When Optimization Isn't Optimal: Aggregation and Information Contagion

David A. Lane
Roberta Vescovini

SFI WORKING PAPER: 1995-04-044

WHEN OPTIMIZATION ISN'T OPTIMAL: AGGREGATION AND INFORMATION CONTAGION

David Lane* and Roberta Vescovini
Department of Political Economy
University of Modena

Abstract

In the information contagion context, agents choose sequentially between two competing products, basing their decisions upon information obtained from a sample of previous adopters. The market shares that each product obtains depend upon the true difference in performance between the products, but also on the number of previous adopters each agent samples and the way in which agents use the sample information to guide their product choice. We highlight some surprising features of these dependencies. First, it is not socially optimal for agents to be Bayesian optimizers. In fact, we show that a simple rule-of-thumb always leads to an asymptotic market share of 100% for the better product, while Bayesian optimization can result in substantial market share for the inferior product. Second, we show that giving agents access to more information can lead to *smaller* market share for the superior product. Third, we show that the ability to predict limiting market shares for the two products, even given knowledge of how the products actually perform and how agents process information, depends upon features of agent psychology and their decision rules.

Acknowledgements: We would like the thank the Santa Fe Institute and the Italian Ministry of Research for support for our research. We benefitted from discussions with Massimo Warglien and Alessandro Narduzzo and our colleagues at the Santa Fe Institute (especially Brian Arthur and Bob Maxfield) and the University of Modena.

* Address any correspondence to lane@c220.unimo.it.

1. INTRODUCTION

Arthur and Lane (1993) introduced a model of information contagion, in which agents sequentially choose between two new, competing products. All agents have access to the same publicly available information about the performance of the two products. Unfortunately, this information fails to distinguish sharply between the two products. Consequently, each agent augments the public information with private information, obtained by sampling n previous adopters. The agent learns two things from each of the previous adopters in his sample: which product the adopter chose, along with an estimate of that product's performance.

In the information contagion model, an agent makes his choice on the basis of information supplied by previous adopters. Thereafter, the agent supplies information about the product he chose to future adopters in whose sample he is included. The model is designed to highlight the effects of this informational feedback. As a result, it excludes strategic manipulation of information or prices.

In the Arthur-Lane version of the model, the agents are Bayesian optimizers: they assimilate the information they obtain from their samples via Bayesian updating, and they decide which product to adopt by maximizing expected utility. To discover what real human beings in the information contagion context might do, Warglien and Narduzzo (1994) carried out an experiment based on the model. As part of the experiment, they 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.

In this paper, we compare what happens in the information contagion model as a function of the decision rules the agents use. Our comparisons focus on *social efficiency*, which has a natural and simple interpretation in the information contagion context. Since the model assumes that all agents have the same utility function, we can unambiguously rank the two products given their true performance characteristics, which are course unknown to the agents themselves. Thus, we can equate social efficiency with the proportion of agents that purchase the superior product. Since this random quantity converges with probability one as the number of adopting agents goes to infinity, we can base our comparisons on its limiting distribution.

Our findings were, for us, full of surprises. In fact, we have still not found satisfactory explanations for several of them. Here are our three most interesting results:

1) **Bayesian optimization is not socially optimal:** The problem facing each agent is to decide, on the available evidence, which product appears superior. It seems plausible that if each agent carries out this task "optimally" -- that is, according to the Bayesian prescriptions for opinion up-dating and decision -- then the proportion of the agents that make the *right* choice will be optimized in turn.

We know that this argument fails for the standard two-armed bandit problem, where there can be a trade-off between *exploration* and *exploitation.* The single agent in the two-armed bandit problem faces the same choice repeatedly, and as a result he may benefit by choosing the currently less attractive arm in order to acquire information that will increase his ability to discriminate between the two arms in the future. In contrast, in the information contagion context, private information obtained from sampling is not passed on from one agent to another, except in so far as it affects the choice the agent makes. Thus, there is no room for deliberate *exploration* in this context. How then could any strategy improve on optimal *exploitation* -- that is, agent-by-agent Bayesian optimization?

Though this argument may be plausible,[1] it turns out to be wrong. Under Bayesian optimization, unless the true difference between the performance of the two products is sufficiently large, the inferior product will retain some share of the market. For some specifications of priors and utility function, that share can be substantial.[2]

That Bayesian optimization does not always lead to market dominance for the superior product is hardly surprising. What is surprising is that one of the Warglien-Narduzzo rules-of-thumb does have this remarkable property. That is, if all agents use the rule, the proportion of agents who choose the superior product converges to 1, *no matter how small the difference between the two products*. Hence this rule-of-thumb always achieves at least as much social efficiency as Bayesian optimization, and out-performs Bayesian optimization over a wide swath of parameter space. What makes this result even more surpising is the fact that the "optimal" rule-of-thumb is based on a statistic that is not a function of the sufficient statistic for the unknown performance characteristics available to the agents when they make their choice!

