CHAPTER 2

INFORMATIONAL INFLUENCES AND THE AMBIVALENCE OF IMITATION

André ORLÉAN

1. Introduction

Imitation is a complex phenomenon that has been studied by economists and other social scientists for a long time. An individual can choose to imitate another individual or a group for a wide array of different reasons: because he is convinced by the other individual's argument; because of peer group pressure as in Leibenstein's (1982) situation where workers within a firm have to conform to the prevailing "effort convention"; because of manias and fads as in Kindleberger's (1978) or Shiller's (1989) analysis of financial instability; because of social sanctions as in Akerlof's (1980) interpretation of social customs; because of coordination externalities as in Arthur's (1989) or Young's (1993) models; to discover a better strategy as in Nelson and Winter (1982) or Axelrod (1984); because of tradition as in Boyd and Richerson's (1985; 1993) study of cultural transmission; for the sake of conformity as in Asch's (1951) experiment; and the list is not exhaustive [1].

The present paper is devoted to "informational influences" [2]. It considers a group situation where agents have no access to others' private information, but can only observe their actions. Since these actions reflect part of their private information, it is rational for each agent to take them into account. Informational influences can give rise to imitative behaviors. Imagine that a sudden fire breaks out in a room, that there are two doors, one being a dead end, the other one being the right exit, and that I do not know which is which. If I see one person leaving the room, it is rational for me to imitate him. If this person has no information and has formed his choice by pure chance, to follow him will not decrease my performance because he is exactly in the same situation than I am. But if this agent has some knowledge about the true distribution of doors, I will improve my performance by imitating him. My ability to evaluate the quality of the information on the basis of which the other has acted is clearly a central parameter in that reasoning. It should be noted that informational influences are

1. Let us remark that these different mechanisms can overlap. A more systematic presentation will be found in Bikhchandani, Hirshleifer, and Welch (1992).
2. The concept of "informational social influence" has been introduced by the social psychologists Deutsch and Gerard (1955). The economical notion of "informational influences" is akin to this concept.

very common in finance because traders usually have no access to others' information but can nevertheless observe prices which are generated by others' actions. The possibility for prices to reveal part or all of the private information is at the core of the notion of informational efficiency (Grossman 1976; Grossman and Stiglitz 1980).

A body of recent work have analyzed informational influences within a dynamical framework. These works have shown that pure informational influences can lead to imitative decision processes such as "herd behavior" (Banerjee 1992; Scharfstein and Stein 1990), "informational cascade" (Bikchandani, Hirshleifer and Welch 1992), or contagion (Arthur and Lane 1993; Kirman 1993). It has been demonstrated that optimizing agents can decide rationally to follow what previous agents have chosen rather than to use their own private information. Such a process leads to a general conformity on a certain behavior or opinion. These results have been obtained in models which assume a sequential decision process: individuals enter the market one by one, observe their predecessors and take a unique and irreversible decision (Banerjee 1992; Bikchandani, Hirshleifer and Welch 1992). The assumption of sequentiality is crucial. It allows the agent entering the market in $(t+1)$ to calculate the true probability $P(\mathbf{a}_t|\boldsymbol{\sigma}_t, \theta)$ of previous choices $\mathbf{a}_t$ conditionally to $\boldsymbol{\sigma}_t$, the set of the t first agents' private information, and θ the state of nature that has been drawn at the beginning of the process at time $(t=0)$. Then, using Bayes' rule, the individual can infer $P(\theta\,|\mathbf{a}_t)$ and then compute his expected utility.

The aim of this paper is to extend these results to a non sequential framework where agents interact simultaneously with each other, modifying their action at each period of time without knowing when "the process has begun". In other words, agents do not know when θ is changing. Such a decision structure is better suited for the modeling of market situations. For instance in Bikchandani *et al.*, "the order of individuals is exogenous and is known to all" (p. 999): the first individual chooses either H or L, according to the value of his private information + or – ; thus, the second individual is able to infer the first agent's private information from his decision. In our model, this is no longer possible. Because agents do not know when θ has changed, they do not know who is the first of the "new round" nor who is the second. We assume that at each period t, an agent' set of information is strictly reduced to his private information σ_t and a publicly available information expressing the group behavior in $(t-1)$, named $f(t-1)$. $f(t-1)$ can be considered in our model as equivalent to a price. In such a framework, a specific problem appears. In order to determine the relative weight μ they will give to their two informations, σ_t and $f(t-1)$, agents have to evaluate the quality of the signal f. But this quality is not exogenous; it depends on the way other agents have weighted the two informations. For instance, if previous agents have mainly relied on their private information, the signal f is very informative and posterior agents should increase the weight of f in their decision. But, because such a behavior will make f less informative, it may lead to an opposite change in μ. In such a context, we have to examine the existence of an equilibrium value of μ.

To handle this problem, we first begin, in section 2, by assuming that all the agents follow a constant rule of decision defined by μ. In our model, μ expresses the propensity to choose according to the majority side of the group rather than to use one's own information; it measures agents' propensity to imitate. Section 3 presents our basic model. It shows that, when μ is small, the individual and collective performance is low because agents neglect the important quantity of information contained in f. When the propensity to imitate increases, the collective and individual performance increases; the average error decreases. The role of imitation is clearly positive here because it allows agents to take into account part of the collective information. Hence, imitation can be viewed as the specific mechanism through which collective information is made available to individual agents within a decentralized information structure. Then we show that imitation improves individual and collective performance *only if the propensity to imitate is not too high*: when imitation gets to a certain threshold, self-validating processes appear which converge on wrong choices. Because agents insufficiently rely on their private information, the collective outcome f is no more informative and imitation becomes counter-productive. The fact that imitation can be either positive or negative according to its intensity is what we refer to as "the ambivalence of imitation". This ambivalence is a very intuitive result: it is efficient for me to imitate the others as long as they are better informed than I am; it becomes inefficient if they are also imitators. Such a situation has been described by Kindleberger (1978): "The action of each individual is rational -or would be, were it not for the fact that others are behaving in the same way" (p. 34). Nevertheless, as far as we know, the fact that imitation can be ambivalent has never been demonstrated.

