IS WHAT IS GOOD FOR EACH BEST FOR ALL? LEARNING FROM OTHERS IN THE INFORMATION CONTAGION MODEL

David Lane
Department of Political Economy
University of Modena
lane@unimo.it

Suppose you are thinking about seeing a movie. Which one should you see? Even if you keep up with the reviews in the newspapers and magazines you read, you will probably also try to find out which movies your friends are seeing -- and what they thought of the ones they saw. So which movie you decide to see will depend on what you learn from other people, people who already went through the same choice process in which you are currently engaged. And after you watch your chosen movie, other people will ask *you* what film you saw and what you thought of it, so your experience will help inform their choices and hence their experiences as well.

Now suppose you are a regular reader of *Variety*, and you follow the fortunes of the summer releases from the major studios. One of the films takes off like a rocket, but then loses momentum. Nonetheless, it continues to draw and ends up making a nice profit. Another starts more slowly but builds up an audience at an ever-increasing rate and works it way to a respectable position on the all-time earners' list. Most of the others fizzle away, failing to recover their production and distribution costs.

The last two paragraphs describe the same process at two different levels: the *individual-level* decision process and the *aggregate-level* market-share allocation process. The first paragraph highlighted a particular feature of the individual-level process: people learn from other people, what they learn affects what they do -- and then what they do affects what others learn. Thus, the process that results in allocation of market-share to the competing films is characterized by an informational feedback, that derives from the fact that individuals learn from the experience of others.

In this paper, I discuss a model designed to isolate the effects of this informational feedback on the market-share allocation process, at least with respect to a particular kind of choice problem and a particular underlying social structure. In the model, agents choose sequentially between two competing products, each of which is characterized by a number that measures how well the product performs (that is, its "intrinsic value"). From each of their informants, agents learn which product the informant selected and an estimate of the product's performance characteristic. The agents' social space has a simple structure: agents randomly sample their informants from the pool of agents who have already made their choice.[1]

The paper focuses on some subtle and surprising ways in which agent characteristics and connectivity structure affect the market-shares of the competing products. In particular, I describe two examples of properties that seem desirable at the individual level, but which turn out to have undesirable effects at the aggregate level: what is good for each is, in a certain sense, bad for all.

The paper proceeds as follows. In section 1, I present the information contagion model, first introduced in Arthur and Lane (1993), and consider what it might mean for an agent to act rationally in the context of this model. In section 2, I show that giving agents access to more information does not necessarily lead to a better outcome at the

---

[1] In fact, social learning is channeled by pre-existing social networks: you get your information about films from friends or at least acquaintances, not random film-goers. Few papers in the economic social learning literature try to deal with this fact. An exception is Ahn and Kiefer (1995), whose agents base their adoption decisions on information obtained from their neighbors on a d-dimensional lattice.

aggregate level: *increasing* the number of informants for each agent can *decrease* the proportion of agents that end up adopting the better product. In section 3, I compare two different procedures for integrating data obtained from informants into a choice between the two products. One of these procedures uses Bayesian updating and maximizes expected utility; the other uses a rule-of-thumb based on an insufficient statistic to estimate the true performance characteristics. It turns out that the rule-of-thumb leads to an asymptotic market-share of 100% for the superior product, as the number of agents goes to infinity, no matter how small the actual difference between the products. In contrast, the procedure based on Bayesian updating and expected utility maximization can result in substantial market-share for the inferior product.

1. THE INFORMATION CONTAGION MODEL

In the information contagion model, a countable set of homogeneous agents chooses sequentially between two products, A and B. The relative value of the two products to agents depends on their respective "performance characteristics", real numbers c_A and c_B respectively. If agents knew the value of these numbers, they would choose the product associated with the larger one.[2]

Agents all have access to the same public information about the two products, which does not change throughout the allocation process. The process begins with an initial "seed" of r A-adopters and s B-adopters. Successive agents supplement the public information by information obtained from n randomly sampled[3] previous adopters. They translate their information into a choice between A and B by means of a decision rule D. Here and elsewhere, agents are homogeneous: they differ only with respect to the information they obtain from their sampled informants.

From each sampled informant, an agent learns two things: which product the informant adopted and the value of that product's performance characteristic perturbed by a standard normal observation error. That is, from informant i, agent j learns the values of two random variables, X_{ji} and Y_{ji}. X_{ji} takes values in {A,B}; Y_{ji} is a normal random variable, with mean $c_{X_{ji}}$ and variance 1. Given $\{X_{j1},\ldots,X_{jn}\}$, $Y_{j1},\ldots Y_{jn}$ are independent. Finally, $Y_1,Y_2,\ldots$ are independent random vectors, given $\{X_1,X_2,\ldots\}$.

A particular information contagion model is specified by two real parameters c_A and c_B; three integer parameters r, s and n; and a decision function D from $\{\{A,B\}xR\}^n \to \{A,B\}$.[4] For any such specification, it can be shown by a simple modification of arguments in Arthur and Lane (1993), based on results from Hill, Lane and Sudderth (1979), that the proportion of adopters of product A converges with probability 1. Moreover, the support of its limiting distribution is {x: x in [0,1], f(x) = x and f'(x) ≤ 1}, where f is given by

$$f(x) = \sum_{i=0}^{n} p(k) \binom{n}{k} x^k (1 - x)^{n-k}$$

---

[2] The prices of the two products do not enter explicitly into the model. If we suppose that these prices are known and fixed throughout the market-share allocation process, then we can suppose that the performance characteristics are measured in a way that adjusts for price differences.

[3] Asymptotically, it does not matter whether the sampling is with or without replacement. Obviously, sampling without replacement is more consistent with the underlying interpretative narrative. However, when I calculate market-allocation distributions after a finite number of adoptions in the next section, I will use sampling with replacement. Among other advantages, I do not then have to worry about whether n is greater than r+s!

[4] Since the agents are homogeneous and share the same public information, the effect of that information on their decision is just incorporated into the form of D. Agents' decisions can differ only on the basis of the private information obtained from sampled informants.

and p(k) is the probability that an agent who samples k previous purchasers of A -- and hence (n-k) previous purchasers of B -- will choose product A; p(k) depends on c_A, c_B, n and D, but not r and s.