2) **More information is not necessarily better:** Our second finding continues the theme adumbrated in the first: what is best for each is not best for all. Suppose the type of decision rule that agents use is fixed. Consider n, the number of previous adopters each agent samples. The larger is n, the more information the agents have available to them, and so the better they can estimate the true performance of the competing products. For any given agent, it is clear that the larger the number of previous adopters he samples, the higher the probability that he will select the better product.

It is therefore plausible that the larger is n, the greater the proportion of agents who will pick the better product. But for several decision rules, including Bayesian optimization, this is not necessarily the case. In fact, we will see that for some parameter values in the Bayesian optimization model, a sample size of 2 guarantees that the superior product will capture the whole market, while the inferior product will retain substantial market share if each agent samples dozens or even hundreds of previous adopters.

3) **Rule-dependence of path-dependence:** Working with Bayesian optimizing agents, Arthur and Lane showed that the informational feedback around which their model is constructed suffices to drive the market to domination by one or the other of the competing products, even when the two products are in fact identical in performance. Information contagion-driven market domination happens when agents are sufficiently risk averse or the products' performance sufficiently exceeds their prior expectations. Which product then takes over the market is determined by historical accidents -- who happened to sample whom early on, what random errors were incorporated in the observations.

We will show that Bayesian optimization can also lead to path-dependence of market share even when the products perform differently, as long as the difference is not large. For sample sizes greater than 2, Bayesian optimization can lead to at most three possible values for the limiting market share of the superior product.

The rules-of-thumb behave quite differently with respect to path-dependence. Two of them never generate path-dependence, except when each agent samples exactly two

---

[1] Attentive readers will have spotted at least one flaw in the argument: it confounds a criterion for optimality from the subjective point of view with an "objective" (that is, conditioned on the true values for product performance) measure of optimality. To us at least, though, this flaw does not detract from the argument's plausibility. As we shall see, Bayesian optimization can be beaten for ALL priors and ALL utility functions by the *same* rule-of-thumb.

[2] As we shall see in section 4 below, the relation between limiting market share for the superior product and the parameters characterizing the prior distribution and utility function is quite complicated. We will reduce a bit the number of these parameters by supposing throughout the paper that the agents assign the *same* prior distribution to the performance of the two products, in keeping with the assumption that the public information does not distinguish between them.

previous adopters. Another rule produces a continuum of possible limiting market shares when the two products are in fact identical, but if the products perform at all differently this rule cannot produce path-dependence. In contrast, the fourth rule *always* produces path-dependent market domination, no matter how different the products are from one another.

The paper proceeds as follows. In the next two sections, we describe the information contagion model and the rules-of-thumb culled from the Warglien-Narduzzo experiment. In section 4, we prove that one of the rules-of-thumb, the so-called max rule, enjoys the optimality property described above. In the next two sections, we investigate by examples the complicated relationship between Bayesian optimization and social efficiency: section 5 illustrates the kinds of asymptotic market share distributions to which Bayesian optimization can give rise, while in section 6 we show that giving Bayesian agents more information does not necessarily lead to greater social efficiency. Section 7 briefly surveys what happens with the other rules-of-thumb. In the final section, we offer some reflections on the meaning of our results.

2. THE INFORMATION CONTAGION MODEL

In this section, we describe the stochastic structure of the information contagion model and show how we analyze the limiting distribution of market share. For a discussion of the model's economic motivation and interpretation, see Arthur and Lane (1993).

The model depends on five numbers and a decision rule:

- c_A and c_B are real numbers, supposed unknown.
- n, R and S are positive integers, with $n \leq R+S$.
- the decision rule D is a function from $\{\{A,B\}xR\}^n \to \{A,B\}$.

We begin with R and S agents who have adopted products A and B respectively. The next agent -- whom we call agent 1 -- selects n of these adopters at random, without replacement. For the i^{th} agent in his sample, agent 1 observes two random variables: X_{1i} and Y_{1i}. X_{1i} takes values in {A,B} and identifies the product that the i^{th} sampled agent adopted. Y_{1i} is a normal random variable, with mean $c_{X_{1i}}$ and variance 1.[3] Given $\{X_{11},\ldots,X_{1n}\}$, $Y_{11},\ldots Y_{1n}$ are independent. Agent 1 then adopts a product of type $D((X_{11},Y_{11}),\ldots,(X_{1n},Y_{1n}))$.

For $j\geq 2$, agent j selects n of the R+S+(j-1) previous adopters. He observes (X_{ji},Y_{ji}), which has the same stochastic structure as (X_{1i},Y_{1i}), and then he determines which product type to adopt by calculating $D((X_{j1},Y_{j1}),\ldots,(X_{jn},Y_{jn}))$.

To complete the stochastic specification of the model, we suppose that $Y_1,Y_2,\ldots$ are independent random vectors, given $\{X_1,X_2,\ldots\}$.