In section 4, we calculate the optimal degree of imitation μ^* and consider the game where each agent i has to determine his strategy $\mu(i)$. We show that a situation where all the agents choose the same value μ^* is a very implausible Nash equilibrium as soon as the number of agents is large. It is always better to be a deviant. This paradoxical structure is quite similar to the one analyzed by Grossman-Stiglitz (1980) and highlights the complex role played by imitation on financial markets. Nevertheless, it should be emphasized that in finance "herding externalities" are only a part of the story because imitation has a cost (the stock market price) which is depending on the number of imitators [1]. Section 5 concludes by showing that our result on the ambivalence of imitation remains true for a very large set of decision rules.

2. The problem

Let us consider an economy composed of N agents noted i, $i \in \{1,2, ..., N\}$. The state of the world, named θ, is either $\{H\}$ or $\{L\}$ with equal

1. This point is discussed in Bikchandani *et al.* (1992, pp. 1012-3).

prior probability. The agents cannot observe directly θ. In order to discover it, each individual independently observes a signal σ defined as follows: σ can be either $\{+\}$ or $\{-\}$, and its value is linked to the state of the economy through the following conditional probabilities:

$$\begin{cases} P(\sigma = + \mid H) = P(\sigma = - \mid L) = p > 0.5 \\ P(\sigma = - \mid H) = P(\sigma = + \mid L) = 1 - p < 0.5 \end{cases} \tag{2.1}$$

σ_i, the value of the independent observation made by agent i, is called his private information. Using Bayes' rule, a rational agent can easily evaluate the probability of being in state $\{H\}$ accordingly to his private information. He obtains:

$$P(H|\sigma_i = +) = P(H| +) = \frac{P(+ \mid H) \,.\, P(H)}{P(+)} = p_i$$

and

$$P(H|\sigma_i = -) = P(H| -) = \frac{P(- \mid H) \,.\, P(H)}{P(-)} = 1 - p_i \tag{2.2}$$

where p_i is agent i's estimation of p. We will assume that agents do not know the exact value of p, but make no mistake about the "direction of the correlation"; i.e. $\forall i, p_i > 0.5$. Then, it follows from equations (2.2) that every agent i will choose $\{H\}$ if he has observed $\{\sigma_i = +\}$ and $\{L\}$ if $\{\sigma_i = -\}$. Thus the probability of making the wrong choice is equal to $(1 - p)$. The greater is the accuracy of the signal σ, the smaller is the probability of making the wrong choice. This general framework is the one proposed by Bikchandani *et al.* (1992).

Let us now consider the whole population. Let us note n, the number of individuals having chosen $\{H\}$ and f equal to n/N, the proportion of such individuals. If every individual is forming his choice on the sole basis of his private and independent information σ_i, the probability of choosing $\{H\}$ will be either p or $(1 - p)$ depending upon θ, the value of the state of the economy. Thus f is a realization of the binomial law $\tilde{A}(\theta)$ defined as follows:

$$\begin{cases} E[\tilde{A}(\theta)] = a(\theta) \\ Var[\tilde{A}(\theta)] = \dfrac{p\,.\,(1-p)}{N} \end{cases} \quad \text{with} \quad \begin{cases} a(H) = p \\ a(L) = 1 - p \end{cases} \tag{2.3}$$

It follows that when agents make their choice on the sole basis of their private information, the average frequency of people having made the wrong choice is equal to $(1 - p)$. It is identical to the average mistake made by an individual

i drawn randomly within the population. Because the rule of choice that has been considered does not take into account the collective information collected by the group, the average collective performance is equal to the one of an isolated agent. Can the use of the collective information improve this result?

It is not difficult to see that the quantity of information contained within the whole group is much richer than the information possessed by an individual. To illustrate this point, let us assume the existence of a center than collects all the private information. Let us note f_+ the proportion of $\{+\}$ that have been observed within the whole population. This value is calculated by the center and then communicated to every agent. Obviously f_+ follows the binomial law $\tilde{A}(\theta)$. Let us assume that all the members of the group adopt the following rule of decision: "if $\{f_+ > 0.5\}$, they choose $\{H\}$; if $\{f_+ < 0.5\}$, they choose $\{L\}$; if $\{f_+ = 0.5\}$, they choose either $\{H\}$ or $\{L\}$ with probabilities equal to 0.5". This rule of choice will lead to a much better performance than the preceding one. Indeed the probability of error is equal to the following probability:

$$P(f_+ < 0.5|H) + 0.5 \,.\, P(f_+ = 0.5|H) = P(f_+ > 0.5|L) + 0.5 \,.\, P(f_+ = 0.5|L)$$

If N is great, this probability is very small. For instance, if $N = 51$ and $p = 0.7$, it is equal to 0.0013, to be compared to 0.30 in the former case. This rule is very efficient, but it needs a center which collects all the private information to be implemented. In this paper we will consider institutional settings where such a centralization of information is impossible or too costly. We will analyze a decentralized structure of interactions, i.e. a structure where agents have no direct access to other agents' private information, but can only observe their choices. More precisely, we will assume that only f, the collective choice, can be observed. This notion of decentralization has been considered because it is close to the way a market functions. In a market, at each date t, every agent only knows his private information and the value of the price in $(t-1)$ which aggregates individual choices. The question we wish to address is how these decentralized interactions can work. To begin with, we have to analyze the way individuals take their decisions.

We define a decentralized rule of decision as a couple of functions $q(f, \sigma_i)$, where σ_i is either $\{+\}$ or $\{-\}$, which gives the probability of choosing $\{H\}$ when the values σ_i and f have been observed. A first rule of decision is the one considered previously where the agent does not take into account the others. He chooses on the sole basis of his private information. We will note this rule q_0. It is defined as follows:

$$\begin{cases} q_0(f, +) = 1 \\ q_0(f, -) = 0 \end{cases} \tag{2.4}$$

This rule does not depend on f. We have seen that it leads to a situation where the average proportion of people having made the right choice, i.e. either

$\{H\}$ if $\{\theta = H\}$ or $\{L\}$ if $\{\theta = L\}$, is equal to p. Our intuition is that taking care of f will improve this result.