Note: In the rest of the paper, assertions about the limiting distribution for the number of adopters of each product are based upon calculating p and then finding the roots of f(x) - x.

RATIONALITY IN THE INFORMATION CONTAGION MODEL A conventionally rational agent in the information contagion model would begin by coding the publicly available information about product performance into a prior distribution for (c_A,c_B). After he observed the value of (X,Y) from his informants, he would compute his posterior distribution for (c_A,c_B):

$$\text{post}(c_A, c_B) \propto \text{Lik}_{(X,Y)}(c_A, c_B)\ \text{prior}(c_A, c_B),$$

where

$$\begin{aligned} Lik_{(X,Y)}(c_A,c_B) &= P[(X,Y)|(c_A,c_B)] \\ &= P[Y|X,(c_A,c_B)]\ P[X|(c_A,c_B)] \end{aligned}.$$

Finally, he would compute his expected utility for c_A (c_B) using his marginal posterior distribution for c_A (c_B), and then would choose the product with the higher expected utility.

Assuming that the agent (correctly) models the components of Y as independent random variables each equal to the appropriate performance characteristic perturbed by a standard normal error, the first factor in the likelihood function is

$$P[y|\,x,(c_A,c_B)] = \prod_{i=1}^{n} \varphi(y_i - c_{x_i})$$

where φ is the standard normal density function.

The second likelihood factor, $P[X|(c_A, c_B)]$, is more problematic. Let n_A be the number of A-adopters among an agent's informants. By exchangeability, $P[X|(c_A, c_B)]$ is proportional to $P[n_A|(c_A, c_B)]$ as a function of (c_A,c_B). How might our agent generate a probability distribution for n_A, given c_A and c_B? Let R and S represent the number of agents who have already adopted A and B respectively. The distribution of n_A given R and S is either hypergeometric or binomial, depending on whether we suppose sampling is with or without replacement. So thinking about the distribution for n_A given c_A and c_B is equivalent to thinking about the distribution of R and S given c_A and c_B -- and that means modeling the market-share allocation process. To do that, agents have to make assumptions about the other agents' sources of information and about the way in which they process that information to arrive at their product choices. How might we imagine the agents do this, in such a way that they all arrive at the same answer -- that is, that agents' sources of information and decision processes become common knowledge? There are only two possibilities: either this common knowledge is *innate*, or it is a product of *social learning* , since knowledge *about* others must be learned through interaction *with* others. The idea that common knowledge of decision processes is innate is implausible; that common knowledge of others' information sources is innate is impossible. So to suppose that our agents share a model of the market-allocation process is to assume the successful accomplishment of a particularly complicated feat of social learning, much more complicated than the one the information contagion model is designed to study.

Basing a model that purports to study the effects of social learning in a restricted context on an assumption that presupposes we know its effects in a grander one seems inadmissibly circular to me.[5] If, on the other hand, we abandon the idea that the agents have common knowledge of each other's information sources and decision processes, but that each has his own private way of modeling these things, then how can we expect the agents to model the market-allocation process at all, depending as it does on their beliefs about others' beliefs about others' beliefs...?[6]

These considerations lead me to suppose that our rational agent gives up the attempt to model how n_A depends on c_A and c_B. That is, I suppose that he regards his knowledge about the initial conditions r and s and about the other agents' information sources and decision processes as sufficiently vague that he just assigns a noninformative uniform distribution to n_A for all possible values c_A and c_B.[7] If he does so, his likelihood function reduces to

$$Lik_{(X,Y)}(c_A, c_B) = P[Y|X,(c_A, c_B)]$$

I think that agents who do Bayesian updating with this likelihood function and then maximize the resulting expected utility are as rational as they ought to be in the information contagion context. I will call such agents "almost-rational."

I now describe a 3-parameter family of decision rules for almost-rational agents, introduced in Arthur and Lane (1993). I will refer to these rules, and to the agents who use them, as "Arthur-Lane Bayesian optimizers" (hereafter ALBO):

ALBO prior distribution: c_A and c_B are independent and normally distributed. If we suppose, as I will henceforth, that the public information does not distinguish between the two products, then the distributions for c_A and c_B have common mean, μ, and variance, σ^2.

---

5 But apparently not to others. Banerjee (1993), Bikhchandani et al (1992), Lee (1993) and Smith and Sorenson (1995) (among others) discuss models in which agents observe the actions of every preceding adopter, so for them the problem of modeling something equivalent to n_A given c_A and c_B does not arise. Still, these authors have to make some strong common-knowledge assumptions in order for their agents to extract information from the decisions other agents make: that all agents know the private information sources (not the information itself) of all the other agents, that they all model these sources in the same way, and that they all update available information Bayesianally. In the setting of these authors' social learning models, these common knowledge assumptions are sufficient to generate rational individual-level decisions that sequentially generate the market allocation process. Smith and Sorenson (1995) are particularly attentive to how the various underlying assumptions drive the results that they and the other authors who have contributed to this literature have obtained.

Banerjee and Fudenberg (BF), on the other hand, suppose that agents obtain information only from a sample of previous adopters as in the information contagion model. As a result, they need to make more stringent rationality and common knowledge assumptions than the authors described in the previous paragraph. In their model, agents carry out their Bayesian updating by first computing a Bayesian equilibrium distribution, supposing the "rationality" of all previous adopters, and then conditioning on this distribution as a datum. BF agents, then, not only have common knowledge of agents' information sources and Bayesianity, but of each other's "desire" to be in an equilibrium state! (Of course, BF merely assume that their agents "are" in the equilibrium; I can interpret this existential claim in process terms only through an assumption the agents make that that is where they all *want* to be.) BF's Bayesian equilibrium, by the way, can only be calculated under the assumption that there are an uncountable number of agents. BF's is a rather strong, not to say bizarre, attribution of common knowledge.

6 Arthur (1995) argues that social learning that results in common knowledge of agent rationality is in fact impossible in circumstances like this.