With this specification, the adoption process can be analyzed by making use of the theory of generalized Polya urn schemes initiated by Hill, Lane and Sudderth (1980). According to this theory, the limiting market share for product A converges with probability one, and its distribution can be calculated from the function

$$(2.1) \qquad f(x) = \sum_{i=1}^{n} p(k)\binom{n}{k}x^k(1-x)^{(n-k)},$$

3 Thus, c_A and c_B represent scalar performance characteristics of products A and B respectively, and observations yield the values of these characteristics, perturbed by independent standard normal errors.

where 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. In particular, the support of the limiting distribution is the set {x: x in [0,1], f(x) = x and f'(x) ≤ 1}. Arguments for these claims are given in Arthur and Lane (1993).

Thus, our analysis in sections 4-7 below reduces to calculating p(k) for particular decision rules and parameter values, and then studying the zeros in [0,1] of the n^{th} degree polynomial f.

3. DECISION RULES

We describe five classes of decision rules, beginning with Bayesian optimization and then turning to the four rules-of-thumb reported by Warglien and Narduzzo.

a) Bayesian optimization

The decision rule we call Bayesian optimization was introduced by Arthur and Lane (1993). It presumes that agents only use the information about product quality that is contained in the observation vector Y. That is, the agents use the information in X to classify observations by product type, but do not try to extract the information about product quality that is *implied* by the choices that the sampled previous adopters have made. Arthur and Lane defend their modelling choice on the grounds that it would be infeasible to construct a model that extracts the missing information -- and certainly unrealistic to presume that such a model[4] could be regarded as common knowledge and hence shared by all agents.

The Arthur-Lane rule supposes that agents know how the observations depend on the true, unknown performance characteristics c_A and c_B. That is, they know that they observe these characteristics perturbed by independent standard normal observation errors.

The rule depends on three parameters: μ, σ and λ. The performance characteristics c_A and c_B have independent $N(\mu, \sigma^2)$ prior distributions. The agents' utility function u belongs to the family of constant risk aversion functions, parameterized by the nonnegative constant λ as follows:

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

With these specifications, we can write the expected utility for a product, A for example, as

$$E(U_A) = \mu_A - \lambda\, \sigma_A^2,$$

where μ_A is the mean of the current posterior distribution for c_A, and σ_A^2 is the variance of this distribution. The Bayesian optimization decision rule is to choose the product with the higher *posterior* expected utility.

Using i to denote either A or B, let n_i represent the number of A adopters a particular agent samples and $\overline{Y}_i$ the average value of the observations obtained from the sampled agents. Then

---

4 Which would entail assumptions about how other agents make their decisions on the basis of their assumptions about how other agents make their decisions, and so on.

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

We can now calculate p(k) as follows:

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

Using our distributional assumptions,

$$(3.1)\quad p(k) = \Phi\left[\frac{(2k-n)[\sigma^{-2}(c_B - \mu) + \lambda] + k(n-k+\sigma^{-2})(c_A - c_B)}{\sqrt{k(\sigma^{-2} + n - k)^2 + (n-k)(\sigma^{-2} + k)^2}}\right],$$

where Φ is the standard normal cdf.

b) **The max rule**

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 $D((x_1,y_1),\ldots,(x_n,y_n)) = x_{max}$.

The Warglien-Narduzzo experimental subjects who invoked this rule justified its use with hearty self-confidence. They felt 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.

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.

When all sampled products are of the same type, according to the max rule the agent must adopt that type. Thus, 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.

c) **The mean rule**

According to this rule, an agent chooses the product associated with the higher mean value in his sample. In contrast to the max rule, the mean rule is based on the sufficient statistic for the two unknown performance characteristics: the means of the respective samples. It is thus plausible that the mean rule should be more socially efficient than the max rule, since the latter rule wastes potentially useful information about the true values of the performance characteristics. The mean rule also corresponds to the limit of Bayesian optimization rules, where the limit is taken as the prior variance goes to 0 and the risk aversion parameter λ is 0.

For 0<k<n, p(k) for the mean rule is just the probability that a $N(c_A, \frac{1}{k})$ random variable exceeds a $N(c_B, \frac{1}{n-k})$ one:

$$p(k) = \Phi[(c_A - c_B)\sqrt{\tfrac{k(n-k)}{n}}].$$

d) **The min rule**

The min rule selects the product associated with the higher *minimum* value in the sample. This rule suggests a strong form of risk aversion, since agents who use it are trying to avoid the worst performance of which they have knowledge. Of course, this interpretation of the rule ignores the dependence on sample size of the distribution of the minimum of a sample of i.i.d. random variables.

We offer two conflicting intuitions about the market share process induced by the min rule. The first is based on an analogy with Bayesian optimization: the more risk averse the agents are, the more tendency there is for information contagion to drive the market to path-dependent domination by one of the products. Hence, we might expect that the min rule leads to market domination. The second intuition is based on the fact that market leadership implies over-representation in samples and hence a stochastic reduction in the value of the sample minimum for that product relative to that of its competitor, at least if the products are not very different. Thus, temporary market leads would tend to disappear, and we would expect nearly equivalent products to end up each with around 50% market share.

