexperienced, regardless of which payoffs are realized. This is because they cannot be sure that they have not already learned what is best in which case any switch would make them worse off. The aim to always weakly increase proportion using better action means for the experienced not to imitate the novices.

What about the novices? Given they observe a more experienced individual, one may think that the inexperienced will choose to use the rule *Always Imitate*, to always imitate the experienced regardless of the payoffs. This is indeed one type of behavior that will make the inexperienced always better off as they adapt the behavior of the experienced from the previous round. Such behavior is consistent with the observance of taboos.

Alternatively the inexperienced could use the *Reverse Proportional Imitation* rule (RPIR) as first defined in this paper. Swap roles and act as if you had the action and payoff of the observed and he or she had chosen your action and received your payoff. Then apply the Proportional Imitation Rule. This causes the individual to always imitate those that perform better, and to sometimes imitate those that perform worse, the probability of not switching being proportional to the difference in payoffs. Thus the probability of imitating an experienced with low payoff is only small if own payoff is high.

How do these two rules compare? The Reverse Proportional Imitation rule has good properties. It does best when there is no difference in experience and does not perform much worse otherwise. In contrast, Always Imitate is very specialized for the situation where difference in experience are large and performs otherwise not very well.

In Section 2 we introduce the model with two actions. In Section 3 we analyze the special case where payoffs are either 0 or 1. The more general case in which any payoffs between 0 and 1 are allowed is investigated in Section 4. In Section 5 we consider how the difference in experience changes over time. In Section 6 we conclude. The appendix contains some insights when there are more than two actions.

2 The Model

Individuals belonging to a large (essentially infinite) population repeatedly face a stationary decision problem. To simplify exposition we assume that there are only two actions, denoted by A and B, from which an individual chooses from. Choice of action $C \in \{A, B\}$ in round t ($t = 1, 2, ...$) yields a random payoff that belongs to a common bounded interval that can be renormalized with an affine mapping to $[0, 1]$. Payoffs are realized independently of other events. Let a and b be the corresponding expected payoffs of actions A and B. All individuals prefer the action that yields the higher expected payoff. Let P_C define the distribution of payoffs realized when choosing action C.

Individuals belong to one of two groups, called novice and experienced. Fix t. Let x_C^t and y_C^t denote the fraction of novices and experienced respectively choosing action C in round t, so $x_C^t, y_C^t \geq 0$ and $x_A^t + x_B^t = y_A^t + y_B^t = 1$. Individuals do not know anything about the distribution of payoffs except that payoffs belong to $[0, 1]$. They do not know the fraction of actions chosen in the population. They only know that the fraction of experienced choosing the better action is larger than the fraction of novices choosing this action. So if $a \geq b$ then $x_A^t \leq y_A^t$, if $a \leq b$ then $x_B^t \leq y_B^t$.

After choosing an action each individual observes the action chosen and payoff realized in the previous round of some randomly chosen individual belonging to the opposite group. Based on own experience in the last round and on this information the individual then chooses an action for the next round. It is assumed that individuals belonging to the same group use the same rule for revising their action.

Consider behavior of the novices. Their learning rule will be denoted by F and is given by a mapping from $\{A, B\} \times [0, 1] \times \{A, B\} \times [0, 1]$ to $[0, 1]$ which is the probability of changing actions. $F(C, u, D, v)$ will denote the probability of not choosing action C in round $t+1$ after choosing action C and receiving payoff u in round t and observing an experienced individual who chose action D an received payoff v in round t.

Let f_{CD} denote the expected probability of switching under rule F after choosing C and observing an experienced choosing D, calculated ex-ante to receiving receiving a payoff from action C. So

$$f_{CD} = \int_0^1 \int_0^1 F(C, u, D, v) \, dP_C(u) \, dP_D(v).$$

We derive the compute the change in play of action B and find

$$x_B^{t+1} - x_B^t = -\left(x_A^{t+1} - x_A^t\right) = x_A^t y_A^t f_{AA} + x_A^t y_B^t f_{AB} - x_B^t y_A^t f_{BA} - x_B^t y_B^t f_{BB}. \quad (1)$$

