Imitation and Learning

Carlos Alós-Ferrer* and Karl H. Schlag†

January 30, 2007

1. Introduction

Imitation is one of the most common forms of human learning. Even if one abstracts from explicit evolutionary or cultural arguments, it is clear that the pervasiveness of imitation can only be explained if it often leads to desirable results. However, it is easy to conceive of situations where imitating the behavior of others is *not* a good idea, that is, imitation is a form of boundedly rational behavior. In this chapter we survey research which clarifies the circumstances in which imitation is desirable. Our approach is partly normative, albeit we do not rely on the standard Bayesian belief-based framework but rather view imitation as a belief-free behavioral rule.

There are many reasons why imitation may be a "good strategy". Some of them are:

(a) To free-ride on the superior information of others.

(b) To save calculation and decision-taking costs.

(c) To take advantage of being regarded similar to others.

(d) To aggregate heterogeneous information.

* Department of Economics, University of Konstanz, Box D-150, D-78457 Konstanz (Germany). Email: carlos.alos-ferrer@uni-konstanz.de

† Department of Economics and Business Administration, Universitat Pompeu i Fabra, Ramón Trías Fargas 25-27. 08005 Barcelona (Spain). Email: karl.schlag@upf.edu

(e) To provide a coordination device in games.

Sinclair (1990) refers to (a) as "information-cost saving." The argument is that, by imitating others, the imitator saves the information-gathering costs behind the observed decisions. Examples range from children learning[1] to research and development strategies.[2] Analogously, (b) refers to the fact that imitation economizes information-processing costs. A classical model building on this point is Conlisk (1980). Another nice example is Rogers (1989), who shows in an evolutionary model how an equilibrium proportion of imitators arises in a society when learning the true state is costly. Pingle and Day (1996) discuss experimental evidence showing that subjects use imitation (and other modes of "economizing behavior") in order to avoid decision costs.

Examples for (c) abound in ecology, where the phenomenon is called *mimicry*. In essence, an organism mimics the outward characteristics of another one in order to alter the behavior of a third organism (e.g. a predator). In line with this argument, Veblen (1899) describes lower social classes imitating higher ones through the adoption of fashion

[1] Experiments in psychology have consistently shown that children readily imitate behavior exhibited by an adult model, even in the absence of the model. See e.g. Bandura (1977). More generally, Bandura (1977, p. 20) states that "[L]earning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do. Fortunately, most human behavior is learned observationally through modeling: from observing others one forms an idea of how new behaviors are performed, and on later occasions this coded information serves as a guide for action."

[2] The protection of innovators from imitation by competitors is the most commonly mentioned justification for patents. Interestingly, Bessen and Maskin (1999) argue that "in a dynamic world, imitators can provide benefit to both the original innovator and to society as a whole."

trends. Clearly, a case can be made for imitation as signalling (see e.g. Cho and Kreps (1987)). In a pooling equilibrium, some agents send a specific signal in order to be mistaken for agents of a different type. In the model of Kreps and Wilson (1982) a flexible chain store that has the option to accommodate entry imitates the behavior of another chain store that can only act tough and fight entry.

The first three reasons we have just discussed are centered on the individual decision-maker. We will focus on the remaining two, which operate at a social level. There are two kind of examples in category (d). Banerjee (1992) and the subsequent herding literature show the presence of intrinsic information in observed choices can lead rational individuals to imitate others, even disregarding conflicting, private signals. A conceptually related example is Squintani and Välimäki (2002). Our objective here, though, is to provide a less-rational perspective on (d) and (e) by showing whether and how individuals who aim to increase payoffs would choose to imitate.

We remind the reader that our approach is not behavioral in the sense of selecting a behavioral rule motivated by empirical evidence and finding out its properties. Although we do not shy away from empirical or experimental evidence on the actual form and relevance of imitation, in this chapter we consider abstract rules of behavior and aim to find out whether some of them have "nicer" properties than others. Thus, our approach to bounded rationality allows for a certain sophistication of the agents. To illustrate, the two central results below are as follows. First, some imitation rules are better than other, e.g. rules which prescribe blind imitation of agents who perform better are dominated by a rule that incorporates the degree of better performance in the imitation behavior. That is, the reluctance to switch when the observed choices are only slightly better than the own might not be due to switching costs, but rather to the fact that the payoff-sensitivity of the imitation rule allows the population to learn the best option. Second, in a large class of strategic situations (games), the results of imitation present a systematic bias from "rational" outcomes, while still ensuring coordination.

2. Social Learning in Decision Problems

Consider the following model of social learning in decision problems. There is a population of decision-makers (DMs) who independently and repeatedly face the same decision in a sequence of rounds. The payoff obtained by choosing an action is random and drawn according to some unknown distribution, independently across DMs and rounds. Between choices, each DM receives information about the choices of other DMs and decides according to this information which action to choose next. We restrict attention to rules that specify what to choose next relying only on own experience and observations in the previous round.[3] Our objective is to identify simple rules which, when used by all DMs, induce increasing average payoffs over time.

More formally, a decision problem is characterized by a finite set of actions A and a payoff distribution P_i with finite mean π_i for each action $i \in A$. That is, P_i determines the random payoff generated when choosing action i, and π_i is the expected payoff of action i. We will make further assumptions on the distributions P_i below. Action j is called *best* if $j \in \arg\max \{\pi_i, i \in A\}$. Further, denote by $\Delta = \left\{(x_i)_{i \in A} \,\middle|\, \sum_{i \in A} x_i = 1,\, x_i \geq 0 \,\forall i \in A\right\}$

[3] This "Markovian" assumption is for simplicity. Bounded memory is simply a (realistic) way to constrain the analysis to simple rules.

the set of probability distributions (mixed actions) on A.

We consider a large population of DMs. Formally, the set of DMs might be either finite or countably infinite. Schlag (1998) explicitly works with a finite framework, while Schlag (1999) considers a countably infinite population. Essentially, the main results do not depend on the cardinality, but the interpretation of both the model and the considered properties does change (slightly). Thus, although the analysis is simpler in the countably infinite case, it is often convenient to keep a large-but-finite population intuition in mind.

All DMs face the same decision problem. While we are interested in repeated choice, for our analysis it is sufficient to consider two consecutive rounds. Let p_i be the proportion of individuals[4] who choose action i in the first round. Let $\bar{\pi} = \sum p_i \pi_i$ denote the average expected payoff in the population. Let p'_i denote the (expected) proportion of individuals that choose action i in the next (or second) round.

We consider an exogenous, arbitrary $p = (p_i)_{i \in A} \in \Delta$ and concentrate on learning between rounds. In the countably infinite case, the proportion of individuals in the population choosing action j is equal to the proportion of individuals choosing j that a given DM faces. In a finite population one has to distinguish in the latter case whether or not the observing individual is also choosing j.

2.1. Single Sampling and Improving Rules. Consider first the setting in which each individual observes one other individual between rounds. We assume *symmetric*

[4]For the countably infinite case, the p_i are computed as the limits of Cesàro averages, i.e. $p_i = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} g_i^k$ where the population is taken to be $\{1, 2, \dots\}$ and $g_i^k = 1$ if DM k chooses action i, zero otherwise. This entails the implicit assumption that the limit exists.

sampling: the probability that individual ω observes individual ω' is the same as vice versa. This assumption arises naturally from a finite-population intuition, when individuals are matched in pairs who then see each other. A particular case is that of *random sampling,* where the probability that a DM observes some other DM who chose action j is p_j.[5]

A learning rule F is a mapping which describes the DMs choice in the next round as a function of what she observed in the previous round. We limit attention to rules that do not depend on the identity of those observed, only on their choice and success. For instance, if each DM observes the choice and payoff of exactly one other DM, a learning rule can be described by a mixed action $F(i,x,j,z) \in \Delta$ where $F(i,x,j,z)_k$ is the probability of choosing action k in the next round after playing action i, getting payoff x and observing an individual choosing action j that received payoff z.

Clearly, the new proportions p'_i are a function of p, F, and the probabilities with which a DM observes other DMs. For our purposes, however, it will be enough to keep the symmetric sampling assumption in mind.

A learning rule F is *improving* if the average expected payoff in the population increases for all p when all DMs use F, i.e. if $\sum p'_i \pi_i \geq \sum p_i \pi_i$, for all possible payoff distributions. If there are only two actions, this is equivalent to requiring that $\pi_i > \pi_j$ implies $p'_i \geq p_i$.[6] We will provide a characterization of improving rules which, surprisingly,

[5] Note that "uniform sampling" is impossible within a countably infinite population. For details on random matching processes for countably infinite populations, see Boylan (1992), whose constructions can be used to show existence of the sampling procedures described here.