7 If we suppose agents have countably additive prior distributions for r and s, these distributions for n_A could only be coherent (in the sense of Heath and Sudderth, 1978) if agents believed that other agents chose each product in a way that disregards the information in Y about product quality and if their prior distribution for (r,s) is symmetric. However, I suspect that the distributions might be coherent no matter what assumptions agents made about each other, if their opinions about r and s were sufficiently vague to be expressed in the form of finitely but not countably additive symmetric prior; or if their assumptions about other agents' knowledge and decision processes were sufficiently diffuse.

ALBO likelihood function: Given X and the product performance characteristics c_A and c_B, the Y's are independent normal random variables with mean the appropriate (as determined by X) performance characteristic and variance equal to 1.

ALBO utility function: ALBOs use a utility function in the constant-risk-aversion family, parameterized by the nonnegative constant λ as follows:[8]

$$u(c) = \begin{cases} -e^{-2\lambda c} & \text{if } \lambda > 0 \\ c & \text{if } \lambda = 0 \end{cases}.$$

ALBO decision rule: The expected utility for product i (i = A or B) is

$$E(U_i) = -\exp[2\lambda(\lambda\, \sigma_i^2 - \mu_i)]$$

where μ_i is the mean of the agent's posterior distribution for c_A, and σ_i^2 is the variance of this distribution. Thus, the ALBO decision rule is: choose A if

$$\mu_A - \lambda\sigma_A^2 > \mu_B - \lambda\sigma_B^2,$$

and if the inequality is reversed, choose B.

Let n_i represent the number of i adopters a particular agent samples and $\overline{Y}_i$ the average value of the observations obtained from the sampled agents. Then

$$\mu_i - \lambda\sigma_i^2 = \frac{1}{n_i + \sigma^{-2}}(n_i\, \overline{Y}_i + \sigma^{-2}\mu - \lambda).$$

To analyze the market-share allocation process with ALBO agents, we need to calculate the appropriate p(k), the probability that an agent will adopt A if he samples k previous A adopters and n-k B adopters:

$$\begin{aligned} p(k) &= P(E(U_A) > E(U_B) \mid n_A = k) \\ &= P(k\, \overline{Y}_A + \sigma^{-2}\mu - \lambda > \frac{k + \sigma^{-2}}{n - k + \sigma^{-2}}[(n-k)\overline{Y}_B + \sigma^{-2}\mu - \lambda]). \end{aligned}$$

Hence,

$$p(k) = \Phi\left[\frac{(2k-n)[\sigma^{-2}(c_B - \mu) + \lambda] + k(n-k+\sigma^{-2})(c_A - c_B)}{\sqrt{k(\sigma^{-2} + n - k)^2 + (n-k)(\sigma^{-2} + k)^2}}\right],$$

where Φ is the standard normal cdf.

[8] When λ equals 0, agents are risk-neutral; the greater is the value of λ, the more risk averse are the agents.

2. MORE CAN BE LESS: INFORMATION INVERSION AND PATH DEPENDENCE

MORE INFORMATION IS GOOD AT THE INDIVIDUAL-LEVEL... Suppose you ask an agent in the information contagion model how many informants he would like to have. Clearly, his answer should depend on how much each interview cost him, in time and trouble as well as money. But suppose that observations are completely costless. Then, from a purely statistical point of view, the agent would like to have as many informants as possible: after all, the more information he gets about each product, the better he can estimate c_A and c_B -- and so the surer he can be about which is really the better product, *no matter what the values of c_A and c_B actually are*. Moreover, for an ALBO agent, additional (costless) information is always desirable on decision-theoretic grounds: his expected value for the expected utility of the product he would choose after an additional observation is always greater than the expected utility of the product he would choose on the basis of his current information.[9]

Thus, if we asked each of an infinite sequence of ALBO agents whether he would rather interview, say, 3 or 20 informants, all of them would opt for 20.

BUT NOT NECESSARILY AT THE AGGREGATE-LEVEL Now suppose we have two societies of ALBO agents, in one of which each agent interviews 3 informants and in the other, 20. According to the arguments in the previous paragraph, agents in the second society on average feel better about their choices than do their counterparts in the first society -- and with justice, since their choices are indeed better founded. Surely, then, a higher proportion of the agents in the second society that in the first will end up adopting the better product?

Not necessarily: Figure 1 shows some typical results.[10] In the comparisons summarized there, product B is slightly superior to product A (c_B-c_A = .1), the agents' prior distribution is well-calibrated for B ($\mu = c_B$), and the prior variance equals 1. The four curves in Figure 1 show how the proportion of B-adopters varies as a function of the number of informants n, for four different values of the agents' risk aversion parameter λ.

The bottom curve in Figure 1, corresponding to the case in which agents are risk-neutral ($\lambda = 0$), is strictly increasing: as we might expect, the proportion of agents adopting the superior product increases with the number of informants per agent. But for larger values of λ,[11] this monoticity no longer holds. Rather, "information inversions" occur -- for some range of values of n, *more* information available to individual agents results in *fewer* adoptions of the superior product . When $\lambda = 1$, the proportion adopting the superior product decreases as the number of informants

---

[9] The argument goes as follows. Let I_i represent the information an agent obtains from his first i informants: $I_i = \{(X_1,Y_1),(X_2,Y_2),\ldots(X_i,Y_i)\}$. Further, let A_i and B_i represent the expected utility the agent calculates for products A and B respectively on the basis of I_i, and let $U_i = \max(A_i,B_i)$. Thus, U_i is the value to the agent of his currently preferred product. It is easy to show that $\{A_j\}$and $\{B_j\}$ are martingales with respect to the σ-fields generated by $\{I_j\}$, so $\{U_j\}$ is a submartingale: $E(U_{i+1}|I_i) \geq U_i$. Since the support of Y_{i+1} is unbounded above and below, an agent will change his mind about which product he prefers on the basis of an additional observation with positive probability. Thus, the above inequality is strict.

[10] Typical in the sense that I have found information inversions corresponding to every value of the paremeters $d=c_A-c_B$, μ and σ^2 that I have investigated, for some range of values of λ. As the λ=0 curve in Figure 1 indicates, inversion is not generic -- at least in this case there was no inversion for n<1000, the largest value I calculated. (That does not rule out the possibility of an inversion somewhere: I have found examples where an inversion occurs for the first time at a very large value of n, of the order of 1000!) In any event, inversion seems to be a very general phenomenon in the information contagion model.