In the following we will search for rules F that are "improving". A rule is called *improving* if the average payoff within the same group increases between rounds when all use this rule. As there are only two actions, this means that the probability of switching to the better action is larger than the probability of switching to the worse action. Formally, F is improving if

$$\begin{aligned} x_B^{t+1} - x_B^t &\geq 0 \text{ if } b > a \text{ and } y_B^t \geq x_B^t, \\ x_B^{t+1} - x_B^t &\leq 0 \text{ if } b < a \text{ and } y_A^t \geq x_A^t. \end{aligned}$$

In particular, the above has to hold when $x_A^t = y_A^t$. With this additional restriction, it is as if individuals are learning from others within the same group. This case has been characterized by Schlag (1998) as follows. F is improving if and only if (i) F is *imitating*, so $F(C, u, C, v) = 0$ for all $C \in \{A, B\}$ and $u, v \in [0, 1]$, and (ii) there exists $\sigma \geq 0$ such that $F(A, u, B, v) - F(B, v, A, u) = \sigma(v - u)$ holds for all $u, v \in [0, 1]$.

Assume that F is improving. As F is imitating, $f_{AA} = f_{BB} = 0$ and $f_{AB} - f_{BA} = \sigma(b - a)$ and there exists $\sigma \geq 0$ such that, given $\delta^t = y_B^t - x_B^t$, (1) becomes

$$\begin{aligned} x_B^{t+1} - x_B^t &= x_A^t x_B^t \sigma(b - a) + \delta^t \left(x_A^t f_{AB} + x_B^t f_{BA} \right) \\ &= x_A^t x_B^t \sigma(b - a) + \delta^t x_A^t \sigma(b - a) + \delta^t f_{BA} \\ x_A^{t+1} - x_A^t &= x_A^t x_B^t \sigma(a - b) - \delta^t x_B^t \sigma(a - b) - \delta^t f_{AB} \\ &= x_A^t x_B^t \sigma(a - b) + \delta^t x_A^t \sigma(a - b) - \delta^t f_{BA} \end{aligned} \quad (2)$$

Note that δ^t measures the difference in experience, where $0 \leq \delta^t \leq x_A^t$ holds if B is the better action (so if $b > a$).

If there is no difference in experience, so if $\delta^t = 0$, then (2) is the same formula as obtained in Schlag (1998). The additional term with δ^t results from the higher likelihood of seeing the better action among the observed than it would be among the novices. More novices that have chosen A get the possibility to observe B and less novices who have already been choosing the better action B are distracted by observing performance of action A.

Following (2), as $\sigma, f_{AB,} f_{BA} \geq 0$ it follows that any improving rule for the case of a single group, so for $\delta^t = 0$, is also improving for novices learning from the experienced (so when $\delta^t \geq 0$).

Note that we can also gain insights from (2) for the experienced by considering $\delta^t \leq 0$ and assuming that x_C^t is the proportion of experienced using action C. Looking at the case where $b-a$ is small and $\delta^t < 0$, (2) implies that $\sigma = f_{BA} = 0$. It is easy to see, as shown below, that this implies that $F(A, u, B, v) = F(B, v, A, u) = 0$ for all $u, v \in [0, 1]$. This means that $F \equiv 0$, experienced ignore performance of the novices.

Proposition 1 *(i) A learning rule for the novices is improving if and only if it is improving when $\delta^t = 0$.*

(ii) A learning rule for the experienced is improving if and only if specifies to never adapt the action of a novice, regardless of how either of them performed.

Proof. All we need to show is (ii). Consider sequences of distributions $\left(P_A^{(n)}\right)_n$ and $\left(P_B^{(n)}\right)_n$ associated to actions A and B such that $a_n = EP_A^{(n)} < EP_B^{(n)} = b$ and $\lim_{n\to\infty}(b_n - a_n) = 0$. Let $f_{BA}^{(n)}$ be the associated expected probability of switching from B to A. Then following (3), $\lim_{n\to\infty} f_{BA}^{(n)} = 0$. In particular, this implies $f_{BA} = 0$ if $P_A = P_B$.