[6] Suppose that a DM replaces a random member of an existing population and observes the previous action and payoff of this individual before observing the performance of others. Then the improving

requires the rule to be a form of imitation. A rule F is *imitating* if a DM only switches to observed actions, formally, if $F(i,x,j,y)_k > 0$ implies $k \in \{i,j\}$. In particular, if the own and observed choice are identical then the DM will choose the same action again.

For imitating rules, it is clear that the change in the population expected payoff depends only on the expected net switching behavior between two DMs that see each other. $F(i,x,j,z)_j - F(j,z,i,x)_i$ specifies this net switching behavior when one individual who chose i and received x sees an individual who chose j and received z. The expected net switching behavior is then given by

$$F(i,j)_j - F(j,i)_i = \int\int \left(F(i,x,j,z)_j - F(j,z,i,x)_i\right) dP_i(x)\, dP_j(y). \tag{1}$$

Thus, an imitating rule F is improving if (and only if) the expected net switching behavior from i to j is positive whenever $\pi_j > \pi_i$. The only if part comes from the fact that one can always consider a state where only i, j are used.

There are of course many different imitating rules. We consider a first example. An imitating rule F is called *Imitate If Better (IIB)* if $F(i,x,j,z)_j = 1$ if $z > x$, $F(i,x,j,z)_j = 0$ otherwise ($j \neq i$). This rule, which simply prescribes to imitate the observed DM if her payoff is larger than the own payoff, is possibly the most intuitive of the imitation rules.

The following illustrative example is taken from Schlag (1998, p.142). Suppose the payoff distributions are restricted to be of the following form. Action i yields payoff $\pi_i + \varepsilon$ where π_i is deterministic but unknown and ε is random independent noise with mean 0. In statistical terminology, we consider distributions underlying each choice that

condition requires that the DM is expected to attain a larger payoff than the one she replaced. This interpretation is problematic in a countably infinite population, due to the inexistence of uniform distributions.

only differ according to a location parameter. We refer to this case as the *idiosyncratic noise* framework. We make no assumptions on the support of ε, in particular the set of achievable payoffs may be unbounded.

Schlag (1988) shows that Imitate If Better is improving under idiosyncratic noise. To see why, assume $\pi_j > \pi_i$. Consider first the case where two DMs who see each other have received the same payoff shock ε. Since $F(i, \pi_i + \varepsilon, j, \pi_j + \varepsilon)_j - F(j, \pi_j + \varepsilon, i, \pi_i + \varepsilon)_i \geq 0$ we find that net switching has the right sign. Now consider the case where the two DMs receive different shocks $\varepsilon_1, \varepsilon_2$. By symmetric sampling, and since payoff shocks follow the same distribution, this case has the exact same likelihood as the opposite case where each DM receives the shock that the other DM receives in the considered case. Thus we can just add these two cases for the computation in (1),

$$\begin{aligned}&\Big(F(i, \pi_i + \varepsilon_1, j, \pi_j + \varepsilon_2)_j - F(j, \pi_j + \varepsilon_1, i, \pi_i + \varepsilon_2)_i\Big) + \\ &\Big(F(i, \pi_i + \varepsilon_2, j, \pi_j + \varepsilon_1)_j - F(j, \pi_j + \varepsilon_2, i, \pi_i + \varepsilon_1)_i\Big)\end{aligned} \tag{2}$$

and the conclusion follows because both terms in brackets are non negative for IIB.

The case of idiosyncratic noise is of course particular; however, it provides us with a first intuition. We now consider an alternative, more reasonable restriction on the payoffs. The distributions P_i are allowed to be qualitatively different, but we assume that they have support on a common, nondegenerate, bounded interval. Without loss of generality, the payoff interval can be assumed to be $[0, 1]$ after an affine transformation. The explicit assumption is that payoffs are known to be contained in $[0, 1]$. We refer to this case as the *bounded payoffs* framework.

We introduce now some other rules, whose formal definitions are adapted to the bounded payoffs case (IIB is of course also well-defined). An imitating rule F is called

- *Proportional Imitation Rule (PIR)* if $F(i,x,j,z)_j = \max\{z-x,0\}$ $(j \neq i)$,
- *Proportional Observation Rule (POR)* if $F(i,x,j,z)_j = z$ $(j \neq i)$, and
- *Proportional Reviewing Rule (PRR)* if $F(i,x,j,z)_j = 1-x$ $(j \neq i)$.

PIR seems a plausible rule, making imitation depend on how much better the observed DM performed. POR has a smaller intuitive appeal, because it ignores the DM's own payoff. PRR is based on an aspiration-satisfaction model. The DM is assumed to be satisfied with her action with probability equal to the realized payoff, e.g. she draws an aspiration level from a uniform distribution on $[0,1]$ at the beginning of each round. A satisfied DM keeps her action while an unhappy one imitates the action used by the observed individual. Schlag (1998) shows:

Proposition 1. *Under bounded payoffs, a rule F is improving if and only if it is imitating and for each $i,j \in A$ such that $i \neq j$ there exists $\sigma_{ij} \in [0,1]$ with $\sigma_{ij} = \sigma_{ji}$ and $F(i,x,j,z)_j - F(j,z,i,x)_i = \sigma_{ij} \cdot (z-x)$ for all $x,z \in [0,1]$.*

The "if" statement is easily verified. The intuition behind the "only if" statement is as follows. Assume F is improving. We first show why F is imitating so why $F(i,x,j,z)_k = 0$ for all other actions $k \neq i,j$. The improving condition must hold in all states so assume that all DMs are choosing action i or j. Now consider an individual who chose i and received x and who observed j receiving z. It could be that all other actions $k \neq i,j$

always generate payoff 0 while $\pi_i = \pi_j > 0$. This means that everyone is currently choosing a best action and if a non-negligible set of individuals switches to any other action then the average payoff decreases in expected terms. Hence our hypothetical DM should not switch.

Given that an improving rule is imitating, as already observed the relevant quantities for the change in population expected payoffs are the expected net switching rates given by (1). Again, if $\pi_j > \pi_i$ then the expected net switching behavior from i to j has to be positive. Of course neither π_i nor π_j are observable. But note that the sign of $\pi_j - \pi_i$ can change when the payoffs that occur with positive probability under P change slightly (recall that the improving condition requires an improvement in population payoffs for all possible payoff distributions). Thus the net switching behavior needs to be sensitive to small changes in the received payoffs. It turns out that this sensitivity has to be linear in order for average switching behavior based on realized payoffs to reflect differences in expected payoffs. The linearity is required by the need to translate behavior based on observations into behavior in terms of expected payoffs.

Given Proposition 1, it is easy to see that average payoffs increase more rapidly for larger σ_{ij}. The upper bound of 1 on σ_{ij} is due to the restriction that F describes probabilities. Note that if payoffs were unbounded or at least if there were no known bound then the above argument would still hold; however the restriction that F describes a probability would imply that $\sigma_{ij} = 0$ for all $i \neq j$. Hence, improving rules under general unbounded payoffs generate no aggregate learning as average payoffs do not change over time (due to $\sigma_{ij} = 0$). As we have seen, this is in turn no longer true if additional constraints are placed on the set of possible payoff distributions.

A direct application of Proposition 1 shows that, for bounded payoffs, PIR, PRR and POR are improving while IIB is not. For example, let $|A| = \{1, 2\}$ and consider P such that $P_1(1) = 1 - P_1(0) = \lambda$ and $P_2(x) = 1$ for some given $x \in (0, 1)$. Then $F(2, 1)_1 = \lambda$ and $F(1, 2)_2 = 1 - \lambda$ so $F(2, 1)_1 - F(1, 2)_2 = 2\lambda - 1$. Consequently, $x < \lambda < 1/2$ and $p_1 \in (0, 1)$ implies $\pi_1 > \pi_2$ but $\bar{\pi}' < \bar{\pi}$.

Higher values of σ_{ij} induce a larger change in average payoffs. Thus it is natural to select among the improving rules those with maximal σ_{ij}. Note that PIR, PRR and POR satisfy $\sigma_{ij} = 1$ for all $i \neq j$. Thus all three can be regarded among the "best" improving rules. Each of these three rules can be uniquely characterized among the improving rules with $\sigma_{ij} = 1$ for all $i \neq j$ (see Schlag 1998). PIR is the unique such rule where the DM never switches to an action that realized a lower payoff. This property is very intuitive although it is not necessary for achieving the improving property. However it does lead to the lowest variance when the population is finite. PRR is the unique improving rule with $\sigma = 1$ that does not depend on the payoff of the observed individual. Hence it can be applied when individuals have less information available, i.e. when they observe the action and not the payoff of someone else. Last, POR is the unique such rule that does not depend on own payoff received. The fact that own payoff is not necessary for maximal increase in average payoffs among the improving rules is here an interesting finding that adds insights to the underlying structure of the problem.