[11] Here, for $\lambda \geq 0.8$.

increases from 3 to 9. Thereafter, it increases. But the increase is very slow; in fact, the proportion adopting the superior product with 20 informants per agent (62.63%) is actually a little lower than it is with 3 (63.97%).

As λ increases, the information inversion becomes much more dramatic. For λ = 1.6, with four informants per agent, the superior product takes more than 96% market-share, while with *twenty*-four per agent it gets only 67.9%. As the number of informants increases beyond 24, the proportion adopting the superior product increases, but very slowly: when each agent interviews *1000* previous adopters, the superior product's asymptotic market-share is only 87.6%, more than 6% smaller than it is when each agent has only 3 informants! When λ = 1.7, the market-share of the superior product is 99.98% with 7 informants per agent, only 69.47% with 20, and hits bottom (68.98%) with 27. With 1000 informants per agent, the superior product obtains 87.7% of the market, about the same as with 2 informants and nearly 10% less than with 3.

PATH-DEPENDENCE For λ larger than 1.7, the situation becomes more complicated: for some values of n, more than one value for the limiting proportion of B-adopters is possible. That is, the market-share allocation process exhibits path-dependence. Small differences in who samples whom early on can result in very large differences in market-shares later in the process.

When λ equals 1.8, the superior product B attains essentially 100% market-share[12] if every agent has 7, 8, 9, 10 or 11 informants -- but if every agent samples 12 or 13 previous adopters, two outcomes are possible: the superior product will take the whole market asymptotically, or it will attain about an 80% market-share! The probability that B takes 100% market-share depends on the initial distribution (r,s); any value between 0 and 1 is possible. If every agent samples 14 previous adopters, there is once again only one outcome: this time, however, the superior product will claim only 76.6% of the market. For all larger numbers of informants, there is also only one possible asymptotic market-share for B. This share is a minimum for n=32, where it is 70%. For larger values of n, the limiting market-share increases, as usual very slowly, to 100%. For n = 1000, it is 87.8%.

For larger values of λ, other kinds of limiting distributions can arise. For example, if λ equals 2.2, the market-share of the superior product is 98.4% when each agent has two informants; with he has three, it is 100%; with 4 to 19 informants per agent, the superior product can have either a 100% or a *0*% market-share; with 20-22, the only possibility again is 100%; with 23-35, two market-shares are possible -- one is 100% , the other decreases as n increases, from 82.7% for n = 23, to less than 75% for n =35; thereafter, only one limiting market-share can obtain -- for 50 informants per agent, it is 73.8%; for 1000, it is 87.3%. For even larger values of λ, three possible market-shares for B are possible: 0%, 100%, or one intermediate value (always greater than 50%).

Figure 2 shows how the support of the market-share distribution depends on λ and n for a set of parameters that differs from the one considered above only in the difference between the prior mean μ and c_B (here the difference is 2, for the example of figure 1 it is 0).

Five different types of market-share distributions appear in the figure.[13] In the first type, which occurs in an irregular-shaped region in the lower left-hand corner, the

---

[12] To be precise, greater than 99.99999%.

[13] I think, but have not yet been able to prove, that these are the only types of market-share distributions that can arise with any parameterizations of the information contagion model, with one exception. Some models can lead to an absolutely continuous distribution on [0,1]: for example, is each agent adopts the product that his only informant adopted; of if the two products are in fact equal and ALBO agents with well-calibrated priors sample exactly two previous adopters.

superior product dominates the market: its limiting market-share is essentially 100%.[14] In the second type, one of the products will attain essentially 100% market-share, but either product can dominate. This type of market-share allocation always obtains for sufficiently large values of the risk aversion parameter λ, for any fixed values of the other model parameters. In the third type of market-structure, there are three possible limiting market-share distributions: essentially 100% market domination by either one of the two products, or stable market-sharing at some intermediate level (> 50% for the superior product). For the particular parameter values used here, the intermediate solution gives between 70% and 85% market-share to the superior product, for values of λ less than 1.5 and n less than 100. In the fourth type of market-structure, either the superior product dominates or there is a single intermediate level of stable market-sharing (here again the market-share of the superior product is between 70% and 85%). In the fifth type of market-structure, only stable market-sharing at one possible intermediate level is possible. This type of market structure obtains for sufficiently large values of n, when the values of the other parameters are fixed at any particular set of values.

When the values of all other parameters are fixed and the number of informants n goes to infinity, the market-share of the superior product increases to 100%. Only in this doubly-asymptotic[15] sense, is it true that the "best possible" outcome at the aggregate level can be achieved by getting more and more information at the individual level.

For a class of decision-rule closely related to ALBO, the diffuse ALBO rules, even this weak asymptotic aggregate-level optimality property for the individual-level adage "more information is better" fails. Diffuse ALBO rules employ the same updating and decision processes as ALBO rules, but with respect to the standard uniform improper prior for μ. This prior distribution yields a proper posterior for c_A (c_B) as long as there is at least one A-(B-)adopter among the agent's informants. If every informant turns out to have adopted the same product, a diffuse ALBO agent adopts the same product, regardless of what he learned about its performance. The diffuse ALBO rule corresponding to just two informants per agent has "optimal" aggregate-level effects: the limiting proportion of agents adopting the superior product is 100% -- actually, not just essentially -- no matter how small the difference between the two products.[16] This level of aggregate-level "performance" is only approached, never attained, as n goes to infinity.

FINITE NUMBER OF ADOPTERS All the results discussed above are asymptotic in the sense that they refer to an infinite sequence of adopters. Since the proportion of agents who adopt each product converges with probability one, asymptotic market-shares are easy to describe: as we have seen, the limiting market-share is often unique, and when it is not (except in very special cases), its distribution has at most 3 points. The situation after a finite number of adoptions -- say 100 -- is much more difficult to summarize succinctly. First, the market-share of the superior product after 100 adoptions is a random variable that can take on any of 101 possible values. Second, the dependence of this distribution on the parameters r and s cannot be ignored. Under these circumstances, it is reasonable to ask whether -- and how -- the "more is less" information inversion phenomenon shows up for finite-sized agent societies.