Assume that $F(B, v_0, A, u_0) > 0$ for some $u_0, v_0 \in [0, 1]$. Assume that $u_0 = v_0$. This gives a contradiction as it implies that $f_{BA} = F(B, u_0, A, u_0) > 0$ holds if A and B realize u_0 almost surely. Assume that $u_0 \neq v_0$. Assume that A and B are identical and realize values u_0 and v_0 equally likely. Then $f_{BA} \geq \frac{1}{4} F(B, v_0, A, u_0) > 0$ which again contradicts $f_{BA} = 0$. ■

3 A Simple Environment

We wish to select among the improving rules for novices. To build some intuition we investigate environments where payoffs lie in $\{0, 1\}$. We also restrict attention to rules where switching probability does not depend on the labelling of the actions, so where $F(C, u, D, v) = F(D, u, C, v)$ for all $u, v \in [0, 1]$ and $C, D \in \{A, B\}$. We argue at the end of this section why this can be done without loss of generality.

Let $s_{ij} = F(A, i, B, j) = F(B, i, A, j)$ for $i, j \in \{0, 1\}$. We drop the superscripts t. Note that a and b are now the probabilities of realizing payoff 1 with actions A and B respectively. Then

$$\sigma = s_{01} - s_{10}$$

and

$$f_{BA} = (1-b)\, a s_{01} + b(1-a)\, s_{10} + (1-a)(1-b)\, s_{00} + ab s_{11}$$

so

$$x_B^{t+1} - x_B^t = x_A^t x_B^t (s_{01} - s_{10})(b-a) + \delta^t x_A^t (s_{01} - s_{10})(b-a) \\ +\delta^t ((1-b) a s_{01} + b(1-a) s_{10} + (1-a)(1-b) s_{00} + ab s_{11}) . \tag{3}$$

As rules do not depend on labels we can assume that B is better than A, so $b \geq a$. It follows from (3) that the increase in fraction of novices choosing the best action is largest if $s_{00} = s_{01} = s_{11} = 1$. Note that there is no single best choice of s_{10} that does not depend on a, b, δ^t and x_A^t as increase in s_{10} raises f_{BA} which is good but decreases σ which is bad. We now wish to investigate the effects of s_{10}. Assuming $s_{00} = s_{01} = s_{11} = 1$ we obtain

$$\begin{aligned} x_B^{t+1} - x_B^t &= x_A^t x_B^t (1 - s_{10})(b-a) + \delta^t x_A^t (1 - s_{10})(b-a) \\ &\quad +\delta^t ((1-b) a + b(1-a) s_{10} + (1-a)(1-b) + ab) \qquad (4) \\ &= x_A^t x_B^t (1 - s_{10})(b-a) + \delta^t x_A^t (1 - s_{10})(b-a) + \delta^t (1 - b(1-a)(1-s_{10})) . \end{aligned}$$

So we are left with the question of how to choose a value for s_{10}. What should a novice do who has obtained the highest possible payoff and observes an expert who has obtained the lowest possible payoff?

Consider best choices for extreme cases of δ^t. If there is no difference in experience, so when $\delta^t = 0$, then it is best to choose $s_{10} = 0$, following (4), which also follows from (Schlag, 1998). If experience is maximal, so when all experienced are choosing the best action ($\delta^t = x_A^t$ and hence $y_B^t = 1$) then it is best for novices to imitate the experienced unconditionally, so best if $s_{10} = 1$.

Now consider for given value of s_{10} performance across all values of δ^t. After all, we assume that x and y are not known, hence that δ^t is not known. Performance is measured in terms of change in payoffs as given by $\left(x_B^{t+1} - x_B^t\right)(b-a)$. We consider loss in this performance that is generated from not knowing x and y (also formally called regret following Savage, 1951). For a given value of s_{10} this is the difference between best performance over all s_{10} when x and y are known and performance for the given value of s_{10}. As (4) is linear in s_{10}, the best performance under the benchmark where x and y are known is given by either $s_{10} = 0$ or $s_{10} = 1$. If best performance is obtained when $s_{10}=0$ then loss in performance is given by

$$\begin{aligned}
& \left(x_A^t x_B^t (b-a) + \delta^t x_A^t (b-a) + \delta^t (1 - b(1-a))\right)(b-a) - \left(x_B^{t+1} - x_B^t\right)(b-a) \\
= & \left(x_A^t x_B^t (b-a) + \delta^t x_A^t (b-a) - \delta^t b (1-a)\right)(b-a)\, s_{10} \\
\leq & \; x_A^t x_B^t (b-a)^2 s_{10} \\
\leq & \; \frac{1}{4} s_{10}.
\end{aligned}$$