Of course, Proposition 1 depends on the assumption of bounded payoffs. As we have illustrated, IIB is not payoff-improving under bounded payoffs but it is under idiosyncratic noise. Additionally, the desirability of a rule depends on the criterion used. Proposition 1 focuses on the improving property. Oyarzun and Ruf (2007) study an alternative property. A learning rule is *first-order monotone* if the proportion of DMs who play actions with first-order stochastic dominant payoff distributions is increasing (in expected terms) in any environment. They show that both IIB and PIR have this property. Actually, all improving rules also have this property. This follows directly from Lemma 1 in Oyarzun and Ruf (2007), which establishes that first-order monotonicity can be verified from two properties (i) imitating and (ii) positive net switching to first order stochastically dominant actions. Oyarzun and Ruf (2007) go on to show that no "best rule" can be identified within this class of rules; the intuition is that the class is too large, or in other words, that the concept of first-order monotonicity is too weak.

2.2. Population Dynamics and Global Learning. Consider a countably infinite population and suppose all DMs use the same improving rule with $\sigma_{ij} = 1$ for $i \neq j$, e.g. one of the three improving rules mentioned above: PIR, PRR, POR. Further, assume random sampling. Direct computation from Proposition 1 (see Schlag 1998) shows that

$$\begin{aligned} p_i' &= p_i + (\pi_i - \bar{\pi}) \cdot p_i \text{ and} \\ \bar{\pi}' &= \bar{\pi} + \sum_i (\pi_i - \bar{\pi})^2 \cdot p_i = \bar{\pi} + \frac{1}{2} \sum_{i,j} (\pi_i - \pi_j)^2 \, p_i p_j. \end{aligned} \tag{3}$$

If the learning procedure is iterated, one obtains a dynamical system in discrete time. Its analysis shows that, for any interior initial condition (i.e. if $p >> 0$), in the long run the proportion of individuals choosing a best action tends to one.

We see in (3) that the expected increase in choice of action i is proportional to the frequency of those currently choosing action i and to the difference between the expected payoff of action i and the average expected payoff among all individuals. The expected dynamic between learning rounds is thus a discrete version of what is known as the *(standard) replicator dynamics* (Taylor, 1979, Weibull, 1995).

For a finite population, the dynamics becomes stochastic as one cannot implicitly invoke a law of large numbers. In (3), p' has to be replaced by its expectation Ep' and the growth rate has to be multiplied by $N/(N-1)$ where N is the number of individuals. Schlag (1998) provides a finite-horizon approximation result relating the dynamics for large but finite population and (3). The explicit analysis of the long-run properties of the finite-population dynamics has not yet been undertaken.

Improving captures local learning much in the spirit of evolutionary game theory where payoff monotone selection dynamics are considered as the relevant class (cf. Weibull, 1995). Selection dynamics refers to the fact that actions not chosen presently will also not be chosen in the future. Imitating rules lead to such dynamics. When there are only two actions, monotonicity is equivalent to requiring that the proportion of those playing the best action is strictly increasing over time whenever both actions are initially present. Thus an improving rule generates a particular payoff monotone dynamics. This is particularly clear for PIR, PRR, and POR in view of (3).

One may alternatively be interested in global learning in terms of the ability to learn which action is best in the long run. Say that a rule is a *global learning rule* if for any initial state all DMs choose a best action in the long run. Say that a rule is a global *interior learning rule* if for any initial state in which each action is present all DMs choose a best action in the long run. The following result is from Schlag (1998).

Proposition 2. *Assume random sampling, countably infinite population, and consider only two actions. A rule is a global interior learning rule if and only if it is improving with* $\sigma_{12} > 0$. *There is no global learning rule.*

A cursory examination of (3) shows that PIR, PRR and POR are global interior learning rules. For an arbitrary global interior learning rule, note that the states in which all DMs use the same action have to be absorbing in order to enable them to be possible long run outcomes. In the interior, as there are only two actions, global learning requires that the number of those using the best action increases strictly. This is slightly stronger than the improving condition, hence the additional requirement that $\sigma_{12} > 0$.

The fact that no improving rule is a global learning rule is due to the imitating property. When all DMs start out choosing a bad action then they will keep doing so forever, as there is no information about the alternative action. However global learning requires some DM to switch if the action is not a best action.[7]

If the population is finite then there is also no global interior learning rule. Any state in which all DMs choose the same action is absorbing and can typically be reached with positive probability. Thus, not all DMs will be choosing a best action in the long run. However, this is simply due to the fact that the definition of global learning is not well-suited for explicitly stochastic frameworks. It would be reasonable to consider rare exogenous mistakes or mutations and then to investigate the long run when these

[7] The argument relies on the fact that the rule is stationary, i.e. it can not condition on the period.

mistakes become small, as in Kandori et al. (1993) and Young (1993) (see also Section 3). Alternatively, one could allow for non-stationary rules where DMs experiment with unobserved actions. Letting this experimentation vanish over time at the appropriate rate one can achieve global learning. The proof technique involves stochastic approximation (e.g. see Fudenberg and Levine, 1998, appendix of chapter 4).

We briefly comment on a scenario involving global learning with *local interactions.* Suppose that each individual is indexed by an integer and that learning only occurs by randomly observing one of the two individuals indexed with adjacent integers. Consider two actions only, fix $t \in \mathbb{Z}$ and assume that all individuals indexed with an integer less or equal to t choose action 1 while those with an index strictly higher than t choose action 2. Dynamics are particularly simple under PIR. Either individual t switches or individual $t+1$ switches, thus maintaining over time a divide in the population between choice of the two actions. Since $F(1,2)_2 - F(2,1)_1 = \pi_2 - \pi_1$ we obtain that $\pi_2 > \pi_1$ implies $F(1,2)_2 > F(2,1)_1$. Thus, the change in the position of the border between the two regions is governed by a random walk, hence in the long run all will choose a best action. PIR is a global interior learning rule. A more general analysis for more actions or more complex learning neighborhoods is not yet available.

2.3. Selection of Rules. Up to now we have assumed that all DMs use the same rule and then evaluated performance according to change in average expected payoffs within the population. Here we briefly discuss whether this objective makes sense from an individual perspective. We present three scenarios. First, there is a *global individual* *learning* motivation. Suppose each DM wishes to choose the best action among those present in the long run. Then it is a Nash equilibrium that each individual chooses the same global learning rule. Second, there is an *individual improving* motivation. Assume that a DM enters a large, finite population, by randomly replacing one of its members where the entering DM is able to observe the previous action and payoff of the member replaced. If the entering DM wishes to guarantee an increase in expected payoffs in the next round as compared to the status quo of choosing the action of the member replaced then the DM can choose an improving rule regardless of what others do. The role of averaging over the entire population is now played by the fact that replacement is random in the sense that each member is replaced equally likely and the improving condition is evaluated ex ante before entry.

Last, we briefly discuss a setting with *evolution of rules*, in which rules are selected according to their performance. A main difficulty with such a setting is that there is no selection pressure once two rules choose the same action provided selection is based on performance only. In particular this means that there is no evolutionarily stable rule.

Björnerstedt and Schlag (1996) investigate a simple setting with two actions in which only two rules are present at the same time. An individual stays in the population with probability proportional to the last payoff achieved and exits otherwise. Individuals entering into the population sample some existing individual at random and adopt rule and action of that individual. Neutral stability is investigated, defined as the ability to remain in an arbitrarily large fraction provided sufficiently many are initially present. It turns out that a rule is neutrally stable in all decision problems if and only if it is strictly improving (improving with $\sigma_{12} > 0$). Note that this result holds regardless of which action is being played among those using the strictly improving rule. Even if some are choosing the better action it can happen, when the alternative rule stubbornly chooses the worse action that in the long run all choose the worst action. Rules that sometimes experiment do not sustain the majority in environments where the action they experiment with is a bad action. Rules that are imitative but not strictly improving lose the majority against alternative rules that are not imitative.

A general analysis with multiple rules has not yet been undertaken.

2.4. Learning from Frequencies. We now consider a model in which DMs do not observe the individual behavior of others but instead know the population frequencies of each action. That is, each individual only observes the population state p. A rule F becomes a function of p, i.e. $F(i,x,p)_j$ is the probability of choosing j after choosing i and obtaining payoff x when the vector of population frequencies is equal to p.