For the min rule, p(0)=0, p(n)=1, and, for 0<k<n,

$$p(k) = 1 - \int (1 - \Phi[x - (c_B - c_A)])^{n-k} k (1 - \Phi(x))^{k-1} \varphi(x)\, dx.$$

e) **The popularity rule**

Unlike the other rules, this rule ignores the information in Y and opts for the product that appears more frequently in the sample. As one might imagine, experimental subjects justified their choice in terms of this rule only when the information in Y failed to distinguish sharply between the two products. Because the rule depends only on the information in X, the limiting market share distribution for this rule cannot depend on the actual values of c_A and c_B.

For the popularity rule,

$$(3.2) \quad p(k) = \begin{cases} 0 & \text{if } k < \frac{n}{2} \\ 1 & \text{if } k > \frac{n}{2} \end{cases}$$

If n is even, we can define $p(\frac{n}{2})$ either as 1/2, if we suppose that the agent randomizes his choice; or as the current proportion of A adoptions in the population, if we suppose that the agent samples an additional agent to break the tie.

The obvious intuition is that the popularity rule leads to market domination, with the probability that product A ends up with 100% of the market depending on R/(R+S). For once, the obvious intuition is correct.

4. THE MAX RULE IS MAXIMALLY SOCIALLY EFFICIENT

In the following proposition, we derive the asymptotic market share distribution under the assumption that all agents use the max rule:

Proposition 1: 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.

Proof: The proposition follows from (4.1) and (4.2) below, using the results from Hill, Lane and Sudderth (1980) and Arthur and Lane (1993) cited in section 2 above:

(4.1) **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.

(4.2) **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 $\{x: x \text{ in } [0,1], f(x) = x \text{ and } f'(x) \leq 1\}$. Hence, the limiting market share of product A is 1 with probability 1.

Thus, our intuition that the max rule ought to lead to information contagion and path-dependent market domination is completely mistaken. Instead, Proposition 1 shows that the max rule leads to a socially efficient outcome, no matter what the values of c_A and c_B: the proportion of agents that have adopted the superior product always converges to 1 as the number of adoptions goes to infinity. We wish we could provide a plausible explanation for this result, but so far we have not succeeded in finding one.

5. BAYESIAN OPTIMIZATION CAN BE INEFFICIENT

Unlike the max rule, Bayesian optimization does not always lead to market domination by the superior product. In this section, we investigate how the limiting market share allocations under Bayesian optimization depend upon the agents' degree of risk aversion λ; the calibration of their prior distributions for the performance characteristics c_A and c_B; and the true but unknown difference in performance characteristics, $d = c_B-c_A$.

Arthur and Lane (1993) showed that Bayesian optimization can lead to three different kinds of market share regimes when the two products perform the same (that is, when d=0). In the first regime, the market share of both products converges to 1/2. In the second regime, there are three possible limiting values for the market share of product A: nearly 0, 1/2, and nearly 1. In the third regime, one or the other product will attain a market share essentially equal to 1.[5]

[5] If n=3, for a very limited range of parameters, another regime is possible -- the two products can have market shares p and (1-p) respectively. For each fixed value of the prior variance σ^2, each possible value p corresponds to a single line in (λ, μ) parameter space.

Which regime obtains depends on the values of the parameters (λ, c_A-μ, σ^2), the first two of which have the most important and the most easily interpretable effects. Fixing the values of the other parameters, agents move from regime 1 to 2 to 3 as either λ or c_A-μ increases. That is, when agents are sufficiently risk averse, or the products have sufficient "unanticipated effectiveness", information contagion leads to market domination by one or the other product. Only a narrow band of parameter values results in the "in-between" second regime.

We now describe what happens when the two products do not perform equally. We start with an example in which $n = 5$ and $\sigma^2 = 10$; any other fixed values for these two parameters would yield the same qualitative results (as long as $n > 2$). The asymptotic behavior of market share differs considerably for the three different regimes described in the previous paragraph:

1) Values of λ between 0 and 1.31 lead to first regime behavior: that is, both products attain 50% of the market in the limit when d=0. In this regime, for any particular values for μ and d, there is only one possible limiting market share for the inferior product, which we will take to be A.

Suppose $\lambda = 0$. Figure 5.1 shows the asymptotic market share of A as a function of d, for three different calibrations of the prior distribution. In the uppermost curve, the prior is well-calibrated for product B: that is, μ=c_B. In the lower curve, the prior is well-calibrated for A, so μ=c_A. The middle curve plots a case in which μ=c_A+1/2, so that the inferior product always performs worse than expected and the superior product does so for values of d less than 0.5, but for larger values of d it outperforms the adopters' expectations.