The general rule that will be considered is determined as follows. When $\{\sigma_i = +\}$ and $\{f \geq 0.5\}$, both informations lead to the same decision $\{H\}$: they are consistant. The situation is different when $\{\sigma_i = +\}$ and $\{f < 0.5\}$ because then, the two signals are contradictory. If we note $\mu \in [0,1]$, the probability to choose according to f, the majority side, we obtain q_μ the following general rule of decision:

$$\begin{cases} \text{if } f \geq 0.5 & q_\mu(f,+) = 1 \\ \text{if } f < 0.5 & q_\mu(f,+) = 1 - \mu \end{cases} \quad \text{and} \quad \begin{cases} \text{if } f > 0.5 & q_\mu(f,-) = \mu \\ \text{if } f \leq 0.5 & q_\mu(f,-) = 0 \end{cases}$$

Then we have the relation:

$$q_\mu(f,-) = 1 - q_\mu(1-f,+)$$

which expresses the symmetry between $\{-\}$ and $\{+\}$. $(1-\mu)$ measures the propensity to choose according to one's own private information.

When μ equals 0, i.e. when agents strictly follow their private information, the average error of the group is $(1-p)$. Our central question is: what happens when every agent follows the rule q_μ with $\mu > 0$? In such a situation, agents' choices are no longer independent: the way agent i chooses depends upon previous agents' choices. It follows that in order to understand how f is determined, we have to be more precise about the definition of the dynamical process of interactions.

3. The basic model

Let us consider the following process. In $(t = 0)$, the state of the economy θ is drawn according to the probabilities $P(H) = P(L) = 0.5$ and will remain constant till date T. Because the sequence $[0,T]$ can be understood as a specific round amid a global process which began before time 0 and will continue after time T, $f(0)$ is depending on what has happened before $(t = 0)$ [1]. Within our model, it will be considered as an arbitrary given parameter. At time $(t > 0)$, one individual i is randomly drawn. He observes $f(t-1)$ and the signal σ. Then he revises his last choice according to the rule $q_\mu(f, \sigma_i)$.

---

1. For instance θ_t can be viewed as a markovian process defined for $t \in Z$ by the following transition matrix:

$$\begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

with ε being as small as wanted.

Let us emphasize that agent i does not take into account in t the information he could have observed before. This simplifying hypothesis can be justified because of the short memory of agents, and/or by the fact that agents do not know when θ is changing and, consequently, do not know if their past information is still relevant. We have also assumed that only one agent is drawn at each instant. We could have assumed that L agents are drawn randomly, with L smaller than N, without qualitatively modifying our results.

This set of hypotheses define a Markovian stochastic process. The variable we are interested in is the probability law $P(f\,;t)$ followed by $f(t)$. To what does it converge? To answer this question, we first have to determine the probability to choose $\{H\}$ at $(t+1)$ when $f(t)=f$. Because the probability to draw $\{+\}$ is equal to $a(\theta)$, we obtain:

$$J(f,\mu,\theta) = a(\theta) \cdot q_\mu(f,+) + [1-a(\theta)] \cdot q_\mu(f,-) \qquad (3.1)$$

Let us emphasize that $J(f,\mu,\theta)$ depends on θ through the value of a, i.e. the average number of $\{+\}$ that agents have observed. Knowing the probability J, we can calculate:

$$\begin{cases} P[f(t+1)=f(t)+1/N] = P(f \to f+1/N) = (1-f) \cdot J(f,\mu,\theta) = W_+(f,\mu,\theta) \\ P[f(t+1)=f(t)-1/N] = P(f \to f-1/N) = f \cdot [1-J(f,\mu,\theta)] = W_-(f,\mu,\theta) \end{cases}$$

These equations completely determined the stochastic process followed by $f(t)$, i.e. the way $P(f\,;t)$ varies through time. (When there is no possible confusion, θ and μ will be omitted for the sake of simplicity). The exact description of this process in terms of discrete numbers n is called the master equation (Weidlich and Haag 1983, chapter 2). To simplify the notations, it is more convenient to use an approximate description in terms of continuous variables. Because the exact form of the stationary distribution can be calculated in both cases, we have been able to verify that the continuous description constitutes a reliable approximation when N is large. The continuous stochastic process is a diffusion process defined by the standard form of a Fokker-Planck equation in one dimension:

$$\frac{\partial P(f\,;t)}{\partial t} = -\frac{\partial}{\partial f}[K(f)\,P(f\,;t)] + \frac{1}{2N}\frac{\partial^2}{\partial f^2}[Q(f)\,P(f\,;t)]$$

with:

$$\begin{cases} K(f) = W_+(f) - W_-(f) \\ Q(f) = W_+(f) + W_-(f) \end{cases}$$

For $0 \leq \mu < 1$, it can be shown that the process is ergodic: whatever $P(f; 0)$ [1], $P(f; t)$ converges to a unique stationary distribution $P_{st}(f, \mu, \theta)$. K, the drift function, is central to understand the dynamics:

$$K(f, \mu, \theta) = W_+(f, \mu, \theta) - W_-(f, \mu, \theta) = -f + J(f, \mu, \theta) \tag{3.2}$$

When K is positive, the probability that f increases is greater than the probability f decreases. More precisely one can associate to our stochastic process the following deterministic system:

$$\frac{df}{dt} = K(f, \mu, \theta) \tag{3.3}$$

This equation describes the deterministic path that would be observed if the fluctuations could be neglected. It is such that its fixed points are the extrema of the stationary distribution $P_{st}(f, \mu, \theta)$. A maximum (resp. minimum) corresponds to a stable (resp. unstable) fixed points. As $K(f, \mu, \theta)$ is equal to $-f + J(f, \mu, \theta)$ (equation 3.2), it is easy to find the extrema of $P_{st}(f, \mu, \theta)$. They are the solutions of the following system:

$$\begin{cases} 0 \leq f < 0.5 & J(f, \mu, \theta) = a(\theta)\,(1 - \mu) = f \\ f = 0.5 & J(f, \mu, \theta) = a(\theta) = f \\ 0.5 < f \leq 1 & J(f, \mu, \theta) = a(\theta) + [1 - a(\theta)]\,\mu = f \end{cases} \tag{3.4}$$

Because $J(f, \mu, L) = 1 - J(1 - f, \mu, H)$, we know that:

$$P_{st}(f, \mu, L) = 1 - P_{st}(1 - f, \mu, H)$$

and then, from now on, we will concentrate our attention on the case $\{\theta = H\}$.