Notice that PRR can be used to create an improving rule even if there is no explicit sampling of others. Each DM can simply apply PRR to determine whether to keep the previous action or to randomly choose an action used by the others, selecting action j with probability p_j. Formally, $F(i,x,p)_i = x + (1-x)p_i$ and $F(i,x,p)_j = (1-x)\,p_j$ if $j \neq i$. The results of this rule as indistinguishable from the situation where there is sampling and then PRR is applied. Thus the resulting dynamic is the one described in (3).

This can be improved upon, though. We present an improving rule that outperforms any rule under single sampling in terms of change in average payoffs when there are two actions. The aim is to eliminate the inefficiencies created under single sampling when two individuals choosing the same action observe each other. The idea is to act as if there is some mediator that matches everyone in pairs such that the number of individuals seeing the same action is minimal. Suppose $p_1 \geq p_2 > 0$. All individuals choosing action 2 are matched with an individual choosing action 1. There are a total of p_2 pairs of individuals matched with a different action, thus an individual choosing action 1 is matched with one choosing action 2 with probability p_2/p_1. After being matched, PRR is used. Thus, individuals using action 2 who are in the minority switch if not happy. Individuals using action 1 that are not happy switch with probability p_2/p_1. Formally, F is imitating and, for $p >> 0$, $F(2, y, p)_1 = 1 - y$ and $F(1, x, p)_2 = (p_2/p_1)(1 - x)$. Consequently,

$$p_1' = p_2 (1 - \pi_2) + p_1 \left(1 - \frac{p_2}{p_1}(1 - \pi_1)\right)$$

and hence

$$p_i' = p_i + \frac{1}{\max\{p_1, p_2\}} p_i (\pi_i - \bar{\pi}). \tag{4}$$

In particular, F is improving and yields a strictly higher increase in average payoffs than under single sampling whenever not all actions present yield the same expected payoff. For instance when there are two actions then its growth rate is up to twice that of single sampling. However its advantage vanishes as the proportion of individuals choosing a best action tends to 1.

2.5. Absolute Expediency. There is a close similarity between learning from frequencies and the model of learning of Börgers et al. (2004), which centers on a refinement of the improving concept. A learning rule F is *absolutely expedient* (Lakshmivarahan and Thathachar, 1973) if it is improving and $\sum p_i' \pi_i > \sum p_i \pi_i$ unless $\pi_i = \pi_j$ for all i, j. That is, the DM's expected payoff increases strictly in expected terms from one round to the next for every decision problem, unless all actions have the same expected payoff. Note that, as a corollary of Proposition 1, when exactly one other DM is observed, absolutely expedient rules are those where, additionally, $\sigma_{ij} > 0$ for all distinct $i, j \in A$.

Börgers et al. (2004) consider a single individual who updates a mixed action (i.e. a distribution over actions) after realizing a pure action based on it. They search for a rule that depends on the mixed action, the pure action realized and the payoff received that is absolutely expedient. Thus, expected payoff in the next round is supposed to be at least as large as the expected payoff in the current round, strictly larger unless all actions yield the same expected payoff. For two actions Börgers et al. (2004) show that a rule is absolutely expedient if and only if there exist $B_{ii} > 0$ and A_{ii} such that $F(i, x, p)_i = p_i + p_j (A_{ii} + B_{ii} x)$. They show that there is a best (or dominant) absolutely expedient rule that achieves higher expected payoffs than any other. This rule is given by $B_{ii} = 1/\max\{p_1, p_2\}$ and $A_{ii} = -\min\{p_1, p_2\} / \max\{p_1, p_2\}$. A computation yields

$$E(p_1' | p) = p_1 + \frac{1}{\max\{p_1, p_2\}} p_1 (\pi_1 - \bar{\pi}).$$

Note that the expected change is equal to the change under our rule shown in (4). This is no surprise as any rule that depends on frequencies can be interpreted as a rule for individual learning and vice versa. Note that Börgers et al. (2004) also show that there is no best rule when there are more than two actions.

Since Börgers et al (2004) work with rules whose unique input is the own received payoff, most imitating rules are excluded. Morales (2002) considers rules where also the action and payoff of another individual are observed, as in Schlag (1998), although as Börgers et al he considers mixed-action rules. He does not consider general learning rules but rather focuses directly on imitation rules, in the sense that the probabilities attached to non-observed actions are not updated.[8] Still, the main result of Morales (2002) is in line with Schlag (1998). An imitating rule is absolutely expedient if and only if it verifies two properties. The first one, unbiasedness, specifies that the expected movement induced by the rule is zero whenever all actions yield the same expected payoff. The second one, positivity, implies that the probability attached to the action chosen by the DM is reduced if the received payoff is smaller than the one of the observed individual, and increased otherwise. Specifically, as in Proposition 1, "the key feature is proportional imitation, meaning that the change in the probability attached to the played strategy is proportional to the difference between the received and the sampled payoff." (Morales 2002, p. 476).

Morales (2002) identifies the "best" absolutely expedient imitating rule, which is such that the change in the probability of the own action is indeed proportional to the payoff difference, but the proportionality factor is the minimum of the probabilities of the own action and the sampled one. Absolute expediency, though, does not imply imitation. In fact, there is a non-imitating, absolutely expedient rule which, in this framework, is able

[8] Morales (2005) shows that no pure-action imitation rule can lead a DM towards optimality for given, fixed population behavior. Recall that in Proposition 1 the rule is employed by the population as a whole.

to outperform the best imitating one. This rule is the classical reinforcement learning model of Cross (1973). The reason behind this result is that, in a framework where the DM plays mixed actions, an imitating rule is not allowed to update the choice probabilities in the event that the sampled action is the same as the chosen one, while Cross' rule uses the observed payoffs in order to update the choice probabilities.

2.6. Multiple Sampling. Consider now the case where each DM observes $M \geq 2$ other DMs. We define the following imitation rules. *Imitate the Best (IB)* specifies to imitate the action chosen by the observed individual who was most successful. *Imitate the Best Average (IBA)* specifies to consider the average payoff achieved by each action sampled and then to choose the action that achieved the highest average. For both rules, if there is a tie among several best-performing actions, the DM randomly selects among them, except if the own action is one of the best-performing. In the latter case, the DM does not switch.

The *Sequential Proportional Observation Rule (SPOR)* is the imitation rule which sequentially applies POR as follows. Randomly select one individual among those observed including those that chose the same action as you did. Imitate her action with probability equal to her payoff. With the complementary probability, randomly select another individual among those observed and imitate her action with probability equal to her payoff. Otherwise, select a new individual. That is, randomly select without replacement among the individuals observed, imitate their action with a probability equal to the payoff and stop once another DM is imitated. Do not change action if the last selected individual is not imitated.[9]

For $M = 1$, IBA and IB reduce to Imitate if Better, while SPOR reduces to POR Note also that when only two payoffs are possible, say 0 and 1, then SPOR and IB yield equivalent behavior. The only difference concerns who is imitated if there is a tie but this does not influence the overall expected change in play.

Consider first idiosyncratic noise. Schlag (1996) shows that IB is improving but IBA is not. To provide some intuition for this result, consider the particular case where there are only two actions, 1 and 2, which yield the same expected payoff. Further, suppose that noise is symmetric and only takes two values. Suppose M is large and consider a group of M DMs where one of them chooses action 1 and all the other $M-1$ choose action 2. We will investigate net-switching between the two actions within this group. If a rule is improving then net switching must be equal to 0. Suppose all DMs use IBA. The average payoff of action 2 will most likely be close to its mean if M is large. Action 1 achieves the highest payoff with probability equal to $1/2$ in which case all individuals switch to action 1 unless of course all individuals choosing action 2 also attained the highest payoff. On the other hand action 1 achieves the lowest payoff with probability $1/2$ in which case the individual who chose action 1 switches. Thus, it is approximately equally likely that all DMs end up choosing action 1 or action 2. It is important to take into account the number of individuals switching. When all DMs switch to action 1 there are $M-1$ switches, when all switch to action 2 there is just a single switch. This imbalance causes IBA not to be

[9] The basic definition of SPOR is due to Schlag (1999) and Hofbauer and Schlag (2000), except that they did not include the stopping rule.

improving. An explicit example is easily constructed (see Schlag, 1996, Example 11). In fact IBA has the tendency to equalize the proportions of the two actions when the difference in expected payoffs is small (Schlag, 1996, Theorem 12).

Now suppose all DMs use IB. Since both actions achieve the same expected payoff and only two payoffs are possible, then the net switching is zero because IB coincides with SPOR with only two possible payoffs. It is equally likely for each individual to achieve the highest payoffs. In $M-1$ cases some individual choosing action 2 achieves the highest payoff in which case only one individual switches while in one case it is the individual who chose action 1 and $M-1$ individuals switch. The latter becomes rare for large M. On average the net switching is zero. Intuitively, idiosyncratic payoffs allow one to infer which action is best by just looking at maximal payoffs.