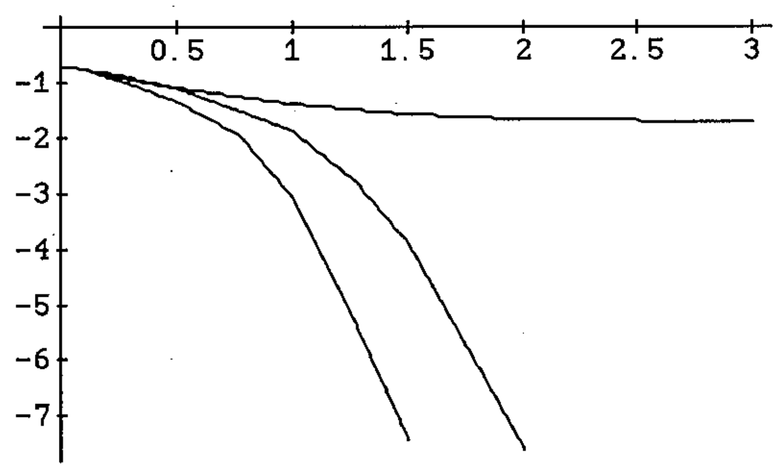

FIGURE 5.1: SOCIAL EFFICIENCY OF BAYESIAN OPTIMIZATION RULES WITH λ = 0.

Abscissa: d=c_B-c_A
Ordinate: logarithm of limiting market share of the inferior product A
Parameters: n=5,s^2=10;l=0. For the upper curve, μ=c_B; for the lower curve, μ=c_A; and for the middle curve, μ=c_A+.5.

For all three prior distributions, the inferior product A obtains substantial market share for small values of D. The dependence of limiting market share on the prior for larger values of d, though, is dramatic. If the prior is well-calibrated for A, the market share of A goes from 27% at d=.5 to 4% at d=1 to 0.06% at d=1.5. In contrast, no matter how large is d, the inferior product always gains an asymptotic market share of at least 18% if the prior is well-calibrated for the superior product B! There is a simple explanation for this initially surprising fact. When the proportion of B adopters becomes sufficiently large, an adopter with high probability has only B adopters in his sample. Then, his expected utility for A is just the prior mean, c_B, while his expected utility for B is a convex combination of c_B and the average of his observations, which will fall below c_B with probability 1/2. Thus, he is as likely to adopt A as he is to adopt B. In the information contagion context, the effects of the prior distribution do not disappear asymptotically, since the prior is "renewed" with each new adopting agent. When the prior is well-calibrated for the superior product, it imposes an upper bound on the sustainable proportion of B adopters.

As the value of λ increases, the efficiency of the resulting Bayesian optimization rule increases as well. For example, if $\lambda=1$, the phenomenon discussed in the last paragraph does not occur: when the prior is well-calibrated for the superior product B, the limiting market share for B is 1, for all values of d greater than 0.225.

2) Only values of λ between 1.31 and 1.33 produce market share allocations in the second regime. The Bayesian optimization rules corresponding to these values are very efficient. For d = 0, there are three possible limits for the market share of A: 0, 1/2, 1. Suppose now that $\lambda = 1.32$. Then, for any value of $d \geq 0.01$, the limiting proportion of A adoptions is 0. For $d < .0021$, there are three possible limit points: 0, 1, and a third, that decreases from 0.5 to 0.32 as d increases. For d between 0.0021 and 0.01, either A or B must dominate the market, and the probability that the superior product B wins increases as d increases.

3) Now we consider a value of λ in the third regime: $\lambda=2$. Surprisingly, the rule is not as efficient as the rule discussed in the previous paragraphs, at least in terms of the possibility of market domination by the inferior product. For all values of d between 0 and 0.475 (if the prior is well-calibrated for c_B) or 0.5 (if the prior is well-calibrated for c_A), either A or B can dominate the market. The probability that the inferior product A dominates of course decreases in this interval from 1/2 to 0. For values of d larger than the cut-off values 0.475 and 0.5, the market share of B converges to 1.

6. BAYESIAN OPTIMIZATION: EFFICIENCY AND QUANTITY OF INFORMATION

In this section, we describe how social efficiency depends on the quantity of information available to adopting agents. This quantity is parameterized as n, the number of previous adopters each agent samples. The relationship between n and efficiency is complicated, and we cannot yet characterize it precisely. Here, we only try to illustrate by example the kinds of relationships we have so far discovered. In all our examples, A is the inferior product.

As we explained in Section 1, it seemed clear to us when we embarked on our research that the greater is n, the lower will be the proportion of agents who end up adopting the inferior product. In one sense, this intuition turned out to be right. Fixing the values of the parameters μ, σ^2, λ, c_A and c_B, the limiting proportion of agents adopting A goes to 0 as n goes to infinity. But in another sense, the intuition is badly mistaken: as the following three examples show, more information need not lead to increased efficiency.