When $p(1 - \mu)$ is greater than $(0.5 - 1/N)$, i.e. if

$$\mu \leq \mu^* = 1 - \frac{0.5 - \frac{1}{N}}{p} \simeq \frac{2p - 1}{2p} \tag{3.5}$$

$P_{st}(f)$ is a unimodal distribution (see Fig. 1.1). Its peak (I) is defined by $f_I(\mu, H) = p + (1 - p)\,\mu$ and $f_I(\mu, L) = 1 - f_I(\mu, H)$. If we term "error" the proportion of agents who make the wrong choice, we can calculate the average

1. We have assumed that $P(f; 0) = \delta(f - f(0))$, a Dirac distribution in $f(0)$, but our result holds for any distribution.

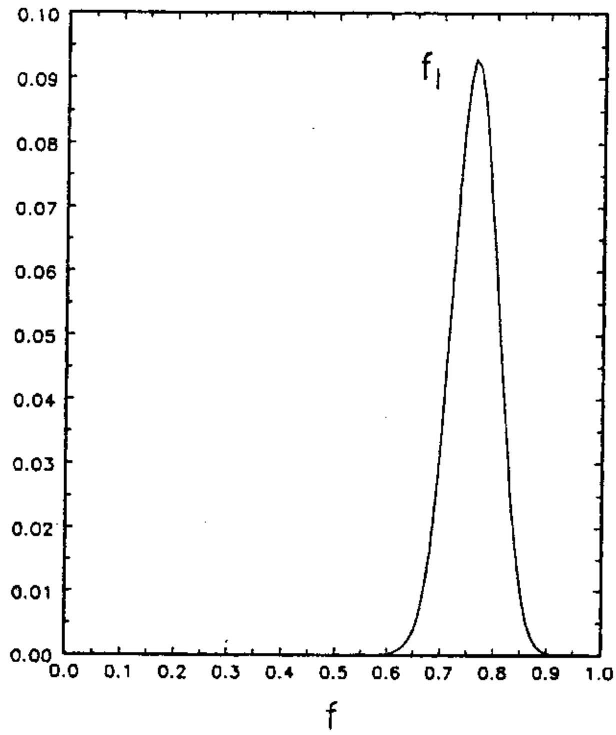

Figure 1.1: Stationary distribution $P_{st}(f)$ with $\mu = 0.2$ and $p = 0.7$

error, $\overline{E}(\mu)$. It is approximately equal to $E_I(\mu)$, the value of the error when f is equal to $f_I(\mu, \theta)$ [1]:

$$\overline{E}(\mu) = E_I(\mu) = 1 - f_I(\mu, H) = f_I(\mu, L) = (1 - \mu)(1 - p) \qquad (3.6)$$

In other words, the average proportion of individuals having made the wrong choice is equal to $(1 - \mu)(1 - p)$: the greater μ, the better the performance of the group. Being imitative is efficient: it leads to better performances than following the independent rule q_0. This result is easy to understand: through imitation, an unlucky individual who has observed $\{-\}$ when the state of the world was $\{H\}$, can nevertheless make the right choice $\{H\}$ because in making his choice he takes into account not only his private "false" information $\{\sigma_i = -\}$, but also the fact that $\{f > 0.5\}$. Imitation can be viewed as the specific manner through which global information is disseminated within decentralized information structures.

When μ becomes greater than μ^* but remains strictly inferior to 1, the shape of the stationary distribution is qualitatively affected. This corresponds to a bifurcation of the equation 3.3. A new peak (M) appears, $f_M(\mu, H) = (1 - \mu)p$. The stationary distribution becomes bimodal (see Fig. 1.2). We can calculate the error in the state f_M:

$$E_M(\mu) = [1 - p(1 - \mu)] \qquad (3.7)$$

1. When $N \to +\infty$, the variance of the stationary distribution is converging toward 0. It follows that, for N large, the stationary distribution is almost concentrated in $f_I(\mu, \theta)$.

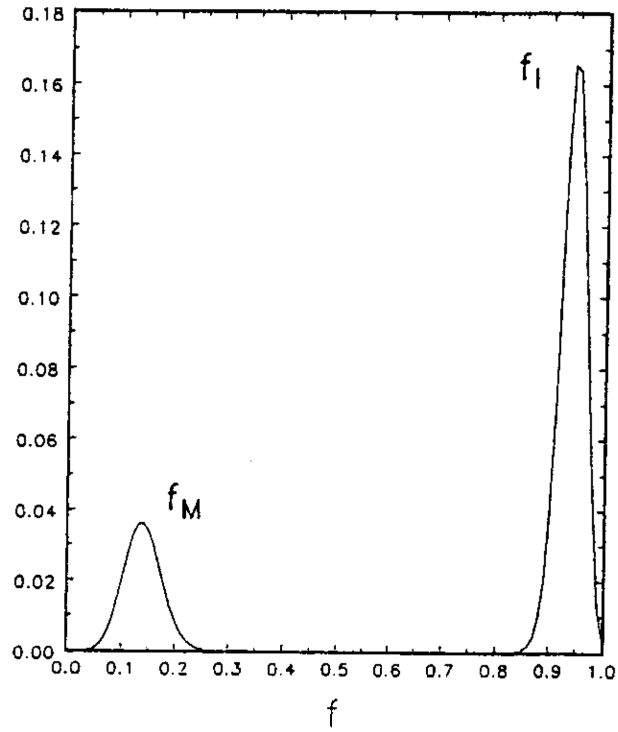

Figure 1.2: Stationary distribution $P_{st}(f)$ with $\mu = 0.8$ and $p = 0.7$

It is always greater than the average error $(1 - p)$ that would have prevailed if there was no imitation at all. The state $f_M(\mu, \theta)$ is the result of a self-validating process: because the propensity to imitate is large, the power of conformism dominates the role of the information, i.e. a large propensity of agents having observed $\{+\}$ will choose $\{L\}$ when $\{f(t) < 0.5\}$. This propensity to conform to the majority can lead the collective opinion toward a configuration where almost everybody has chosen the wrong opinion. For μ close to 0 and $\{\theta = H\}$, we can observe a quasi unanimity on $\{L\}$!

To have an accurate understanding of the way our system behaves when μ is greater than μ^*, note that the probability of transition from one peak to the other is very very small when N is great. The transition time is proportional to e^N. It follows that the process is "quasi" non ergodic for large N: $f(t)$ remains either in the vicinity of $f_I(\mu, \theta)$ or in the vicinity of $f_M(\mu, \theta)$. For a plausible time of observation T, we shall not observe transition from one peak to the other.