Consider now bounded payoffs, i.e. payoffs are known to be contained in a bounded closed interval which we renormalize into $[0,1]$. The following Proposition summarizes the results for the three considered rules.

Proposition 3. *Neither IBA nor IB are improving. SPOR is improving, with*

$$\begin{aligned} p_i' &= p_i + \left(1 + (1-\bar{\pi}) + \cdots + (1-\bar{\pi})^{M-1}\right)(\pi_i - \bar{\pi})\, p_i \\ &= p_i + \frac{1-(1-\bar{\pi})^M}{\bar{\pi}}(\pi_i - \bar{\pi})\, p_i. \end{aligned} \tag{5}$$

The reason why neither IBA nor IB are improving is as in the case of Imitate If Better their unwillingness to adjust to small changes in payoffs. To see this, consider again the example where P satisfies $P_1(1) = 1 - P_1(0) = \lambda$ and $P_2(x) = 1$ for some given $x \in (0,1)$. Consider a group of $M+1$ DMs seeing each other where one is choosing action 1 while the other M are choosing action 2. Then behavior is the same under IB as under IBA, $F(2,1,2,...,2)_1 = \lambda$ and $F(1,2,...,2)_2 = 1-\lambda$. The net switching in this group from action 2 to action 1 is equal to $M\lambda - (1-\lambda)$ which clearly does not reflect which action is best. Notice however that improving requires that $\pi_1 > (<)\, \pi_2$ implies $MF(2,1,2,...,2)_1 \geq (\leq)\, F(1,2,...,2)_2$ when $p_2 < 1$ but $p_2 \approx 1$.

As an illustration, and still within this example, consider SPOR when $M = 2$. We find that $F(2,1,2)_1 = \frac{1}{2}\lambda + \frac{1}{2}(1-x)\lambda$ and $F(1,2,2)_2 = x + (1-x)x$. So net switching from 2 to 1 is equal to $2F(2,1,2)_1 - F(1,2,2)_2 = (2-x)(\lambda - x)$ which is ≥ 0 if $\pi_1 = \lambda \geq \pi_2 = x$ which ensures the necessary condition for being improving.

The expression (5) for SPOR is derived in Hofbauer and Schlag (2000). If the first sampled DM is imitated then it is as if POR is used, and hence $p_i' = p_i + (\pi_i - \bar{\pi})\, p_i$. However if the first DM is not imitated, in practice she receives a further chance of being imitated, hence $(\pi_i - \bar{\pi})\, p_i$ is added on again, only discounting for the probability $1-\bar{\pi}$ of getting a second chance, etc. Note that when $\bar{\pi}$ is small then the growth rate of SPOR is approximately M times that of single sampling while when $\bar{\pi}$ is large then the growth rate only marginally depends on M. Given (5) it is clear that SPOR is improving. Note also that the growth rate of SPOR is increasing with M. The dynamics that obtains as M tends to infinity is called the *adjusted replicator dynamics,* due to the scaling factor in the denominator (Maynard Smith, 1982).

For $M \geq 2$, there are other strictly improving rules. Indeed, unlike in the single sampling case, a unique class of best rules cannot be selected as under single sampling. An interesting research agenda would be to identify appropriate criteria for selecting among improving rules with multiple sampling.

We conclude with some additional comments on IB and IBA. First, note that IB is improving if only two payoffs are possible. This follows immediately from our previous observation that SPOR and IB yield the same change in average payoffs whenever payoffs are binary valued.

Second, for any given p and any given decision problem, if M is sufficiently large then $\bar{\pi}' \geq \bar{\pi}$ under IBA. This is because, due to the law of large numbers, most individuals will switch to a best action. In fact, this result can be established uniformly for all decision problems with $|\pi_1 - \pi_2| \geq d$ for a given $d > 0$. This statement does not hold for IB, though. For sufficiently large M, and provided distributions are discrete, the dynamics under IB are solely driven by the largest payoff in the support of each action, which need not be related to which action has the largest mean.

2.7. Correlated Noise. It is natural to imagine that individuals who observe each other also face similar environments. This leads us to consider environments where payoffs are no longer realized independently. Instead they can be correlated. We refer to this as *correlated noise* as opposed to independent noise.

Consider the following model of correlated noise. There is a finite number Z of states. Let $q_{\alpha\beta}$ be the probability that a DM is in state α and observes a DM in state β, with $q_{\alpha\beta} = q_{\beta\alpha}$. States are not observable by the DMs. The probability that a given DM is in state α is given by $q_\alpha = \sum_\beta q_{\alpha\beta}$. Apart from the finiteness of states, the previous case with independent payoffs corresponds to the special case where $q_{\alpha\beta} = q_\alpha q_\beta$ for all α, β.

Another extreme is the case with *perfectly correlated noise* where $q_{\alpha\beta} > 0$ implies $\alpha = \beta$. Let $\pi_{i,\alpha}$ be the deterministic payoff achieved by action i in state α. So $\pi_i = \sum_\alpha q_\alpha \pi_{i,\alpha}$ is the expected payoff of action i. Consider first the setting of bounded payoffs.

Proposition 4. *Assume that payoffs $\pi_{i,\alpha}$ are known to be contained in $[0,1]$. If $M = 1$ then the improving rules under independent noise are also improving under correlated noise. If $M \geq 2$ then there exists $(q_{\alpha\beta})_{\alpha\beta}$ such that SPOR is not improving.*

To understand this result, consider first $M = 1$ and let F be improving under independent noise. Then net switching between action i and action j is given by

$$\sum_{\alpha\beta} q_{\alpha\beta} \left(F\left(i, \pi_{i,\alpha}, j, \pi_{j,\beta}\right)_j - F\left(j, \pi_{j,\beta}, i, \pi_{i,\alpha}\right)_i \right) = \sum_{\alpha} q_{\alpha\beta}\sigma_{ij}\left(\pi_{j,\alpha} - \pi_{i,\alpha}\right) = \sigma_{ij}\left(\pi_j - \pi_i\right)$$

and hence F is also improving under correlated noise. Clearly, it is the linearity of net switching behavior in both payoffs that ensures improving under correlated noise. Analogously, it is precisely the lack of linearity when $M \geq 2$ what leads to a violation of the improving condition for SPOR. For, let $M \geq 2$ and consider a group of $M+1$ DMs in which one DM chooses action 1 and the rest choose action 2. Suppose $q_{\alpha\alpha} = q_{\beta\beta} = 1/2$, $\pi_{1,\alpha} = 0$, and $\pi_{2,\alpha} = 1$. The DM who chose action 1 is the only one who switches, thus the net switching from action 1 to action 2 in state α is $s_{12}(\alpha) = 1$. If in state β we have $\pi_{1,\beta} = 1$ and $\pi_{2,\beta} = 0$, thus $s_{12}(\beta) = -M$. Consequently the expected net switching from action 1 to action 2 is equal to $\frac{1}{2}(1 - M)$ which means that SPOR is not improving.

Idiosyncratic payoffs can be embedded in this model with correlated noise. In fact, we can build a model where a first order stochastic dominance relationship emerges by assuming that the order of actions is the same in each state. Specifically, we say that payoffs *satisfy common order* if for all α, β it holds that $\pi_{i,\alpha} \geq \pi_{j,\alpha} \Longleftrightarrow \pi_{i,\beta} \geq \pi_{j,\beta}$. The same calculations used in Section 2.1 (see (2)) then show that, if payoffs are known to satisfy common order, then IIB is improving.

We now turn to population dynamics. In order to investigate the dynamics in a countably infinite population one has to specify how the noise is distributed within the population. The resulting population dynamics can be deterministic or stochastic. We discuss these possibilities and provide some connections to the literature.

Consider first the case where the proportion of individuals in each state is distributed according to $(q_\alpha)_{\alpha \in A}$. Then the population dynamics is deterministic and we obtain the equivalence of improving with $\sigma_{ij} > 0$ and global interior learning (see Proposition 2). The special case where payoffs are perfectly correlated is closely related to Ellison and Fudenberg (1993, Section II) which we now briefly present. In their model there are two actions, payoffs are idiosyncratic, in each period a fraction of the population is selected and receives the opportunity to change their action, an agent revising observes the population average payoffs of each action and applies IBA (i.e. it is as if they had an infinite sample size so $M = \infty$). The consequence is that the time averages of the proportion of DMs using each action converge to the probability with which those actions are best. That is, global learning does not occur, since in general the worst action can outperform the best one with positive probability due to noise. As stated above, global learning occurs when $M = 1$ and all use IB. Thus it is the ineffectiveness of the learning rule combined with the information available that prevents global learning.