The answer to "whether" is: of course, but not uniformly in r and s. Figure 3 shows the distribution for the proportion of adopters of the superior product B, with the same values for d, μ - c_B, and σ^2 as in the Figure 1 example, λ equal to 1.6, and two different

---

[14] See footnote 12.

[15] In the number of adopters and in the number of informants per adopter.

[16] If the two products have equal performance characteristics, the limiting share of product A may be any number between 0 and 1. The market-share allocation process is a Polya urn process, so the a priori distribution for A's limiting market-share is Beta(r,s).

values of n: 3 and 20. The asymptotic market-shares of the superior product in these two cases are 93.88% and 68.04% respectively. In the example illustrated in Figure 3, the initial proportion of B-adopters is 2/3. The means of the two distributions are nearly identical: 0.712 and 0.707 for n=3 and n=20 respectively. Not surprisingly, the n=3 distribution is much more spread out than is the n=20 distribution (sd's 0.24 and 0.13 respectively). The only sense in which the information inversion phenomenon is observable in Figure 3 is in the difference in the modes of the two distributions: 1 for n=3 and about 2/3 for n=20. For the n=3 distribution, the superior product will have a market-share in excess of 90% with probability 0.28, while for the n=20 distribution, this probability is only 0.0915.

The market-share distributions change slowly with the number of adopters, of course in the direction predicted by the asymptotic results. After 5000 adoptions, the mean of the n=3 distribution has increased, but only to 0.75, while the mean of the n=20 distribution has decreased to 0.697; the probability that the superior product will have a market-share over 90% for the n=3 distribution has risen to 0.3, while that for the n=20 distribution, it has fallen to 0.0185.

Information inversion, yes; but not nearly as striking as in the infinite agent case. And if we start with more concentrated initial distributions or ones more favorable to the inferior product, it takes much larger agent societies for the aggregate-level advantage of less information to make itself evident.

A NOTE ON RELATED WORK Two other "more information/less efficiency" results arise in social learning models in the economic literature, but both differ in key respects from those I have reported here. In the model of Ellison and Fudenberg (1995), agents choose between two competing products. Each agent in the uncountably infinite population uses one of the two products, and in each time period a fixed fraction of them consider switching on the basis of information obtained from other users. In contrast with the information contagion model, in the Ellison-Fudenberg model the actual product performance characteristics change over time, in response to exogenous shocks. The agents in this model use a decision rule similar to the diffuse ALBO rule with $\lambda = 0$: they choose the product with the highest average performance in their sample, as long as they have information on both products; otherwise, they choose the product about which they have information.

Ellison and Fudenberg find that when the number of informants per agent is small, the proportion adopting the superior product converges to 1, which does not happen when the number of informants is larger. This Ellison-Fudenberg version of information inversion is driven by their model's assumed "shock" to the products' performance characteristics: a currently unpopular product can be "resurrected" in a period in which its performance is particularly high only if many agents find out about this good performance, which happens with higher probability the more informants each agent has. Thus, more informants per agent guarantee that neither product is eliminated. More subtle relations between individual- and aggregate-level do not arise in the Ellison-Fudenberg model, because their assumption of an uncountable number of agents lets the strong law of large numbers sweep the delicate probabilistic effects of feedback under the rug.

In Banerjee (1993), heterogeneous agents encounter an investment opportunity only by hearing about others who chose to invest. The yield of the investment depends on a state of the world unknown to the agents. For some of the agents, the investment is profitable no matter what the state of the world, but for others it is not. Agents update their state of the world distribution rationally when they learn that someone has invested, and then decide whether or not to invest themselves. The probability that an agent will learn about an investment in a particular time interval is an increasing function of the number of agents who have invested. The key to deciding whether to invest is the time at which one first learns about the investment.

Banerjee's process begins when a certain fraction x of agents learn about the investment opportunity *and* the true state of the world; this is a one-time revelation, not repeated for later investors who must infer the state of the world as described above.

Surprisingly, the aggregate-level efficiency (measured as the fraction of those who should invest that do, as a function of the true state of the world) does not increase with x, the amount of "hard information" injected into the system. To me, this result is interesting and counter-intuitive, but the phenomenon is different from what happens in the information contagion model, since in Banerjee's model the relevant "more information" comes in at the aggregate- not the individual-level.

If we range further afield than social learning models, we can encounter a number of results that bear a certain resemblance to what I found in the information contagion context. The general issue is: whether (and if so when) does more information at the micro-level of a multi-level system results in degraded higher-level system performance? Many examples of this phenomenon come to mind. Think of stochastic Hopfield networks (see Hertz, Krogh and Palmer, 1991, Chapter 2): their performance as associative memories is enhanced when each unit responds to its input signals from other units *perturbed by noise*. Here, I interpret "less noise" as "more information": thus, more information at the individual unit level, about the state of other units, can result in poorer aggregate- or system-level performance.[17] Or the "complexity catastrophe" that happens in the NK model, for sufficiently large values of the connectivity parameter K (Kaufmann, 1992). It would be interesting to speculate about whether there is some common underlying processes linking these different phenomena.

3. RATIONALITY, SUFFICIENCY AND EFFICIENCY: MAX DOMINATES ALBO

In section 1, I argued that ALBO agents were as rational as we have a right to expect agents to be in the information contagion context. In section 2, I showed that the aggregate-level "efficiency" -- the proportion of agents that adopts the superior product -- of ALBO rules depends in a complicated way on the model parameters. In particular, the market-share of the superior product can be substantially less than 100%, at least when the real difference between the products' performance characteristics is not too large.[18] In this section, I introduce another decision rule, which I call the Max rule. The Max rule has the remarkable property that it always leads to 100% market-share for the superior product, no matter how small the difference between the two products. Thus, from a social efficiency perspective, it dominates ALBO rules, however attractive the Bayesian updating and maximizing expected utility prescriptions may be at the individual-level. Moreover, the Max rule violates one of the central canons of statistical inference and decision theory: it is based on a function of the data (X,Y) that is not sufficient for the unknown parameters c_A and c_B!

---

[17] In Hopfield networks, too much noise always results in poor performance, while in the information contagion context very little -- but not *no* -- information at the individual-level can in some circumstances result in very good aggregate-level results.