Figures 2 and 3 show two simulations of our process where p is equal to 0.7 and N is equal to 100. 5 000 periods are calculated. In $(t = 0)$, $f(0)$ is supposed to be equal to 0.5, and the state $\{\theta = H\}$ is drawn. Then every 1 000 periods, the state θ is changing. In figure 2, we assume that μ is equal to 0.2 which is inferior to $\mu^* = 0.28$. In that case, the stationary distribution is unimodal: $f_I(\mu, H) = 0.76$ and $f_I(\mu, L) = 0.24$. When θ changes from $\{H\}$ to $\{L\}$, the proportion of choices $\{H\}$ is moving from the neighborhood of the mode $f_I(\mu, H)$ to the neighborhood of the mode $f_I(\mu, L)$.

In Figure 3, μ is assumed to be equal to 0.8. The stationary distribution becomes bimodal and the dynamics is then very different. In the first round

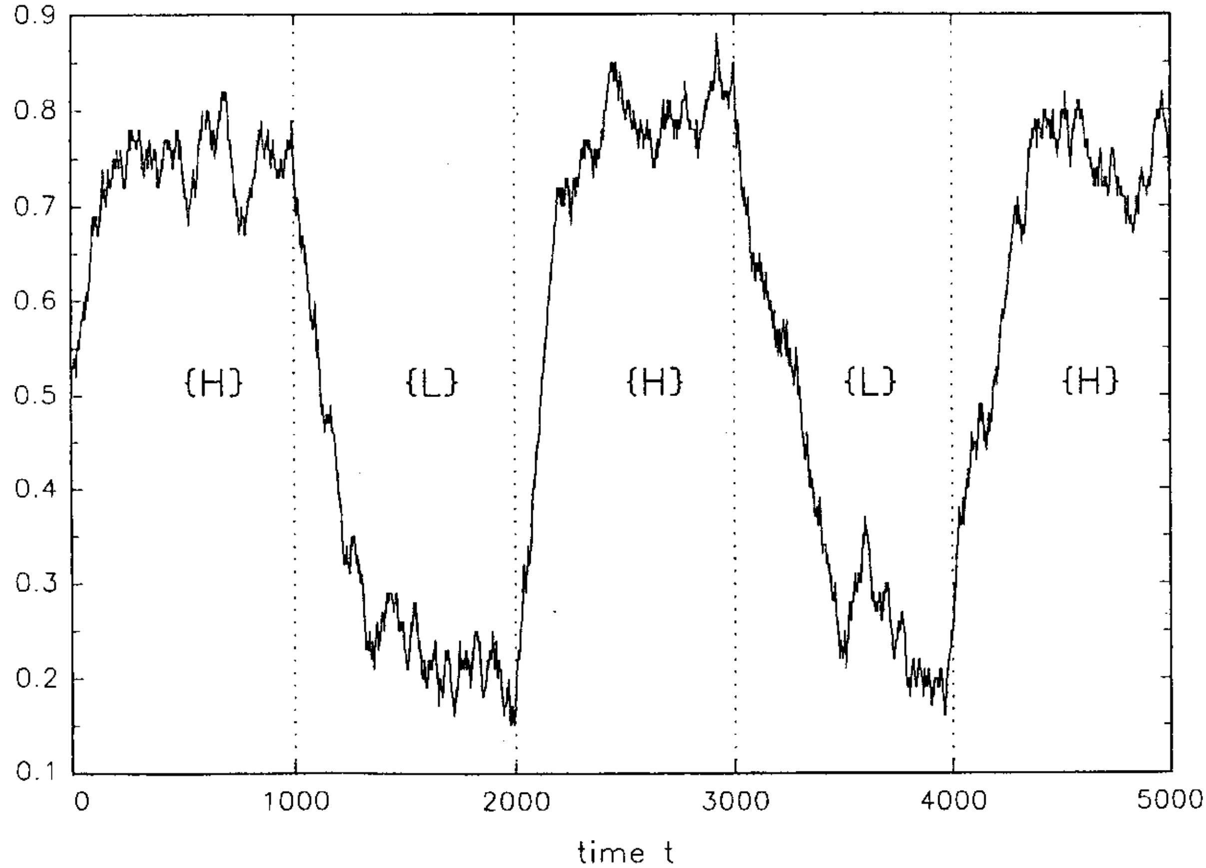

Figure 2: $f_I(0.2, H) = 0.76$ and $f_I(0.2, L) = 0.24$

($t < 1\,000$), $f(t)$ is converging to the neighborhood of the right mode $f_I(\mu, H)$. Its value is equal to 0.94 [1]. Almost every agent has found the right value of the state θ. The average error is inferior to the one prevailing in the previous case. But when θ changes from $\{H\}$ to $\{L\}$, $f(t)$ does not converge anymore to $f_I(\mu, L)$, but to the wrong mode $f_M(\mu, L) = 0.86$. In that state, only 14% of the population makes the right choice $\{L\}$. When the propensity to imitate is large, the changes in the state of the world have only a small impact on the collective choice because agents give an insufficient weight to their private information relatively to f. The majority side of the population remains unchanged.

When $\mu = 1$, the process is strictly non ergodic: $\{f = 0\}$ and $\{f = 1\}$ are absorbing states. There are two stationary distributions: δ_0 and δ_1, the Dirac distributions in $\{f = 0\}$ and $\{f = 1\}$.

4. The ambivalence of imitation

Imitation is ambivalent: to imitate is efficient only if the average propensity to imitate is small; it is getting counter-productive otherwise. This result is quite

1. Of course, we could have observed the convergence of $f(t)$ to the neighborhood of the other mode $f_M(\mu, H)$. In that case, when θ changes from $\{H\}$ to $\{L\}$, $f(t)$ goes to $f_I(\mu, L)$.