Example 1: Unimodal inefficiency distribution. For this example, $\sigma^2=100$, $\lambda=.5$, $c_B-\mu= 2$ and $c_B-c_A=0.1$. The limiting market shares of A for some values of n are given in Figure 6.1:

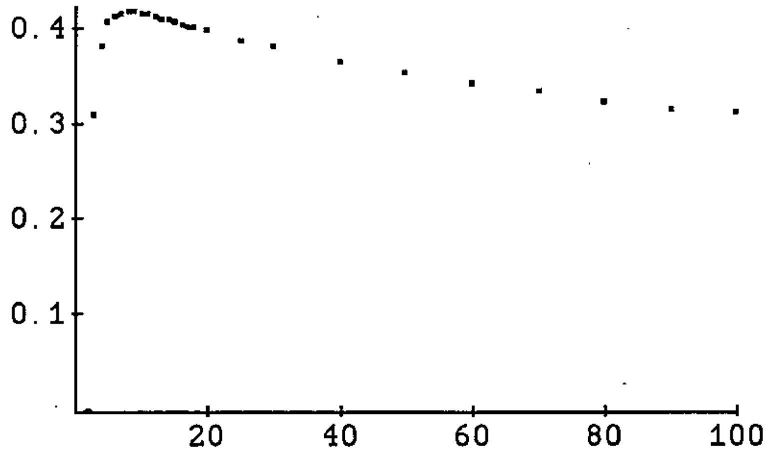

FIGURE 6.1: UNIMODAL INEFFICIENCY DISTRIBUTION
Abscissa: n
Ordinate: limiting market share of the inferior product A
Parameters: d=0.1; σ^2=100; λ=.5; c_B-μ=2.

The maximum social efficiency is attained for the minimum value, n=2 (for which the limiting market share of A is around 10^{-20}). In contrast, for n=3, 30.9% of agents adopt A. The percentage adopting A continues to rise as n increases, reaching a maximum value of 41.5% for n=8. After that, it begins to decrease, very slowly: for n=100, the percentage adopting A is 31%, greater than the value associated with n=3, and for n=200 it is 25.5%.

Note the following qualitative behavior we see in this example: 1) for each value of n, there is only one possible limiting proportion for A; 2) the greatest social efficiency occurs at the minimum value n=2; 3) there is a large jump in social efficiency at some point (here, at n=3); 4) there is a single mode in the function relating n to proportion adopting A; 5) the right-hand tail of this function is extremely long, so that the efficiency attained for small values of n is not matched until n becomes extremely (generally "unrealistically") large.

Of these, property 5) seems the most general. We have found it in all the examples we have computed in which $\sigma^{-2}(c_B - \mu) + \lambda$ is sufficiently large[6] and c_B-c_A is not so large that B dominates the market for all values of n. Example 2 below presents a common exception to property 4), while example 3 does the same for property 1).

Example 2: Bimodal inefficiency distribution. In this example, σ^2=10, λ=0, c_B-μ= 2 and c_B-c_A=1:

[6] Sufficiently large is very small, but positive. Of the examples we have computed, the largest value of this function that yields a strictly decreasing inefficiency function is 0.0005.

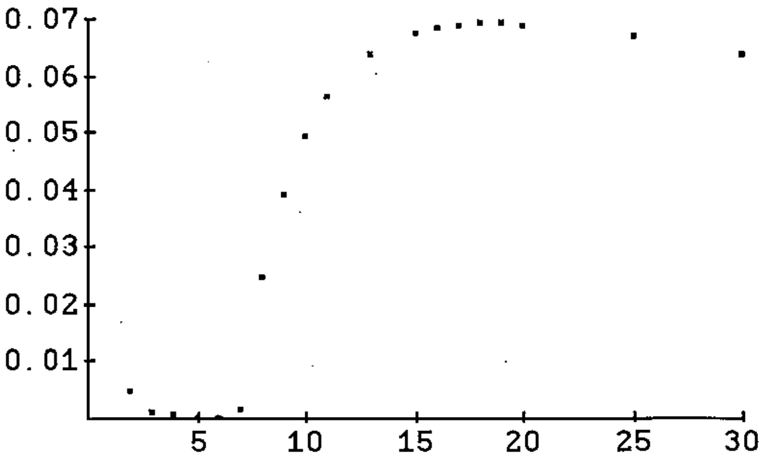

FIGURE 6.2: BIMODAL INEFFICIENCY DISTRIBUTION.

Abscissa: n
Ordinate: limiting market share of the inferior product A
Parameters: d=1; σ^2=10; λ=0; c_B-μ=2.