[18] In fact, the proportion of ALBO agents that adopt the superior product can be substantially less than 1 even when the real difference between the products is arbitrarily large. For example, suppose that the agents are well-calibrated for the superior product B (that is, $\mu = c_B$) and, as usual, the public information does not distinguish between the two products, so the agents have the same prior distribution for c_A as for c_B. In such a situation, the market share for the inferior product A need not be essentially 0, no matter how large is the actual difference between the two products. The reason is simple: unlike the usual single-agent Bayesian asymptotics, in the information contagion context the effect of the prior distribution does not go away as the number of adopters increases. When the proportion of adopters of the superior product B becomes sufficiently large, an agent with high probability samples only B adopters. Then, his posterior mean for c_A is just the prior mean $\mu = c_B$. Of course, the expected value of the average of his Y-observations is also c_B, so the probability that his posterior mean for c_B is less than his posterior mean for c_A is 1/2. Thus, if he is not very risk averse and the number of his informants is not too large, he will end up adopting A with appreciable probability. For example, if λ=0, n=5 and σ^2 = 10, the inferior product will have a market-share of at least 18%, not matter how low its actual performance characteristic.

The Max rule was "discovered" in the course of an experiment carried out by Narduzzo and Warglien (1996). This experiment reproduced the context of the information contagion model, with Venetian business school students as agents. As part of their experiment, Narduzzo and Warglien conducted a protocol analysis, in which subjects explained why they preferred one product to the other. Not surprisingly, none of the subjects claimed to reach their decision by Bayesian optimization. Instead, they typically invoked one or another of four simple rules-of-thumb to account for their choices. One of these rules, by the way, corresponds to the Ellison-Fudenberg rule cited in the last section, the diffuse ALBO rule with $\lambda = 0$. Another, in complete contrast to my rationality argument in section 1, uses only the information in X and completely ignores the information in Y: choose whichever product the majority of your informants chose.[19]

The Max rule can be stated as follows: choose the product associated with the highest value observed in the sample. That is, denoting by max the index of $y_{(n)} = \max(y_1,\ldots,y_n)$, the rule is given by $D((x_1,y_1),\ldots,(x_n,y_n))=x_{max}$. Note that if all informants adopted the same product, then the max rule opts for that product, regardless of the magnitude of the observations $y_1,\ldots,y_n$.[20]

The Warglien-Narduzzo experimental subjects who invoked the max rule justified it with a curious argument. They claimed that they ought to be able to obtain product performance "as good as anyone else", and so they figured that the best guide to how a product would work for them was the best it had worked for the people in their sample who had used it. Narduzzo and Warglien (1996) describe this argument as an example of a “well-known bias in human decision-making”, which they follow Langer (1975) in calling the “illusion of control”.

Of course, this justification completely fails to take into account sample size effects on the distribution of the maximum of a set of i.i.d. random variables. Since the current market leader tends to be overrepresented in agents' samples, its maximum observed value will tend to be higher than its competitors', at least as long as the true performance of the two products are about the same. Thus, it seems plausible that in these circumstances a market lead once attained ought to tend to increase. According to this intuition, the max rule should generate information contagion and thus path-dependent market domination.

This intuition turns out to be completely wrong. In fact, for any $n \geq 2$, the Max rule always leads to a maximally socially efficient outcome:

Proposition: Suppose $n \geq 2$. If the two products in reality have identical performance characteristics, the market-share allocation process exhibits path-dependence: the limiting distribution for the share of product A is Beta(r,s). Suppose, however, that the two products are different. Then, the better product attains 100% limiting market-share with probability 1.

A proof of this proposition, from Lane and Vescovini (1996), is given in the appendix.

Why ALBO should fail to lead to market domination by the superior product is no mystery. But it is not easy to understand why the Max rule does allow agents collectively to ferret out the superior product. The proof gives no insight into the

---

19 Not surprisingly, this "imitation" rule leads to completely market domination by one or the other product. If r=s, each product has probability 1/2 of dominating. See Lane and Vescovini (1996). Kirman (1993) provides an interesting analysis of the effects of this rule when n=1, the population of agents is finite, and agents can change their product choices.

20 p(k) can be calculated for the Max rule as follows: p(n) equals 1 and p(0) = 0; for 0<k<n, p(k) is just the probability that the maximum of a sample of size k from a $N(c_A,1)$ distribution will exceed the maximum of an independent sample of size (n-k) from a $N(c_B,1)$ distribution:

$$p(k) = \int \Phi[y - (c_B - c_A)]^{n-k}\, k\, \Phi[y]^{k-1}\, \phi[y]\, dy,$$

where φ is the standard normal density function.

reason, though it does make clear that the larger is n, the faster will be the convergence to market domination. In fact, I am not sure there is a real "reason" more accessible to intuition than the fact that drives the proof: with the Max rule, the probability of adopting the inferior product given the current proportion of inferior product adopters is always smaller than that proportion. In fact, the superiority of the Max rule to ALBO rules is not particularly evident in "real time". The associated distributions for the superior product's market-share after a finite number of adoptions tend to be both more spread-out and more dependent on initial conditions (r and s) than those associated with ALBO rules with the same number of adopters and number of informants per adoption. Nonetheless, these distributions do, sooner or later, inexorably converge to point mass at 100%.

4. SUMMARY

This paper has focussed on a simple instance of a deep and general problem in social science, the problem of micro-macro coordination in a multilevel system. Behavior at the micro-level in the system gives rise to aggregate-level patterns and structures, which then constrain and condition the behavior back at the micro-level. In the information contagion model, the interesting micro-level behavior concerns how and what individual agents choose, and the macro-level pattern that emerges is the stable structure of market-shares for the two products that results from the aggregation of agent choices. The connection between these two levels lies just in the agents' access to information about the products.

The results described in Sections 2 and 3 show that mechanism at the micro-level has a determining and surprising effect at the macro-level. In particular, what seems good from a micro-level point of view can turn out to have bad effects at the macro-level, which in turn can materially affect micro-level agents. The paper highlighted two specific instances of this phenomenon. First, although individual decision-makers should always prefer additional costless information, performance at the aggregate level can decrease as more information is available at the micro-level. Second, despite the fact that individual decision-makers "ought" to maximize their expected utility, a very large population of Max rule followers will achieve a higher level of social efficiency than will a population of Bayesian optimizers.