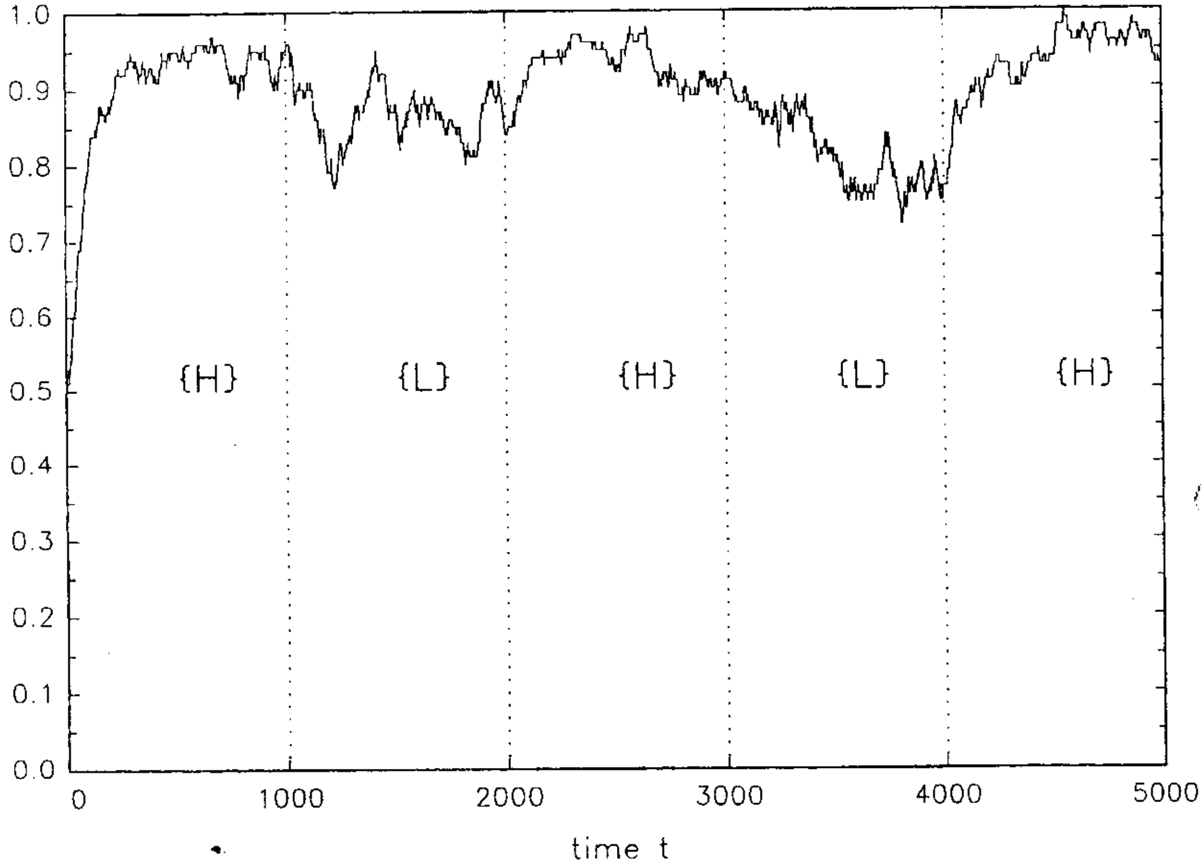

Figure 3: $f_I(0.8, H) = 0.94$ and $f_M(0.8, L) = 0.86$

intuitive: imitation is efficient if the individual I imitate is well-informed; it is not if the individual I imitate is himself an imitator. More precisely, we have shown that, when the propensity to imitate μ is smaller than μ^*, being imitative increases the accuracy of individual opinions: through imitation individuals have access to the global information which allows the agents having observed the "wrong" information to correct their error. Nevertheless, imitation is efficient only if the collective opinion embodies enough information. If μ is getting greater than μ^*, imitation gives rise to a self-validating dynamics where collective opinion becomes deconnected from fundamental information. This result is summarized in Figure 4[1]. It shows that there are two types of imitative processes, a positive one associated with $f_I(\mu, \theta)$ in which the error $E_I(\mu)$ is a decreasing function of μ; and a negative one associated with $f_M(\mu, \theta)$ where the error $E_M(\mu)$ is an increasing function of μ.

1. Figure 4 is also the bifurcation diagram of the equation 3.3.

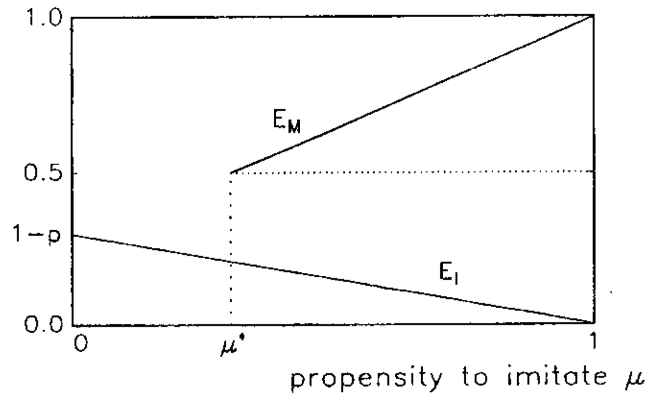

Figure 4: E_I and E_M

We have already seen that, for N large and $\mu \leq \mu^*$, the expected average error $\overline{E}(\mu)$ is close to $E_I(\mu)$ (equation 3.6). When $1 > \mu > \mu^*$ and N is large, it is more difficult to evaluate $\overline{E}(\mu)$ because the mathematical expectation:

$$\begin{cases} \int (1-f) \cdot \mathrm{d}P_{st}(f, \mu, H) & \text{if } \{\theta = H\} \\ \int f \cdot \mathrm{d}P_{st}(f, \mu, L) & \text{if } \{\theta = L\} \end{cases}$$

would be a good evaluation only if the dynamics could be observed during an infinite length of time in order to observe a great number of transitions from one peak to the other. For a plausible time T of observation, the following quantity will be a "satisfying" approximation:

$$\overline{E}(\mu) = \frac{E_M(\mu) + E_I(\mu)}{2} = (1-p) + \frac{(2p-1)\mu}{2} \geq (1-p)$$

because, in the situation under consideration, $f(t)$ oscillates only between neighborhoods of $f_I(\mu, H)$ and $f_M(\mu, L)$, (respectively between $f_I(\mu, L)$ and $f_M(\mu, H)$), as it has been illustrated by figure 3, with $\{H\}$ and $\{L\}$ being equiprobable. This measure is better suited to our problem because the probability to observe a transition from $f_I(\mu, \theta)$ to $f_M(\mu, \theta)$ for the given time T is negligible. In other words, the probability for the majority to be on the right side, either $\left\{f > \frac{1}{2}\right\}$ if $\{H\}$ or $\left\{f < \frac{1}{2}\right\}$ if $\{L\}$, is close to 1/2.