If instead best performance is obtained when $s_{10}=1$ then loss in performance is given by

$$\begin{aligned}
& \delta^t (b-a) - \left(x_B^{t+1} - x_B^t\right)(b-a) \\
= & -\left(x_A^t x_B^t (b-a) + \delta^t x_A^t (b-a) - \delta^t b (1-a)\right)(b-a)(1-s_{10}) \\
\leq & -\left(x_A^t x_B^t (b-a) + x_A^t x_A^t (b-a) - x_A^t b (1-a)\right)(b-a)(1-s_{10}) \\
= & \; a(1-b)\, x_A^t (b-a)(1-s_{10}) \\
\leq & \; \frac{1}{27}(1-s_{10}).
\end{aligned}$$

Hence, loss in performance is given by

$$\max\left\{\frac{1}{4}s_{10}, \frac{1}{27}(1-s_{10})\right\}.$$

For instance, if $s_{10}=0$ then maximal loss is $1/27 \approx 0.037$ while if $s_{10}=1$ then maximal loss is $1/4$. So when comparing these two extremes we find $s_{10}=0$ in the sense that it attains smaller maximal loss. Maximal loss is minimized if one takes a convex combination between these two rules, by choose $s_{10}=4/31 \approx 0.129$ as solution to $\frac{1}{4}s_{10}=\frac{1}{27}(1-s_{10})$, which yields maximal loss equal to $1/31 \approx 0.032$ which is only slightly better than the value $1/27$ attained when $s_{10}=0$.

To summarize, we find the most appealing improving rule to satisfy $s_{00}=s_{01}=s_{11}=1$ and $s_{10}=0$. The simplicity of setting $s_{10}=0$ is chosen to favor the slightly better rule that requires $s_{10}=4/31$. In the next section we consider general payoffs.

We now comment on the above restriction to rules that do not depend on labels. It is intuitive that such rules will be chosen, given there is nothing known about the environment. The formal reasoning is as follows. Consider some more general improving rule F_1. Define a rule F by $F(C,x,D,y)=\frac{1}{2}F_1(C,x,D,y)+\frac{1}{2}F_1(D,x,C,y)$. Then F does not depend on labels and it is easily seen that F is improving. Moreover, the maximal regret of F is bounded above by the maximal regret of F_1 (maximal regret of a sum is less than or equal to the sum of the maximal regrets). Hence, according to our criteria for selecting rules we can restrict attention to those that do not depend on labels.

4 More General Payoffs

Above we selected a rule that for environments where payoffs are 0 or 1 by considering loss in performance from not knowing the parameters. This rule satisfies $s_{00} = s_{01} = s_{11} = 1$ and $s_{10} = 0$. Clearly, maximal loss in more general environments is weakly larger. In the following we present several rules that satisfy $s_{00} = s_{01} = s_{11} = 1$ and $s_{10} = 0$ where maximal loss is not larger, hence where maximal loss is equal to $1/27$.

The simplest rule with this property is imitating and satisfies $F(A,u,B,v) = 1-u(1-v)$. The following slightly more complicated rule maximizes switching among all improving rules that satisfy $\sigma = 1$. It is defined by $F(A,u,B,v) = 1$ if $v \geq u$ and $F(A,u,B,v) = 1-(u-v)$ if $u > v$. This rule is called the *Reverse Proportional Imitation rule* (RPIR) and can equivalently be defined by $F(A,u,B,v) = 1-\max\{u-v,0\}$. The name comes from the fact that it is as if one switches roles with the observed and chooses the strategy that the observed would choose if the observed would be using the Proportional Imitation Rule. The property of maximal switching can be derived as follows. Let F be improving rule with $\sigma = 1$. Since

$$F(A,u,B,v) - F(B,v,A,u) = v - u$$

we find that switching is maximal when both $F(A,u,B,v)$ and $F(B,v,A,u)$ are maximal subject to satisfying the above constraint. As $F(C,u,D,v) \leq 1$ for $C \neq D$ and $v \geq u$ with equality holding when F is the RPIR the claim is proven.