Acknowledgements

I would like to thank the Santa Fe Institute and the Italian Ministry of Research for support for this research. I benefitted from discussions with Massimo Warglien and Alessandro Narduzzo and my colleagues at the Santa Fe Institute (especially Brian Arthur and Bob Maxfield) and the Univerity of Modena (especially Andrea Ginzburg). Roberta Vescovini, Giordano Sghedoni and Raffaella Barozzi contributed invaluable research assistance in the course of their baccalaureate thesis projects.

References

An, M and N Kiefer (1995), Local externalities and societal adoption of technologies, *J Evol Econ* 5, 103-117.

Arthur, WB (1995), Complexity in economic and financial markets, *Complexity*, 1, 20-25.

Arthur, WB and D Lane (1993), Information contagion, *Structural Change and Econ Dynamics* 4, 81-104.

Banerjee, A (1992), A simple model of herd behavior, *Quarterly J Econ* 107, 797-817.

Banerjee, A (1993), The economics of rumours, *Review of Econ Studies* 60, 309-327

Banerjee, A and D Fudenberg (1995), Word-of-mouth learning, *preprint*

Bikhchandani, S, D Hirshleifer and I Welch (1992), A theory of fads, fashion, custom, and cultural change as informational cascades, *J Pol Econ* 100, 992-1026.

Ellison, G and D Fudenberg (1993), Rules of thumb for social learning, *J Pol Econ* 101, 612-643.
Ellison, G and D Fudenberg (1995), Word of mouth communication and social learning, *Quarterly J Econ* 110, 93-125.
Heath, D and W Sudderth (1978), On finitely additive priors, coherence, and extended admissibility, *Ann Statistics* 6, 333-345.
Hertz, J, A Krogh and R Palmer (1991), *Introdution to the Theory of Neural Computation* (Addison-Wesley Press)
Hill, B, D Lane and W Sudderth (1979), A strong law for some generalized urn models, *Ann Prob* 6, 214-226.
Kaufmann, S (1992), *The Origins of Order: Self Organization and Selection in Evolution* (Oxford University Press).
Kirman, A (1993), Ants, rationality and recruitment, *Quarterly J Econ* 108, 137-156.
Lane, D and R Vescovini (1996), Decision rules and market-share: aggregation in an information contagion model, *Industrial and Corporate Change* 5, 127-146.
Langer, EJ (1975), The illusion of control, *Journal of Personality and Social Psychology* 32, 311-328.
Lee, I (1993), On the convergence of informational cascades, *J Econ Theory* 61, 395-411.
Narduzzo, A and M Warglien (1996), Learning from the experience of others: an experiment on information contagion, *Industrial and Corporate Change* 5, 113-126.
Smith, L and P Sorensen (1995), Pathological outcomes of observational learning, MIT Working Paper.

APPENDIX: PROOFS

1. ALBO agents expect to gain from more information

Let I_i represent the information an agent obtains from his first i informants: $I_i = \{(X_1,Y_1),(X_2,Y_2),\ldots(X_i,Y_i)\}$. Further, let A_i and B_i represent the expected utility the agent calculates for products A and B respectively on the basis of I_i, and let $U_i = \max(A_i,B_i)$. Thus, U_i is the value to the agent of his currently preferred product. I now show that an agent always expects to increase the value of U by obtaining information from an additional informant.

Claim 1: $\{A_i\}$ and $\{B_i\}$ are martingales with respect to the σ-fields generated by I_i.

Proof of claim: I show that $\{A_i\}$ is a martingale.

$$\begin{aligned} E(A_{i+1}|I_i) &= E(A_{i+1}1_{\{X_{i+1}=A\}}|I_i) + E(A_{i+1}1_{\{X_{i+1}=B\}}|I_i) \\ &= E(A_{i+1}|\{X_{i+1} = A\},I_i)P(X_{i+1} = A|I_i) + \\ &\quad A_iP(X_{i+1} = B|I_i) \end{aligned}$$

Thus, I need to show that $E(A_{i+1}|\{X_{i+1} = A\},I_i) = A_i$:

Let n be the number of A-adopters among the first i informants and let Z1,…Zn denote the Y-values obtained from these informants. Let a denote σ^{-2}. Then, given I_i,

$$c_A \sim N(\mu_n, \frac{1}{n + a}),$$

where

$$\mu_n = \frac{1}{n + a}(\textstyle\sum_{j=1}^{n} Z_i + a\mu).$$

Given I_i and $\{X_{i+1} = A\}$,

$$A_i = -\exp(-2\lambda(\mu_n - \frac{\lambda}{n + a}))$$

and

$$A_{i+1} = -\exp(-2\lambda(\frac{n + a + 1}{n + a} A_i + \frac{Y_{i+1}}{n + a + 1}))$$

Thus,

$$E(A_{i+1} | \{X_{i+1} = A\}, I_i) = -\exp(-2\lambda(\frac{n + a}{n + a + 1}\mu_n - \frac{\lambda}{n + a + 1}))$$

$$\sum E(\exp(-2\lambda \frac{Y_{i+1}}{n + a + 1}) | \{X_{i+1} = A\}, I_i))$$

Given I_i and $\{X_{i+1} = A\}$, $Y_{i+1} \sim N(\mu_n, \frac{n + a + 1}{n + a})$; thus the second factor above evaluates to $\exp(-2 \frac{\lambda}{n + a + 1}(\mu_n - \frac{\lambda}{n + a}))$, and so

$$E(A_{i+1} | \{X_{i+1} = A\}, I_i) = A_i.$$

Claim 2: $E(U_{i+1}|I_i) > U_i$ with probability 1: Since U = max(A,B), U is a submartingale: $E(U_{i+1}|I_i) \geq U_i$. Moreover, since the support of Y_{i+1} is unbounded, the information from the (i+1)st informant will induce the agent to change his mind about which product to obtain with positive probability: thus, the inequality is strict.