Ellison and Fudenberg (1993, Section III) enrich the model by adding popularity weighting (a form of payoff-independent imitation) to the decision rule. Specifically, a DM who receives revision opportunity chooses action 1 if $\pi_1+\varepsilon_1 \geq \pi_2+\varepsilon_2+m(1-2p_1)$, where m is a popularity parameter. That is (for $m > 0$), if action 1 is "popular" ($p_1 > 1/2$), then it is imitated even if it is slightly worse than action 2, and vice versa. Then they further assume that the distribution of the payoff shock $\varepsilon_1 - \varepsilon_2$ is uniform on an interval $[-a, a]$ and find that global learning (convergence to the best action with probability one) requires $a - \Delta\pi \leq m \leq a + \Delta\pi$, where $\Delta\pi = \pi_1 - \pi_2$.[10] In contrast, if payoffs are bounded and all use the rule described in Section 2.4 then global learning emerges under the same informational conditions: agents revising know average payoffs of each action and population frequencies.

Now assume that all DMs are in the same state in each period. Then the dynamic is stochastic and convergence is governed by logarithmic growth rates. Ellison and Fudenberg (1995) consider such a model with idiosyncratic and aggregate shocks where DMs apply IBA to a sample of M other agents. For $M = 1$, IBA is equivalent to IB (up to ties which play no role). Global learning occurs when the fraction of DMs who receive revision opportunities is small enough and the two actions are not too similar. The intuition is that in this case the ratio of the logarithmic growth rates has the same sign

[10] Juang (2001) studies the initial conditions under which an evolutionary process on rules will lead the population to select popularity weighting parameters ensuring global learning. In a society with two groups of agents, these conditions require that either one group adopts the optimal parameter from the beginning, or the optimal parameter lies between those of both groups. That is, "a society does not have to be precise to learn efficiently, as long as the types of its members are suficiently diverse" (Juang 2001, p. 735).

as the difference in expected values. For $M \geq 2$, global learning no longer necessarily obtains when few DMs revise and actions are sufficiently different. Given the relationship between expected change and logarithmic growth rates mentioned above, this inefficiency is due to the fact that IBA is not improving under idiosyncratic payoffs.

3. GAME PLAYING

We now turn to imitation and learning in strategic environments. While above we were interested in whether the population would learn which action is best we are now interested in whether play approaches a Nash equilibrium or another suitable convention.

3.1. Imitation of Kin, Play against Others. There is a straightforward way to translate the framework we considered above to game-playing and thus allow us to investigate imitation in general (here two person) games.

Consider a two player game with two actions for each player. Associate a different population to each player role in the game. Each round individuals are randomly matched in pairs, one from each population, to play the game. Between rounds individuals observe choice and success of someone else within the same population. That is, there is explicitly no strategic interaction between individuals that observe each other. Individuals belonging to the same population are ex-ante in identical situations and hence imitation can be useful. Further, from the point of view of an individual, the other population's play can be viewed as inducing a distribution of outcomes for each possible (own) action. That is, we are (myopically) back in the framework considered above.

Suppose that all individuals in the same population use the same rule. The improving condition becomes the objective to increase average payoffs in one population, for all distributions of own and opponent play and for all games, *under the (mistaken) belief that the opponent population's behavior does not change.* Equivalent, the rule must induce a better reply dynamics in each population. Thus we need no separate analysis of which rules are selected and immediately can proceed to the analysis of dynamics.

Consider the dynamics which results when each population uses a single rule, namely a strictly improving single sampling rule or SPOR when $M \geq 2$. It is easily shown that if play starting in the interior converges then it converges to a Nash equilibrium. This is, of course, unsurprising, since the rules aim to increase individual payoffs and the multi-population setting abstracts from strategic effects. If there is no expected movement then each action chosen yields the same expected payoff. For an interior initial state, actions not chosen in the limit must achieve lower expected payoff (see Weibull 1995 for a formal argument). However, trajectories need not converge to a single point.

By a standard argument, reducing the time between rounds and the proportion of individuals that observe someone else between rounds, one obtains a continuous-time dynamics. Strictly improving rules under single sampling induce the standard replicator dynamics of Taylor (1979). Convergence from an interior starting point to a Nash equilibrium holds in all types of 2×2 games except in those that have best reply structure as in Matching Pennies. In this Matching Pennies-type of game the replicator dynamics are known to cycle forever around the interior Nash equilibrium. Hofbauer and Schlag (2000) show that the dynamics under SPOR with $M \geq 2$ starting in the interior converge to the Nash equilibrium. Observing two others is sufficient to lead the population to the Nash equilibrium from an interior initial state in all 2×2 games. However this result should not be expected to hold in all more general games. For instance it will not hold in the generalized Shapley game of Hofbauer and Swinkels (1995, cf. Balkenborg and Schlag, 2007) that involves a $2 \times 2 \times 2$ game among three populations.

Of course, the true dynamics is not continuous but discrete, driven by the jumps associated to a strictly positive proportion of individuals changing actions. Hofbauer and Schlag (2000) investigate the discrete dynamics induced by SPOR and show that the Nash equilibrium of Matching Pennies is repelling for all M. Under single sampling the population state spirals outwards to the boundary. When $M \geq 2$ then the dynamics will circle close to and around the Nash equilibrium if sufficiently few individuals observe the play of others between rounds.[11]

Pollock and Schlag (1999) consider individuals that know the game they play, so uncertainty is only about the distribution of actions. They investigate conditions on a single sampling rule that yield a payoff monotone dynamics in a game that has a cyclic best response structure as in Matching Pennies. They find that the rule has to be imitating and that the continuous version of the population dynamics will have—like the standard replicator dynamics—closed orbits around the Nash equilibrium. They contrast this to the finding that there is no rule only based on a finite sample of opponent play that will lead to a payoff monotone dynamics. This is due to the fact that information on success of play has to be stored and recalled in order to generate a payoff monotone dynamics.

[11] Cycling can have a descriptive appeal, for such cycles might describe fluctuations between costly enforcement and fraud (e.g. see Cressman et al., 1998).

Dawid (1999) considers two populations playing a battle-of-the-sexes game, where each agent observes a randomly selected other member of the same population and imitates the observed action if the payoff is larger than the own and the gap is large enough. For certain parameter values, this model includes PIR. The induced dynamics is payoff monotone. In games with no risk-dominant equilibrium, there is convergence towards one of the pure-strategy coordination equilibria unless the initial population distribution is symmetric. In the latter case, depending on the model's parameters, play might either converge to the mixed-strategy equilibrium or to periodic or complex attractors. If one equilibrium is risk dominant, it has a larger basin of attraction than the other one.

3.2. Imitating your Opponents. In the following we consider the situation where player roles are not separated. There is a symmetric game and agents play against and learn from agents within the same population. Environments where row players cannot be distinguished from column players include oligopolies and financial markets. Here it makes a difference whether we look for rules that increase average payoffs or whether we look for rules that induce a better reply dynamics.

Consider first the objective to induce a better reply dynamics. Rules that we characterized as being improving in decision problems have this property. To induce a (myopic) better reply dynamic means that, if play of other agents does not change, an individual agent following the rule should improve payoffs. Thus this condition is identical to the improving condition for decision problems. Specifically, a rule induces a better reply dynamic if and only if it is improving in decision problems. The condition of bounded payoffs translates to considering the set of all games with payoffs within these bounds. The decision setting with idiosyncratic payoffs translates into games where all pure strategies can be ordered according to dominance.

Now turn to the objective of finding a rule that always increases average payoffs. Ania (2000) presents an interesting result that shows that this is not possible unless average payoffs remain constant. The reason is as follows. When a population of players are randomly matched to play a Prisoners' Dilemma, in a state with mostly cooperators and only a few defectors, increase in average payoffs requires that more defectors switch to cooperate than vice versa. However note that the game might just as well not be a Prisoners' Dilemma but one where mutual defection yields a superior payoff to mutual cooperation. Then cooperators should switch more likely to defect than vice versa. Note that the difference between these two games does not play a role when there are mostly cooperators and hence the only way to solve the problem is for there to be no net switching. Thus, the strategic framework is fundamentally different from the individual decision framework of e.g. Schlag (1998).