This leads us to the following result. First, for completeness, we define *maximal loss in performance* of a rule F (given x^t and y^t) among improving rules by

$$\max_{P_A,P_B}\left\{\max_{F_0 \text{ improving}}\left\{\left(x_B^{t+1}(F_0)-x_B^t\right)(b-a)\right\}-\left(x_B^{t+1}(F)-x_B^t\right)(b-a)\right\}$$

where $x_B^{t+1} = x_B^{t+1}(F_0)$ is given by (1).

Proposition 2 *(i) The Reverse Proportional Imitation Rule maximizes the probability of switching among all improving rules with $\sigma = 1$. Maximal loss in performance equals $1/27$.*

(ii) The rule F^* that is imitating and satisfies $F^*(A,u,B,v) = F^*(B,u,A,v) = 1-\frac{27}{31}\max\{u-v,0\}$ minimizes the maximal loss in performance among all improving rules. Maximal loss in performance equals $1/31$.

Proof. Part (i). Let F_{RPIIR} be the Reverse Proportional Imitation rule. Above we showed that that RPIR maximizes switching probability among improving rules with $\sigma = 1$. In other words, we have shown that if F is improving with $\sigma = 1$ then $F_{RPIIR}(C,u,D,v) \geq F(C,u,D,v)$ for all $C, D \in \{0,1\}$ and $u, v \in [0,1]$. Consider the rule F_1 given by $F_1(A,u,B,v) = 1-u(1-v)$. This rule satisfies $F_1(A,i,B,j) = 1$ if $(i,j) \in \{(0,0),(0,1),(1,1)\}$ and $F_1(A,1,B,0) = 0$. Following our analysis in Section 3 its maximal loss in performance equals $1/27$ if we limit attention to payoffs in $\{0,1\}$. However, the linearity of F_1 shows that maximal loss in performance is no larger if we consider payoffs in $[0,1]$ which shows the claim.

Part (ii). Consider the linear rule F_2 that is imitating and defined by $F_2(A,u,B,v) = 1-\frac{27}{31}u(1-v)$. Note that F_2 is improving with $\sigma = \frac{27}{31}$ and $F_2(A,i,B,j) = 1$ if $(i,j) \in \{(0,0),(0,1),(1,1)\}$ and $F_2(A,1,B,0) = 4/31$. Following Section 3 this rule attains the smallest maximal loss in performance among all improving rules if payoffs are in $\{0,1\}$. The linearity of F_2 implies that this statement also holds when payoffs belong to $[0,1]$. As the rule F^* given in (ii) is identical to F_2 when payoffs belong to $\{0,1\}$ but always switches more likely than F_2 it follows that maximal loss under F^* cannot be larger than under F_2 which completes the proof. ∎

5 Change in Experience

Above we investigated improving rules for individuals learning from others that have different experience in period t. The difference in experience was formally given by the condition that $(y_B^t - x_B^t)(b-a) \geq 0$. We found that experienced should not switch and argued that novices should use RPIR. We now investigate whether and how differences in experience my change over time. This will depend on whom individuals learn from, whether from the same group or from someone belong to the other group. Assume that individuals choose some improving rule with $\sigma = 1$ when learning from others in their same group. Given our above analysis, experienced do not wish to learn from novices. To fix ideas, let λ be the probability that novices learn from experienced, so they learn from other novices with probability $1-\lambda$. Then population dynamics are given by

$$\begin{aligned}
x_B^{t+1} - x_B^t &= (1-\lambda)\, x_A^t x_B^t (b-a) + \lambda \left(x_A^t x_B^t (b-a) + \delta^t x_A^t (b-a) + \delta^t f_{BA} \right) \\
&= x_A^t x_B^t (b-a) + \lambda \delta^t \left(x_A^t (b-a) + f_{BA} \right), \\
y_B^{t+1} - y_B^t &= y_A^t y_B^t (b-a)
\end{aligned}$$

where f_{BA} describes the switching behavior of novices learning from experienced.