2. The max rule is maximally socially efficient

Proof: The proposition follows from the two claims below, using the results from Hill, Lane and Sudderth (1980) and Arthur and Lane (1993) cited in section 1 of this paper:

Claim 1: If $c_A = c_B$, then p(k) = k/n and hence f(x) = x for all x in [0,1]: In this case, p(k) is just the probability that the maximum of the sample of the n i.i.d. random variables $Y_1,\ldots,Y_n$ is one of the k of them whose associated X-value is A. The value of f now follows from (3.1).

Since f(x) = x, the market allocation process is a Polya urn scheme, from which the Beta(r,s) limiting distribution follows.

Claim 2: If $c_A > c_B$, then $p(k) \geq k/n$ for all k, and for 0<k<n the inequality is strict; hence f(x) > x for all x in (0,1). We now consider a sample of size k from a $N(c_A,1)$ distribution and a sample of size (n-k) from a $N(c_B,1)$; p(k) is the probability that the maximum of the combined sample comes from the first sample, which is clearly a non-decreasing (increasing for 0<k<n) function of c_A-c_B. The first assertion in (4.2) then follows from the first assertion of (4.1), which calculates p(k) when c_A-c_B = 0. Since f is just a linear combination of the p(k)'s with positive coefficients in (0,1), the inequality f(x) > x follows from the inequalities for each p(k) and the second assertion in (4.1), which calculates f when c_A-c_B = 0. The inequality f(x) > x, together with the evaluations f(0) = 0 and f(1) = 1, implies that 1 is the only value in the set {x: x in [0,1], f(x) = x and f'(x) ≤ 1}. Hence, the limiting market-share of product A is 1 with probability 1.

FIGURE 1: THE ASYMPTOTIC PROPORTION OF ADOPTERS OF THE SUPERIOR PRODUCT B: d = -0.1; $\mu = c_B$, $\sigma^2 = 1$. Each of the four curves corresponds to a different value of l: for bottom curve, λ=0, and for the others, in ascending order, λ=1,1.6,1.7.

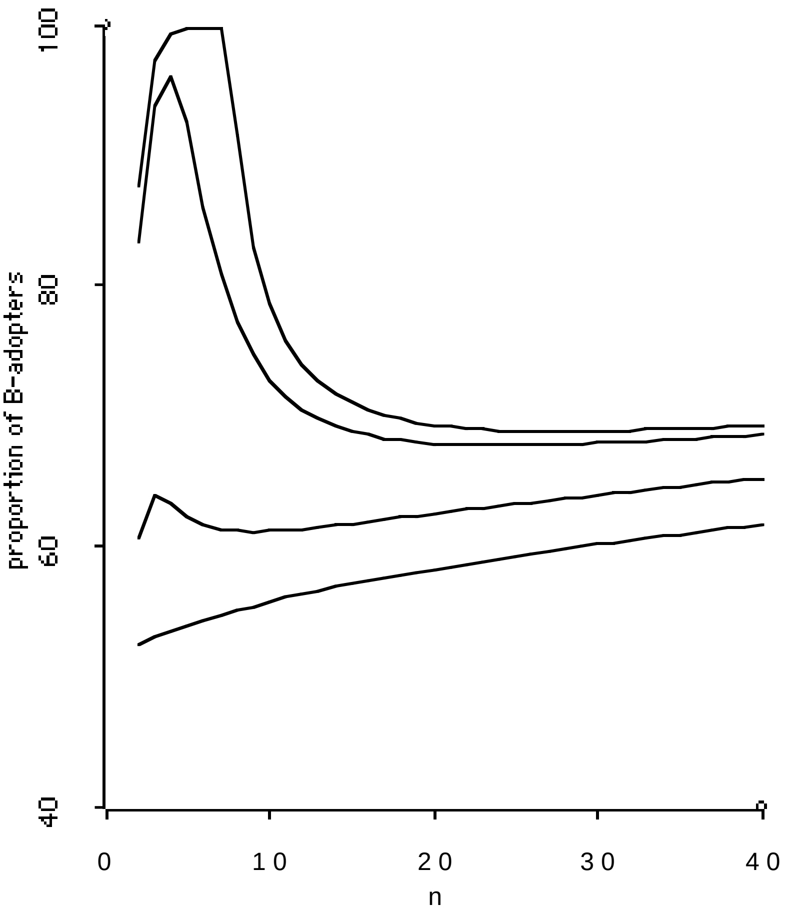

FIGURE 2: SUPPORT OF THE DISTRIBUTION FOR THE MARKET-SHARE OF THE INFERIOR PRODUCT: $d = -0.1$, $\mu = c_B+2$, $\sigma^2 = 1$.

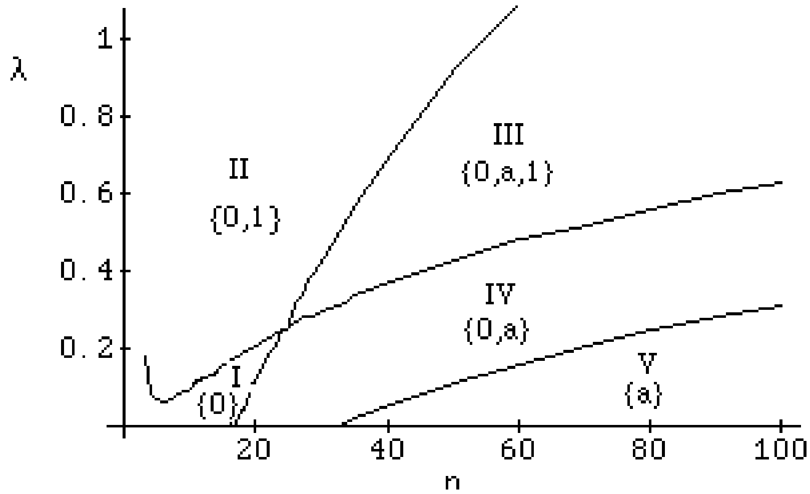

FIGURE 3: DISTRIBUTION OF THE PROPORTION OF ADOPTERS OF THE SUPERIOR PRODUCT B AFTER 100 ADOPTIONS: d = -0.1; $\mu = c_B$, $\sigma^2 = 1$, λ=1.6, r = 1, s=2. For the density with mode at 1, n=3; for the other density, n=20.