Given this negative result it is natural to directly investigate the connection between imitation dynamics and Nash equilibria. The following dynamics, which we will refer to as the *perturbed imitation dynamics*, has played a prominent role in the literature. Each period, players receive revision opportunities with a given, exogenous probability $0 < 1 - \delta \leq 1$, i.e. δ measures the amount of inertia in individual behavior. When allowed to revise, players observe either all or a random sample of the strategies used and payoffs attained in the last period (always including the own) and use an imitation rule, e.g. Imitate the Best. Additionally, with an exogenous probability $0 < \varepsilon < 1$, players mutate (make a mistake) and choose a strategy at random, all strategies having positive probability. Clearly, the dynamics is a Markov chain in discrete time, indexed by the mutation probability. The "long-run outcomes" (or stochastically stable states) in such models are the states in the support of the (limit) invariant distribution of the chain as ε goes to zero. See Kandori et al (1993) or Young (1993) for details.

The first imitation model of this kind is due to Kandori et al (1993), who show that the long-run outcome when N players play an underlying two-player, symmetric game in a round-robin tournament, the long-run outcome corresponds to the symmetric profile where all players adopt the strategy of the risk-dominant equilibrium, even if the other pure-strategy equilibrium is payoff-dominant. A clever robustness test was performed by Robson and Vega-Redondo (1996), who show that when the round-robin tournament is replaced by random matching, the perturbed IB dynamics leads to payoff-dominant equilibria instead.

We concentrate now on proper N-player games. When considering imitation in games, it is natural to restrict attention to symmetric games, i.e. games where the payoff of each player k is given through the same function $\pi(s_k|s_{-k})$ where s_k is the strategy of player k, s_{-k} is the vector of strategies of other players, all strategy spaces are equal, and $\pi(s_k|s_{-k})$ is invariant to permutations in s_{-k}.

The consideration of N-player, symmetric games immediately leads to a departure from the framework in the previous sections. First, DMs imitate their opponents, so that there is no abstracting away from strategic considerations. Second, the population size has to be N, i.e. we will deal with a finite population framework and no large-population limit can be meaningfully considered for the resulting dynamics.

It turns out that the analysis of imitation in N-player games is tightly related to the concept of finite population ESS (Evolutionarily Stable Strategy), which is different from the classical infinite-population ESS. This notion was developed by Schaffer (1988). A finite population ESS is a strategy such that, if it is adopted by the whole population, any single deviant (mutant) will fare worse than the incumbents after deviation. Formally, it is a strategy a such that $\pi\left(a \left| b, a, \overset{N-2}{\ldots}, a\right.\right) \geq \pi\left(b \left| a, \overset{N-1}{\ldots}, a\right.\right)$ for any other strategy b. An ESS is strict if this inequality is always strict. Note that, if a is a finite-population ESS, the profile $(a, \ldots, a)$ does not need to be a Nash Equilibrium. Instead of maximizing the payoffs of any given player, an ESS maximizes relative payoffs—the difference between the payoffs of the ESS and those of any alternative 'mutant' behavior.[12]

An ESS a is (strictly) globally stable if

$$\pi\left(a \left| b, \overset{m}{\ldots}, b, a, \overset{N-m-1}{\ldots}, a\right.\right) (>) \geq \pi\left(b \left| b, \overset{m-1}{\ldots}, b, a, \overset{N-m}{\ldots}, a\right.\right)$$

for all $1 \leq m \leq N-1$, that is, if it resists the appearance of any fraction of such experimenters. We obtain:

Proposition 5. *For an arbitrary, symmetric game, if there exists a strictly globally stable finite-population ESS a, then $(a, \ldots, a)$ is the unique long-run outcome of all perturbed imitation dynamics where the imitation rule is such that actions with maximal payoffs*

[12] An ESS may correspond to spiteful behavior, i.e. harmful behavior that decreases the survival probability of competitors (Hamilton 1970).

are imitated with positive probability and actions with worse payoffs than the own are never imitated, e.g. IB or PIR.

Alós-Ferrer and Ania (2005a) prove this result for IB. However, the logic of their proof extends to all the rules mentioned in the statement. The intuition is as follows. If the dynamics is at $(a, \ldots, a)$, any mutant will receive worse payoffs than the incumbents and hence will never be imitated. However, starting from any symmetric profile $(b, \ldots, b)$, a single mutant to a will attain maximal payoffs and hence be imitated with positive probability. Thus, the dynamics flows towards $(a, \ldots, a)$.

Schaffer (1989) and Vega-Redondo (1997) observe that, in a Cournot oligopoly, the output corresponding to a competitive equilibrium—the output level that maximizes profits at the market-clearing price—is a finite population ESS. That is, a firm deviating from the competitive equilibrium will earn lower profits than its competitors after deviation. Actually, Vega-Redondo's proof shows that it is a strictly, globally stable ESS. Additionally, Vega-Redondo (1997) shows that the competitive equilibrium is the only long-run outcome of a learning dynamics where players update strategies according to Imitate the Best and occasionally make mistakes (as in Kandori et al, 1993).

Possajennikov (2003) and Alós-Ferrer and Ania (2005a) show that the results for the Cournot oligopoly are but an instance of a general phenomenon. Consider any *aggregative game*, i.e. a game where payoffs depend only on individual strategies and an aggregate of all strategies (total output in the case of Cournot oligopolies). Suppose there is strategic substitutability (*submodularity*) between individual and aggregate strategy. For example, in Cournot oligopolies the incentive to increase individual output decreases the higher the total output in the market. Define an aggregate-taking strategy (ATS) to be one that is individually optimal given the value of the aggregate that results when all players adopt it. Alós-Ferrer and Ania (2005a) show the following.

Proposition 6. *Any ATS is a finite population ESS in any submodular, aggregative game. Further, any strict ATS is strictly globally stable, and the unique ESS.*

This result has a natural counterpart in the supermodular case (strategic complementarity), where any ESS can be shown to correspond to aggregate-taking optimization.[13]

As a corollary of the last two propositions, any strict ATS of a submodular aggregative game is the unique long-run outcome of the perturbed imitation dynamics with e.g. IB, hence implying the results in Vega-Redondo (1997).

These results show that, in general, imitation in games does not lead to Nash equilibria. The concept of finite-population ESS, and not Nash equilibrium, is the appropriate tool to study imitation outcomes.[14] In some examples, though, the latter might be a subset of the former. Alós-Ferrer et al (2000) consider Imitate the Best in the framework of a Bertrand oligopoly with strictly convex costs. Contrary to the linear costs setting, this game has a continuum of symmetric Nash equilibria. Imitate the Best selects a proper

[13]Leininger (2006) shows that, for submodular aggregative games, every ESS is globally stable.

[14]For the inertialess case, this assertion depends on the fact that we are considering rules which only depend on the last period's outcomes. Alós-Ferrer (2004) shows that, even with just an additional period of memory, the perturbed IB dynamics with $\delta = 0$ selects all symmetric states with output levels between, and including, the perfectly competitive outcome and the Cournot-Nash equilibrium.

subset of those equilibria. As observed by Ania (2006), the ultimate reason is that this subset corresponds to the set of finite-population ESS.[15]

The work just summarized focuses mainly on Imitate the Best. As seen in Proposition 5, there are no substantial differences if one assumes PIR instead. The technical reason is that the models mentioned above are finite population models with vanishing mutation rates. For these models, results are driven by the existence of a strictly positive probability of switching, not by the size of this probability. Behavior under PIR is equivalent to that of any other imitative rule in which imitation only takes place when observed payoff is strictly higher than own payoff. Whether or not net switching is linear plays no role. Rules like IBA or SPOR would produce different results, though, although a general analysis has not yet been undertaken.

We would like to end this chapter by reminding the reader that our aim was to concentrate on learning rules, and in particular imitating ones, that can be shown to possess appealing optimality properties. However, we would like to point out that a large part the

[15] Alós-Ferrer and Ania (2005b) study an asset market game where the unique pure-strategy Nash equilibrium is also a finite population ESS. They consider a two-portfolio dynamics on investment strategies where wealth flows with higher probability into those strategies that obtained higher realized payoffs. Although the resulting stochastic process never gets absorbed in any population profile, it can be shown that, whenever one of the two portfolios corresponds to the ESS, a majority of traders adopt it in the long run. The dynamics can also be interpreted as follows: each period, an investor updates her portfolio. The probability that this revision results in an investor switching from the first portfolio to the second, rather than the opposite, is directly proportional to the difference in payoffs between the portfolios. That is, those probabilities follow the Proportional Imitation Rule.

literature on learning in both decision problems and games has been more descriptive. Of course, from a behavioral perspective we would expect certain, particularly simple rules like IB or PIR to be more descriptively relevant than others. For example, due to its intricate definition, we think of SPOR more as a becnhmark. Huck et al (1999) find that the informational setting is crucial for individual behavior. If provided with the appropriate information, experimental subjects do exhibit a tendency to imitate the highest payoffs in a Cournot oligopoly. Apesteguía et al (2007) elaborate on the importance of information and also report that the subjects' propensity to imitate more successful actions is increasing in payoff differences as specified by PIR. Barron and Erev (2003) and Erev and Barron (2005) discuss a large number of decision-making experiments and identify several interesting behavioral traits which oppose payoff maximization. First, the observation of high (foregone) payoff weighs heavily. Second, alternatives with the highest recent payoffs seem to be attractive even when they have low expected returns. Thus, IB or PIR might be more realistic than IBA.