In the following we investigate how difference in experience changes over time. We compute

$$\begin{aligned}
y_B^{t+1} - x_B^{t+1} &= y_B^t + y_A^t y_B^t (b-a) - \left(x_B^t + x_A^t x_B^t (b-a) + \lambda \delta^t \left(x_A^t (b-a) + f_{BA} \right) \right) \quad (5) \\
&= \delta^t + \left(x_A^t x_B^t + x_A^t \delta^t - \delta^t x_B^t - \left(\delta^t\right)^2 \right) (b-a) \\
&\quad - \left(x_A^t x_B^t (b-a) + \lambda \delta^t \left(x_A^t (b-a) + f_{BA} \right) \right) \\
&= \delta^t \left[1 + \left(x_A^t - x_B^t - \delta^t \right) (b-a) - \lambda f_{BA} - \lambda x_A^t (b-a) \right] \\
&\geq \delta^t (1-\lambda) \left(1 - x_B^t (b-a) \right) \geq 0
\end{aligned}$$

as $\delta^t \leq x_A^t$ and $f_{BA} \leq 1 - (b-a)$.

Note that even if novices only observe experienced, so if $\lambda = 1$, then $\delta^{t+1} > 0$ if $\delta^t > 0$ and $y_B^t < 1$. This proves the following.

Proposition 3 *Experienced in round t remain more experienced also in all later rounds. In particular, novices never completely catch up when not all experienced choose the better action in round t.*

We hasten to point out there is an alternative natural way to model who learns from whom. Assume that individuals see the performance each other. Then λ is the probability of that a novice observes an experienced and vice versa. As experienced choose not to change behavior in such situations we obtain

$$y_B^{t+1} - y_B^t = (1-\lambda)\, y_A^t y_B^t (b-a).$$

Learning among the experienced is slower than it was above. In particular, in this case novice can "over take" experienced. This happens given $x_B^t \in (0,1)$, $\lambda > 0$ and $b > a$ when δ^t is sufficiently small as then $\delta^{t+1} = y_B^{t+1} - x_B^{t+1} \approx -\lambda x_A^t x_B^t (b-a) < 0$. Of course δ^t will not be small for given t and given $b > a$ if novices only enter the population much later than the experienced. However, due to the slower rate of learning among the experienced, eventually individuals will not know whether those that were initially more experienced are still more experienced. A more detailed analysis of this model is beyond the scope of this paper. We expect that experienced will remain more experienced for a given number of periods after entry of novices if experienced have been facing the environment sufficiently long prior to this entry.

6 Conclusion

The concept of improving has proven to be useful for identifying and comparing social learning rules and establishing formalisms that lead to imitation. The methodology is easily adapted to specific settings as demonstrated in this paper (see also Alos-Ferrer and Schlag, 2009). New insights emerge. We compare. The Proportional Imitation Rule was selected by Schlag (1998) for individuals learning from others in the same group, due to its maximal change in average payoffs among those rules that are improving and for its minimal variance of behavior in finite populations. It has the natural property that individuals that performed worse are never imitated. In this paper we have considered learning across different experience and find that novices will also sometimes imitate those that perform worse, trying to overtake the experienced. The rule selected for the novices, the Reverse Proportional Imitation Rule, maximizes switching. In fact, it maximizes variance in any finite population among all improving rules with maximal change in average payoffs (this follows easily from Proposition 2i). Switching is maximal in order to adapt the behavior of the experienced as often as possible without eliminating the ability to learn from what is observed in case difference in experience is negligible.

Being more experienced is identified with a larger fraction choosing the better action. Difference in experience can result when some individuals enter the population later than others, all start in the same initial state and all those entering together use the same improving rule. Whether or not the difference in experience persists over time depends on how one models learning opportunities between making choices. The difference persists when experienced learn from experienced (see Proposition 3). It does not persist forever when individuals learn from each other. When novices learn from experienced, the experienced learn from novices and do not change behavior. This "waists" a learning opportunity of the experienced. Eventually, difference in experience need not hold (see comment after Proposition 3).