It follows that the minimum value for the average error $\overline{E}(\mu)$ is obtained for $\mu = \mu^*$. If the agents were acting cooperatively, they would decide collectively to choose the decision rule q_{μ^*}. In that situation, the average collective error would be:

$$E_{\min}(p) = \overline{E}(\mu^*) = \frac{1-p}{2p}$$

which is a decreasing function in p. The more the signal σ is accurate, the smaller is the average collective error.

What happens if each agent i can choose independently his degree $\mu(i)$ of imitation? Let us now consider the non-cooperative game defined as follows: (i) the N agents are the N players; (ii) each agent i has to choose a strategy $\mu(i)$ with $\mu(i)$ belonging to [0, 1]; (iii) the payoff for playing the strategy $\mu(i)$ is given by the probability of making the right choice when the stationary distribution is obtained. Let us examine if there exists a Nash equilibrium for this game. Before addressing this question, it should be noted that if every agent chooses a propensity $\mu(i)$ to imitate, the global process which is obtained is the one defined by $J(f, \mu_m, \theta)$ (equation 3.1), with μ_m being the average value of the distribution $\{\mu(i)\}_{i=1,\dots,N}$. This result is easy to prove if one remarks that:

$$\sum_{i=1}^{i=N} \frac{1}{N} \cdot q_{\mu(i)}(f, \sigma_i) = q_{\mu_m}(f, \sigma_i) \quad \text{with} \quad \mu_m = \frac{1}{N} \sum_{i=1}^{i=N} \mu(i)$$

with $1/N$ being the probability to draw agent i at time t.

Let us consider agent N's choice. He is facing a set of strategies $\{\mu(i)\}_{i=1,\dots,N-1}$ such that:

$$\mu_{-N} = \frac{1}{N-1} \sum_{i=1}^{i=N-1} \mu(i) \tag{4.1}$$

It follows that:

$$\mu_m = \frac{N-1}{N} \cdot \mu_{-N} + \frac{1}{N} \cdot \mu(N) \tag{4.2}$$

Agent N's optimal choice depends on the value of μ_{-N}. The central intuition is the following: if other individuals are mostly relying on their private information, i.e. μ_{-N} is "small", agent N's best choice is to be a full imitator, i.e. $\mu(N) = 1$, because the collective signal f is more precise that his private signal σ; whereas if other individuals are mostly imitators, i.e. μ_{-N} is "large", agent N's best choice is to ignore f, i.e. $\mu(N) = 0$, because the collective signal is then less precise than σ. If $\mu_{-N} = \mu^*$, then agent N's best choice is $\mu(N) = \mu^*$. $\{\mu(i) = \mu^*, \forall i\}$ is the only Nash equilibrium. To prove this statement, we have to take into account the fact that the value of μ_m is depending on agent N's choice, $\mu(N)$ (equation 4.2).

First let us consider small values of μ_{-N}, i.e. values such that:

$$\frac{N-1}{N} \cdot \mu_{-N} + \frac{1}{N} \leq \mu^* \quad \text{which is equivalent to} \quad \mu_{-N} \leq \frac{N\mu^* - 1}{N-1} \tag{4.3}$$

In that case, agent N's best choice is $\mu(N)=1$. The proof of this result goes as follows. Because of inequality 4.3, μ_m will remain inferior or equal to μ^*, whatever the value of $\mu(N)$ chosen by agent N. This implies that the stationary distribution $P_{st}(f, \mu_m, \theta)$ will always be unimodal. Then it is easy to calculate the payoff of the strategy μ, i.e. the probability of making the right choice. When $\theta=\{H\}$, the right choice is made with probability 1 when $\{+\} \cup \{f_{st} \geq 0.5\}$ is observed; with probability $(1-\mu)$ when $\{+\} \cup \{f_{st} < 0.5\}$ is observed; and with probability μ when $\{-\} \cup \{f_{st} > 0.5\}$ is observed. Then the payoff m is equal to:

$$m(\mu|H) = pB + (1-\mu)pA + \mu(1-p)C \approx p + \mu(C-p) \qquad (4.4)$$

with:

$$A = \text{Prob}(f_{st} < 0.5), B = \text{Prob}(f_{st} \geq 0.5) \text{ and } C = \text{Prob}(f_{st} > 0.5) \approx 1 - A$$

when $\theta=\{H\}$. Because $C>p$, this payoff is an increasing function in μ. Agent N's optimal choice is then $\mu(N)=1$. When f is revealing an important part of the collective information, imitation is the best strategy.

Secondly, let us consider large values of μ_{-N}, i.e. values such that:

$$\frac{N-1}{N} \cdot \mu_{-N} > \mu^* \quad \text{which is equivalent to} \quad \mu_{-N} > \frac{N\mu^*}{N-1} \qquad (4.5)$$

Whatever the value of $\mu(N)$, μ_m will remain superior to μ^*. The payoff is still given by equation 4.4. Because of the quasi non ergodicity of the stochastic process, for "plausible" values of T, the quantity C is now depending on $f(0)$. As it has been emphasized previously, a rough estimation of C is 1/2 [1]. It follows that agent N's best strategy is $\mu(N)=0$. Because the collective signal has become a "bubble", the private information σ is more reliable and should be privileged.

If $\frac{N\mu^*-1}{N-1} < \mu_{-N} \leq \frac{N\mu^*}{N-1}$, agent N's best choice is such that μ_m is equal to μ^*:

$$\mu(N) = N\mu^* - (N-1)\mu_{-N}$$

Thus $\{\mu(i)=\mu^*, \forall i\}$ is the only Nash equilibrium of our game. For N large, the ability for the group to converge on this equilibrium seems very implau-

1. Another way to justify this approximation is to consider the drift term $K(f)$ and the equation 3.3. For $\mu > \mu^*$, this deterministic equation has two stable fixed points: f_l and f_M. It converges to f_l if $f(0) < 0.5$; it converges to f_M if $f(0) > 0.5$. For N large, the drift term dominates the process. The greater μ, the more this approximation holds. For $\mu = 1$, the process is non ergodic and our evaluation of C is exact.

sible [1] but for the special case where agents know the exact value μ^*. More probably, we will observe cycles. At the beginning of the cycle, agents mostly rely on their private information, μ_m is close to 0. Because agents learn that the collective signal reveal an important part of the information, they will give a larger weight to f in their decision. μ_m will then increase. The agents who have chosen a large μ obtain better performances than others (equation 4.4). This situation generates a strong incentive to become more imitative. In a first step, this process is collectively positive because it improves the collective efficiency. But when μ goes beyond the threshold μ^*, the collective efficiency suddenly decreases. This new situation is not immediately perceived by the agents. But after a while they understand that the collective signal has become a "bubble": it does not reveal fundamental information anymore, but is the consequence of collective imitation. Then μ will decrease rapidly. This kind of cyclical process can explained certain features of financial dynamics where a succession of "normal" and "pathological" periods can be observed.