Note that in this case, we obtain a bimodal distribution. The most efficient outcome associated with small values of n is attained at n=6, where the limiting proportion of A is 0.0000005, in contrast to 0.0045 for n=2 and 0.0012 for n=7. The big jump here occurs at n=8, associated with a limiting proportion of 0.0245. The right-hand tail of this distribution can be seen in Figure 6.3:

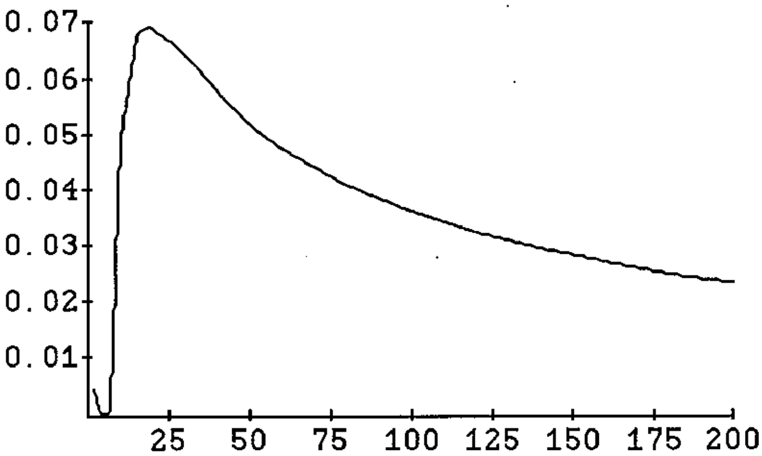

FIGURE 6.3: BIMODAL INEFFICIENCY DISTRIBUTION.

Abscissa: n
Ordinate: limiting market share of the inferior product A
Parameters: d=1; σ^2=10; λ=0; c_B-μ=2.

Example 3: Regime switches. In both the previous two examples, each value of n produced only one possible limiting proportion of A adopters. In this example, we obtain a very complicated pattern of behavior, in which all three of the regimes discussed in the previous section make their appearance. The parameter values of this example are: σ^2=10, λ=1.5, c_B-μ= 0.5 and c_B-c_A=0.1.

For n=2, A's market share is essentially 0. For n=3-6, both 0 and 1 are possible limits. The probability that A dominates the market in these cases depends on the initial proportion of A adopters; if this proportion is sufficiently high, the probability that A dominates is also high for each n between 3 and 6, but for fixed initial concentration, the probability decreases as n increases. For n=7-11, B again attains essentially 100% market share (regardless of the initial proportion of A adopters). In contrast, when n=12 or 13, two limiting market shares are possible: either 0 or 18.9% for n=12 and 22.8% for n=13. Unless R is much less than S, the larger market share for A is substantially more probable in both these cases. For n≥14, A has only one possible limiting market share. For n=14, it is 25.4%, and it continues to increase up to a maximum value of 31.3% for n=30! The market share of A decreases after 30, very slowly as usual: for n=100, A gets 27.3% of the market, and for n=200, its limiting market share is 23.1%.

7. SOME OBSERVATIONS ON THE OTHER RULES-OF-THUMB

Here we summarize the asymptotic market shares associated with the other Warglien-Narduzzo rules-of-thumb:

a) **The mean rule**

When n=2, the mean rule enjoys the same social efficiency properties as the max rule. For d=0, the market share process is a Polya urn scheme and hence the market share for A has a Beta[R,S] limiting distribution; while for d>0, the market share for the inferior product A is asymptotically 0. For larger values of n, the mean rule is less efficient. Figure 7.1 shows a nearly linear decrease in the market share of the inferior product as a function of d:

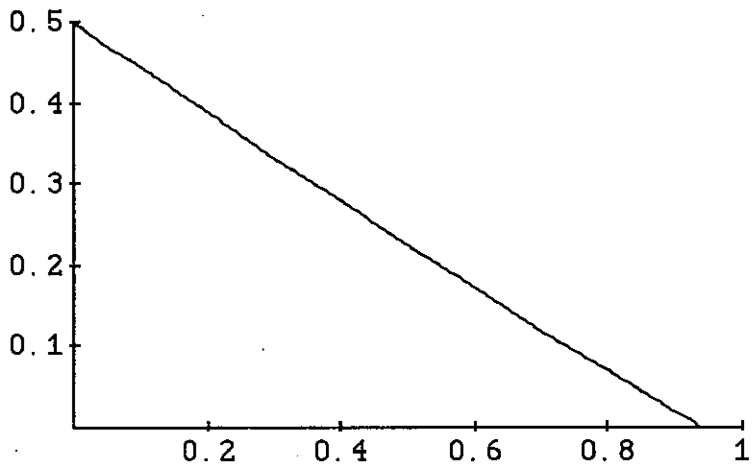

FIGURE 7.1: SOCIAL EFFICIENCY OF MEAN RULE

Abscissa: $d=c_B-c_A$
Ordinate: limiting market share of the inferior product A
Parameter: n=5

Note that for d>0.945, the limiting market share of A is 0.

Like Bayesian optimization, more information may result in decreased social efficiency for the mean rule. For example, if d=1, the limiting market share of A is 0 for n=3-5; it is 2.5% for n=6; it reaches its maximum value of 9.2% for n=14; and at n=100 it is 4%.

b) **The min rule**

When we introduced the min rule in section 2, we offered two opposing plausibility arguments. According to one of them, the min rule ought to lead to information contagion. According to the other, it should lead to relatively stable market sharing, even when the products are quite different. The second argument turns out to be right, at least for n>2.