If individuals could choose whom to observe, then upon entry novices would choose to observe an experienced. If the novices would use the rule Always Imitate then this can result in large "regret" if the difference in experience is only small (see Section 3). However, if novices use Always Imitate in their first round then in the following round it would be commonly known that there is no longer any difference in experience after which it would not matter for either type whom to observe. All could minimize variance and switching by using PIR. On the other hand, when using RPIR, that attains lower regret, the difference in experience will persist forever.

The methodology used in this paper is not able to quantify the importance of innovation. One would have to allow behavior to be time dependent and assume that the initial state is known. These assumptions are plausible but the analysis quickly becomes intricate as demonstrated in (Schlag, 2004) for the case of learning from others in the same group.

References

[1] Alós-Ferrer, C. and K.H. Schlag (2009), Imitation and Social Learning, chapter in The Handbook of Rational and Social Choice, P. Anand, P. Pattanaik and C. Puppe (eds.), Oxford University Press.

[2] Apesteguia, J., Huck, S. and J. Oechssler (2007), Imitation - theory and experimental evidence, Journal of Economic Theory **136**, 217-235.

[3] Galef, B.G. and K.N. Laland (2005) Social Learning in Animals: Empirical Studies and Theoretical Models, BioScience **55(6)**, 489-499.

[4] Henrich, J., and J. Broesch (2011) On the nature of cultural transmission networks: Evidence from Fijian villages for adaptive learning biases. Philosophical Transactions of the Royal Society **366**, 1139-1148.

[5] Henrich, J. and Gil-White, F. (2001) The Evolution of Prestige: freely conferred status as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior **22**, 1-32.

[6] Savage, L. J. (1951), "The Theory of Statistical Decision," *J. Amer. Stat. Assoc.* **46(253)**, 55-67.

[7] Schlag, K.H. (2004), Imitation and Minimax Regret, Mimeo, http://homepage.univie.ac.at/karl.schlag/research/imitation/regretim47.pdf.

[8] Schlag, K. H. (1998), "Why Imitate, and if so, How? A Boundedly Rational Approach to Multi-Armed Bandits," *J. Econ. Theory* **78(1)**, 130-156.

Appendix: More than Two Actions

Consider the case where there are $n > 2$ actions denoted by $A_i, ..., A_n$, each yielding a random payoff in $[0, 1]$, the mean of action i denoted by a_i. To simplify notation we will write x_i^t instead of $x_{A_i}^t$, y_j^t instead of $y_{A_j}^t$ and f_{ij} instead of $f_{A_i A_j}$. Consider first the case where individuals learn from someone else with the same experience. A rule is called *improving* if average payoffs increase over time when used by all, so if the associated dynamics satisfy

$$\sum_{j=1}^{n} \left(z_j^{t+1} - z_j^t\right) a_j \geq 0$$

for all z^t and P_{A_i}, $i = 1, .., n$. Following Schlag (1998) such a rule is improving if and only if it is imitating and for each $i \neq j$ there exists $\sigma_{ij} = \sigma_{ji} \geq 0$ such that $F(A_i, u, A_j, v) - F(A_j, v, A_i, u) = \sigma_{ij}(v - u)$ for all $u, v \in [0, 1]$. Change in average payoffs is highest if $\sigma_{ij} = \sigma_{ji} = 1$, such rules are hence best among the improving rules. We will show that any such rule is also improving for novices learning from the experienced.

Now we define what it means to be more experienced which is motivated by the following scenario. Experienced start in the some initial state $\bar{y}$ at time 1 and learn from other experienced. Novices start later in the same initial state $\bar{y}$ and learn from other novices. All individuals of the same type use the same rule that is best among the improving rules, so $\sigma_{ij} = 1$ for $i \neq j$. Hence they both independently follow the dynamics described by

$$z_j^{s+1} - z_j^s = \left(a_j - \sum_{k=1}^{n} z_k^s a_k\right) z_j^s \text{ for } s = 1, 2, ...$$

with the same initial starting point. At time t novices start also observing experienced and vice versa. They know when they are observing an experienced and when they are observing a novice. As both novices and experienced started in the same initial state, and $\sigma_{ij} = 1$ holds for both, but experienced have faced the environment longer there exists s with $1 \leq s \leq t$ such that $y_j^t = z_j^t$ and $x_j^t = z_j^s$ for all j. We show what this implies for (x^t) and (y^t). As

$$\frac{z_j^{s+1}}{z_i^{s+1}} = \left(\frac{1 + a_j - \sum_{k=1}^n z_k^s a_k}{1 + a_i - \sum_{k=1}^n z_k^s a_k}\right) \left(\frac{z_j^s}{z_i^s}\right) \geq \frac{z_j^s}{z_i^s} \text{ if } a_j \geq a_i$$

it follows that

$$\frac{y_j^t}{y_i^t} \geq \frac{x_j^t}{x_i^t} \text{ if } a_j \geq a_i. \tag{6}$$