5. Conclusion

The decision rule $q_\mu(f, \sigma_i)$ we have considered until now is particular. It seems plausible that the propensity to imitate is an increasing function of f. If f is close to 0, the probability to choose $\{L\}$ can be greater than when f is equal to 0.45. We can even assume that when $\{+\}$ and $\{f > 0.5\}$ are observed some agents can choose $\{L\}$ because they wrongly interpret the signal σ. Will our result on the ambivalence of imitation remain true for a different family of decision rules? To answer this question, we have to determined the properties that rational private decision rules must satisfy. If we note s the degree of confidence in f, and $q(f, \sigma_i, s)$, the probability to choose $\{H\}$, the general family of decision rules $q(f, \sigma_i, s)$ must verify the following properties:

$$\begin{cases} \text{if } f \in \left]\frac{1}{2}, 1\right[& q(f, \sigma_i, s) \text{ is a decreasing function in } s \\ \text{if } f \in \left]\frac{1}{2}, 1\right] & q(f, \sigma_i, s) \text{ is an increasing function in } s \end{cases} \tag{5.1}$$

Thus the family of decision rules $q(f, \sigma_i, s)$ indexed by s will verify:

$$\begin{cases} q(f, \sigma_i, s) \text{ satisfies the condition } (5.1) \\ q(f, \sigma_i, s) \text{ is an increasing function in } f \\ q(f, \sigma_i, 0) = q_0(\sigma_i) \quad (\text{equation } 2.4) \\ q(f, -, s) = 1 - q[(1-f), +, s] \quad (\text{Symmetry}) \end{cases} \tag{5.2}$$

1. A more suggesting and intuitive argument can be proposed. If agent N believes that his action has no effect on μ_m (agents are price-takers), the preceeding analysis leads him to choose $\mu(N) = 1$, if $\mu \leq \mu^*$ and $\mu(N) = 0$ if $\mu > \mu^*$. In such a situation, there is no Nash equilibrium at all.

$q(f, \sigma_i, s)$ corresponds to $q_\mu(f, \sigma_i)$ within our model. Then we can show that if this family of rules verifies:

$$\exists \epsilon \text{ such that } \forall f \in [0, \epsilon[, \quad \lim_{s \to +\infty} q(f, \sigma_i, s) = 0$$

imitation remains ambivalent. In others words, the stationary distribution has always a good peak, whatever the value of s; and a new peak will appear on the wrong side for large values of s. Depending on the form of q, more than two peaks may exist.

REFERENCES

[1] Akerlof George A., "A Theory of Social Custom, of Which Unemployment May Be One Consequence", *Quarterly Journal of Economics* 94 (June 1980): 599-617.

[2] Arthur W. Brian, "Competing Technologies, Increasing Returns, and Lock-in by Historical Events", *Economic Journal* 99 (March 1989): 116-31.

[3] Arthur W. Brian, and Lane D. A., "Information Contagion", *Structural Change and Economic Dynamics* 4 (1993): 81-104.

[4] Asch Solomon E., "Effects of Group Pressure upon Modification and Distorsion of Judgment". In *Groups, Leadership, and, Men*, edited by H. Guetzkow, Pittsburgh: Varnegie Press, 1951.

[5] Axelrod Robert, *The Evolution of Cooperation*, New York: Basic Books, 1984.

[6] Banerjee Abhijit V., "A Simple Model of Herd Behavior", *Quarterly Journal of Economics* 107 (August 1992): 797-817.

[7] Bikhchandani Sushil; Hirshleifer David and Welch Ivo, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economy* 100, no. 5 (1992): 992-1026.

[8] Boyd Robert and Richerson Peter J., *Culture and the Evolutionary Process*, Chicago: Chicago University Press, 1985.

[9] —, "Rationality, Imitation, and Tradition". In *Nonlinear Dynamics and Evolutionary Economics*, edited by Richard H. Day and Ping Chen, New York, Oxford: Oxford University Press, 1993.

[10] Deutsch Morton and Gerard Harold B., "A Study of Normative and Social Influences upon Individual Judgment", *Journal of Abnormal and Social Psychology* 51 (1955): 629-36.

[11] Grossman Sanford J., "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information", *Journal of Finance* 31 (May 1976): 573-85.

[12] Grossman Sanford J. and Stiglitz Joseph E., "On the Impossibility of Informationally Efficient Markets", *The American Economic Review* 70, no. 3 (June 1980): 393-408.

[13] Kindleberger Charles P., *Manias, Panics and Crashes*, New York: Basic Books, 1978.

[14] Leibenstein Harvey, "The Prisoners' Dilemma in the Invisible Hand: an Analysis of Intrafirm Productivity", *The American Economic Review* 72 (May 1982): 92-97.

[15] Kirman Alan, "Ants, Rationality and Recruitment", *Quarterly Journal of Economics* 108 (February 1993): 137-56.

[16] Nelson Richard R. and Winter Sidney G., *An Evolutionary Theory of Economic Change*, Cambridge (Mass.) and London: Belknap Press of Harvard University Press, 1982.

[17] Scharfstein David S. and Stein Jeremy C., "Herd Behavior and Investment", *The American Economic Review* 80 (June 1990): 465-79.

[18] Shiller Robert J., *Market Volatility*, Cambridge (Mass.) and London: The M.I.T. Press, 1989.

[19] Weidlich W. and Haag G., *Concepts and Models of a Quantitative Sociology*, Berlin, Heidelberg, New York: Springer-Verlag, 1983.

[20] Young Peyton H., "The Evolution of Conventions", *Econometrica* 61, no. 1 (January 1993): 57-84.