For n=2, the min rule behaves like the mean rule and the max rule. For n>2, the min rule is less efficient than the mean rule. Figure 7.2 shows how the limiting market share of the inferior product depends on d for n=5:

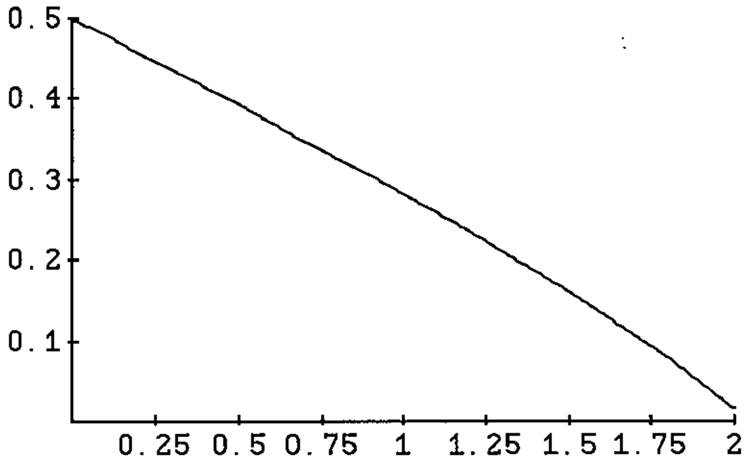

FIGURE 7.2: SOCIAL EFFICIENCY OF MIN RULE

Abscissa: $d=c_B-c_A$
Ordinate: **limiting market share of the inferior product A**
Parameter: $n=5$

For d>2.1, the limiting market share of A is 0.

The mean rule can also generate decreasing social efficiency as available information increases. For example, when d=1, the inferior product gets only 5.6% market share if n=4, compared to the maximum value of 31.5% if n=8.

c) **The popularity rule**

For n≥3, the probability that the next adoption is of type A given that the current proportion of A adoptions is x is greater than x for x>1/2 and less than x for x<1/2. This follows from the observation that the probability of an A adoption is just the probability that a Binomial[n,x] random variable exceeds n/2 (+ 1/2 or x times the probability that the random variable equals n/2, if n is even); a simple recursion shows that this probability increases with n for all x>1/2. Moreover, for n=2 and the "1/2" version of the popularity rule, the probability of an A adoption given x is just x.

Thus, according to the results cited in section 2 above, for n>2 (or n=2 and the "x" version of the rule) the limiting market share of A can be either 0 or 1. Which limit is obtained depends on the initial parameters R and S. If R=S, either product dominates the market with probability 1/2. Of course, the limiting distribution does not depend on d.

8. DISCUSSION

In this paper, we have 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 world, 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.

What we found is that mechanism at the micro-level has a determining and surprising effect at the macro-level. In particular, a population of max rule followers will in general enjoy a greater level of social welfare than will a population of Bayesian optimizers. We do not know any other result in the literature in which a specific rule-of-thumb has been shown to outperform Bayesian optimization at the aggregate level. Unfortunately, we cannot explain why the max rule performs as it does.

Our results showing that performance at the aggregate level can decrease as more information is available at the micro-level have a number of precedents in somewhat different contexts (see, for example, Bassan and Scarsini, 1995, and Ellison and Fudenburg, 1993). While it is possible to give a heuristic interpretation of why this happens in the information contagion context, we forebear from doing so, since we cannot yet explain anything like the complicated dependence of social efficiency on sample size that we observed in Example 3 of Section 6.

The final result to which we called attention in Section 1 is interesting for another reason. One of the motivations for the Warglien-Narduzzo experiment was to see if the path-dependence predicted in Arthur and Lane (1993) could be empirically verified. It was; but as we showed in Section 7, three of the rules that the agents claimed to use in making their choices do not lead to path-dependence. So where was the observable path-dependence coming from? Were the agents really crypto-Bayesian optimizers? Of course, the experimental subjects used different rules when they faced different data, and we have not analyzed what kinds of "meta-rules" determine which rule to use when, and what limiting market-shares would be associated with such meta-rules. We are currently awaiting further experimental results the analysis of which might shed light on what real agents in this kind of context actually do -- and how much social efficiency they attain by doing it.

REFERENCES

Arthur, W.B. and D.A. Lane (1993), Information contagion, *Economic Dynamics and Structural Change*, 4, 81-104.

Bassan, B. and M. Scarsini (1995), On the value of information in multi-agent decision theory, to appear in *Journal of Mathematical Economics*.

Ellison, G. and D. Fudenberg (1993), Word of mouth communication and social learning, *Harvard Institute of Economic Research Economic Theory Discussion Paper* #12.

Hill, B.M., D.A. Lane and W.D. Sudderth (1980), A strong law for some generalized urn processes, *Annals of Probability*, 8, 214-226.

Warglien, M. and A. Narduzzo (1994), Learning from the experience of others, Universita' degli Studi di Venezia, Dip. di Economia aziendale.