References

[1] Alós-Ferrer, C. (2004), "Cournot vs. Walras in Oligopoly Models with Memory," *Int. J. Ind. Organ.* 22, 193-217.

[2] Alós-Ferrer, C. and A.B. Ania (2005a), "The Evolutionary Stability of Perfectly Competitive Behavior," *Econ. Theory* **26**, 497-516.

[3] Alós-Ferrer, C. and A.B. Ania (2005b), "The Asset Market Game," *J. Math. Econ.* **41**, 67-90.

[4] Alós-Ferrer, C., A.B. Ania and K.R. Schenk-Hoppé(2000), "An Evolutionary Model of Bertrand Oligopoly," *Games Econ. Behav.* **33**, 1-19.

[5] Ania, A.B. (2000), *Learning by Imitation when Playing the Field*, Working paper 0005, Department of Economics, University of Vienna.

[6] Ania, A.B. (2006), "Evolutionary stability and Nash equilibrium in finite populations, with an application to price competition," *J. Econ. Behav. Org.* (forthcoming).

[7] Apesteguía, J., S. Huck and J. Oechssler (2007), "Imitation—Theory and Experimental Evidence," *J. Econ. Theory* **136**, 217-235.

[8] Balkenborg, D. and K.H. Schlag (2007), "On the Evolutionary Selection of Nash Equilibrium Components," *J. Econ. Theory* **133**, 295-315.

[9] Bandura, A. (1977). *Social Learning Theory.* Prentice-Hall, Englewood Cliffs, N.J.

[10] Banerjee, A. (1992), "A Simple Model of Herd Behavior," *Quart. J. Econ.* **107**, 797–817.

[11] Barron, G., and I. Erev (2003), "Small Feedback-Based Decisions and their Limited Correspondence to Description-Based Decisions," *J. Behav. Dec. Making* **16**, 215-233.

[12] Bessen, J., and E. Maskin (1999), "Sequential Innovation, Patents, and Imitation," *Quart. J. Econ.* (forthcoming)

[13] Björnerstedt, J. and K.H. Schlag (1996), *On the Evolution of Imitative Behavior*, Discussion Paper No. B–378, Sonderforschungsbereich 303, University of Bonn.

[14] Börgers, T., A. Morales and R. Sarin (2004), "Expedient and Monotone Rules," *Econometrica* **72(2)**, 383–405.

[15] Boylan, R.T. (1992), "Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals," *J. Econ. Theory* **57**, 473-504.

[16] Cho, I.-K. and D. Kreps (1987), "Signaling Games and Stable Equilibria," *Quart. J. Econ.* **102**, 179-221.

[17] Conlisk, J. (1980), "Costly Optimizers versus Cheap Imitators," *J. Econ. Behav. Org.* **1**, 275-293.

[18] Cressman, R., W.G. Morrison and J.F. Wen (1998), "On the Evolutionary Dynamics of Crime," *Can. J. Econ.* **31**, 1101-1117.

[19] Cross, J. (1973), "A Stochastic Learning Model of Economic Behavior," *Quart. J. Econ.* 87, 239-266.

[20] Dawid, H. (1999), "On the Dynamics of Word of Mouth Learning with and without Anticipations," *Ann. Oper. Research* **89**, 273-295.

[21] Ellison, G., and Fudenberg, D. (1993), "Rules of Thumb for Social Learning," *J. Polit. Economy* **101**, 612-643.

[22] Ellison, G., and Fudenberg, D. (1995), "Word of Mouth Communication and Social Learning," *Quart. J. Econ.* **110**, 93-125.

[23] Erev, I., and G. Barron (2005), "On Adaptation, Maximization, and Reinforcement Learning among Cognitive Strategies," *Psychological Review* 112, 912-931.

[24] Fudenberg, D. and D.K. Levine (1998), *The Theory of Learning in Games*, MIT Press, Cambridge.

[25] Hamilton, W. (1970), "Selfish and Spiteful Behavior in an Evolutionary Model," *Nature* **228**, 1218-1220.

[26] Hofbauer, J. and K.H. Schlag (2000), "Sophisticated Imitation in Cyclic Games," *J. Evol. Econ.* **10(5)**, 523-543.

[27] Hofbauer, J. and J. Swinkels (1995), *A Universal Shapley-Example*, Unpublished Manuscript, University of Vienna and Washington University in St. Louis.

[28] Huck, S., H.T. Normann and J. Oechssler (1999), "Learning in Cournot Oligopoly—An Experiment," *Econ. J.* **109**, C80-C95.

[29] Juang, W.-T. (2001), "Learning from Popularity", *Econometrica* 69, 735-747.

[30] Kandori, M., G. Mailath and R. Rob (1993), "Learning, Mutation and Long Run Equilibria in Games," *Econometrica* **61**, 29-56.

[31] Kreps, D. and R. Wilson (1982), "Reputation and Imperfect Information," *J. Econ. Theory* **27**, 253–279.

[32] Lakshmivarahan, S. and M. A. L. Thathachar (1973), "Absolutely Expedient Learning Algorithms for Stochastic Automata," *IEEE Trans. Syst., Man. Cybernetics*, vol. **SMC-3**, 281–286.

[33] Leininger, W. (2006), "Fending off one means fending off all: evolutionary stability in submodular games," *Econ. Theory* **29**, 713-719.

[34] Maynard Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge Univ. Press, Cambridge.

[35] Morales, A.J. (2002), "Absolutely Expedient Imitative Behavior," *Int. J. Game Theory* **31**, 475-492.

[36] Morales, A.J. (2005), "On the Role of Group Composition for Achieving Optimality," *Ann. Oper. Research* 137, 378-397.

[37] Oyarzun and Ruf (2007), *Monotone Imitation*, Unpublished Manuscript, Texas A&M and Columbia University.

[38] Pingle, M., and R.H. Day (1996), "Modes of economizing behavior: experimental evidence," *J. Econ. Behav. Org.* **29**, 191-209.

[39] Pollock, G. and K.H. Schlag (1999), "Social Roles as an Effective Learning Mechanism," *Rationality and Society* **11**, 371-397.

[40] Possajennikov, A. (2003), "Evolutionary Foundations of Aggregate-Taking Behavior," *Econ. Theory* **21**, 921-928.

[41] Robson, A.J., and F. Vega-Redondo (1996), "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *J. Econ. Theory* **70**, 65-92.

[42] Rogers, A. (1989), "Does Biology Constrain Culture?" *Amer. Anthrop.* **90**, 819–831.

[43] Schaffer, M.E. (1988), "Evolutionarily Stable strategies for a Finite Population and a Variable Contest Size," *J. Theor. Biol.* **132**, 469-478.

[44] Schaffer, M.E. (1989), "Are Profit-Maximisers the Best Survivors?," *J. Econ. Behav. Org.* **12**, 29-45.

[45] Schlag, K.H. (1996) *'Imitate Best' vs 'Imitate Best Average'*, Unpublished Manuscript, University of Bonn.

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

[47] Schlag, K. H. (1999), "Which one should I imitate?" *J. Math. Econ.* **31**, 493-522.

[48] Sinclair, P.J.N. (1990), "The Economics of Imitation," *Scottish J. Pol. Econ.* **37**, 113-144.

[49] Squintani, F. and J. Välimäki (2002), "Imitation and Experimentation in Changing Contests," *J. Econ. Theory* **104**, 376–404.

[50] Taylor P.D. (1979), "Evolutionarily Stable Strategies with Two Types of Players," *J. Appl. Prob.* **16**, 76–83

[51] Veblen, T. (1899), *The Theory of the Leisure Class, An Economic Study of Institutions,* The Macmillan Company, New York.

[52] Vega-Redondo, F. (1997), "The Evolution of Walrasian Behavior," *Econometrica* **65**, 375-384.

[53] Weibull, J. (1995) *Evolutionary Game Theory*, MIT Press, Cambridge.

[54] Young, P. (1993), "The Evolution of Conventions," *Econometrica* **61**(1), 57-84.