In the following we say that those behaving according to (y^t) are *more experienced* than those following (x^t) if (6) holds. Novices learning from experienced use an improving rule if average payoffs increase over time, so if

$$\sum_{j=1}^n \left(x_j^{t+1} - x_j^t\right) a_j \geq 0$$

for all $(P_{A_i})_i$ and all x^t and y^t that satisfy (6).

Consider now any improving rule for novices learning from other novices with $\sigma_{ij} = 1$. We now compute how this rule fairs when learning from experienced. Using the fact $f_{ij} - f_{ji} = a_j - a_i$ that we obtain

$$x_j^{t+1} - x_j^t = \sum_{i \neq j} \left(x_i^t y_j^t f_{ij} - x_j^t y_i^t f_{ji}\right) = \sum_{i \neq j} x_i x_j \left(a_j - a_i\right) + \sum_{i \neq j} \left(x_i \delta_j^t f_{ij} - x_j \delta_i^t f_{ji}\right).$$

We compute

$$\begin{aligned}
\sum_{j=1}^n \left(x_j^{t+1} - x_j^t\right) a_j &= \sum_{j=1}^n \sum_{i \neq j} x_i^t x_j^t \left(a_j - a_i\right) a_j + \sum_{j=1}^n \sum_{i \neq j} \left(x_i^t \delta_j^t f_{ij} - x_j^t \delta_i^t f_{ji}\right) a_j \\
&= \frac{1}{2} \sum_{i<j} x_i^t x_j^t \left(a_j - a_i\right)^2 + \sum_{i<j} \left(\left(x_i^t \delta_j^t f_{ij} - x_j^t \delta_i^t f_{ji}\right) \left(a_j - a_i\right)\right).
\end{aligned}$$

If $a_j \geq a_i$ then $x_i^t \delta_j^t \geq x_j^t \delta_i^t$ and hence

$$\left(x_i^t \delta_j^t f_{ij} - x_j^t \delta_i^t f_{ji}\right) \left(a_j - a_i\right) \geq \frac{1}{2} \left(x_i^t \delta_j^t + x_j^t \delta_i^t\right) \left(f_{ij} - f_{ji}\right) \left(a_j - a_i\right) = \frac{1}{2} \left(x_i^t \delta_j^t + x_j^t \delta_i^t\right) \left(a_j - a_i\right)^2.$$

If $a_j \leq a_i$ then $x_i^t \delta_j^t \leq x_j^t \delta_i^t$ and hence

$$\left(x_i^t \delta_j^t f_{ij} - x_j^t \delta_i^t f_{ji}\right) \left(a_j - a_i\right) \geq \frac{1}{2} \left(x_i^t \delta_j^t + x_j^t \delta_i^t\right) \left(f_{ij} - f_{ji}\right) \left(a_j - a_i\right) = \frac{1}{2} \left(x_i^t \delta_j^t + x_j^t \delta_i^t\right) \left(a_j - a_i\right)^2.$$

Consequently,

$$\sum_{j=1}^n \left(x_j^{t+1} - x_j^t\right) a_j \geq \frac{1}{2} \sum_{i<j} x_i^t x_j^t \left(a_j - a_i\right)^2 + \frac{1}{2} \sum_{i<j} \left(x_i^t \delta_j^t + x_j^t \delta_i^t\right) \left(a_j - a_i\right)^2 \geq 0.$$

Proposition 4 *Any improving rule for learning from someone with the same experience that satisfies* $\sigma_{ij} = 1$ *for* $i \neq j$ *is also improving for the novices. The only improving rule for the experienced is Never Switch.*