Real Word-of-Mouth Interaction and Organization of Behavior*

Zakaria Babutsidze[†‡] and Robin Cowan[‡♮]

[‡] MERIT - Maastricht University

[♮] BETA - Université Louis Pasteur

Abstract

We present a model of word-of-mouth interaction. Agents interact on fixed, regular social network and exchange information in order to make a choice between multiple alternatives. Communicated information includes rumors. The model also includes the inertia in economic behavior. We analyze the organization of behavior in the long run. We show that for a large range of initial conditions clustering in economic behavior in case of two-option environment emerges and persists indefinitely. Multi-option setup is also investigated. Here it turns out that long-run option popularity distribution is heavily skewed with few options dominating large portion of social space.

Key Words: *Choice* · *Interaction* · *Inertia* · *Clustering*

JEL codes: D71 · D83 · C65

*The authors are grateful to Michael Ben-Gad, Alex Coad, Steve Childress, Giorgio Fagiolo, Emin Karagozoglu, Bulat Sanditov, Jamsheed Shorish, Marco Valente and Vladimir Yankov for helpful discussions. Comments from other participants of various meetings in Ancona, Jerusalem, Kiel, London, Maastricht, New York, Paris and Tbilisi are also appreciated.

[†]Corresponding author. Address: MERIT, School of Business and Economics, Maastricht University. Keizer Karelplein 19, 6211 TC, Maastricht, The Netherlands. E-mail: *babutsidze@merit.unu.edu*.

1 Introduction

Information is the most valuable asset when making choices. It can be acquired in many different ways, in many different formats. Therefore, at any point in time any agent can possess discrete bits of diverse information about available options rather than one, unified informational landscape. Human cognitive and computational capabilities do not allow for processing the information to the extent that it can be used as input to the optimization problem. Therefore, many types of economic behavior are based on heuristics instead of solutions to optimization problems.

Significant amount of information streams through social networks by means of word-of-mouth communication. Research has demonstrated that social networks are thought to be the least corrupt, the most reliable sources of external information (Hansen, 1972). It also seems that information transfered through word-of-mouth communication has a special form. Psychological research has shown that people do not remember particular features of options (in this case consumption products), but rather maintain and communicate general impressions (Wyer and Srull, 1989; Park and Wyer, 1993). This finding becomes central in our choice of modeling approach for this paper.

There are two more reasons for our choice of modeling strategy. One is the observation that the structure of the word-of-mouth communication is characterized by temporal stability. Due to geographical and/or social constraints we tend to speak to the same small set of other people. This implies that social interaction is necessarily of local nature, that there is an underlying social network that describes how information flows through the population. This feature of word-of-mouth communication has been overlooked in research. This kind of interaction has usually been modeled by means of random matching (e.g. Ahn and Suominen, 2001; Rob and Fishman, 2005) which, we believe, is not an accurate description of social ties that are required for word-of-mouth.

The other is the fact that people can talk about products with which they have no direct experience. In this sense word-of-mouth can generate certain rumors that may induce people to choose certain options. We believe this is the major characteristic feature of word-of-mouth communication. Unfortunately it has been left unexplored in the literature. Although Banerjee and Fudenberg (2004) allude to this feature in the motivation to their work, they model a different characteristic, which is the fact that only small (and perhaps the most important) bits of information are transfered through word-of-mouth. Their setup allows for communication of the current state of affairs with regard to the options that agents have experienced, but omits the transfer of "rumors." It is important to make a distinction about the rumors that we have described, and rumors as understood by Banerjee (1993). In the latter case rumors diffuse only though practices, while, we believe, the essence of the former is the diffusion of information without explicit diffusion of practices.

Implications of behavior with choice heuristics has been examined before (e.g. Grabowski, 1970; Smallwood and Conlisk, 1979). Implications of word-of-mouth interaction has not been left without attention either. Fundamental contribution in this respect is due to Ellison and Fudenberg (1995). They analyze the setup where society learns about the merits of the alternative technologies through word-of-mouth interaction. Choices are based on rules-of-thumb. The parameter controlling the interaction structure in this model is the number of peers with whom an agent is randomly matched to learn from. The result is that in this environment socially efficient choices can be reached. However, it is also possible that the economy ends up in an inefficient allocation. This is somewhat similar to informational cascade models where learning from others' experience in sequential choice models can easily result in inefficient outcome (e.g. Bikhchandani et al., 1992; Banerjee, 1992).

Bala and Goyal (2001) take word-of-mouth learning one step further by imposing the constant interaction structure. They analyze the model with fixed social network instead of random matching. Here available options are of the same quality. Authors show that even if the society is fully connected, word-of-mouth communication can result in diversity of choices among homogenous agents. In an earlier contribution Bala and Goyal (1998) showed that diversity in equilibrium cannot be obtained when options' qualities differ. In this case society can end up in an inferior allocation, but this allocation will always be characterized by the global conformity.

One feature of studies on social interactions is that the concern is the learning with the help of others. This usually involves communication of returns one agent receives to other agents. However, returns of certain actions can be uncertain even after the action has been taken or they can change over time (think of consumption of a product). Therefore, consumers might be confronted by the learning process with the moving target. Furthermore, perception of returns can be subjective, which can further complicate the process and make the idea of learning irrelevant. Besides, considerations other than the direct payoff can be involved in the choice process. For example, the need to "fit into" some social group might induce an agent to choose a different option from the one she would have chosen considering only the direct payoffs (Cowan et al., 1997).

Therefore, in this paper we propose the framework to analyze implications of word-of-mouth communication without learning about payoffs. We rather model choices being based on subjective valuations, which are themselves affected by the behavior of society. We concentrate on the long-run outcomes of the model.

Following Bala and Goyal (2001) interaction in our model has a fixed structure. Agents have to make choices between multiple alternatives that are available at all times at constant and equal costs. They do not engage in full-fledged utility maximization. They utilize behavioral rules-of-thumb instead. Similar to Ellison and Fudenberg (1993), our modeling makes use of another empirical observation. This is the existence of inertia in behavior. However, the form taken by inertia in our model is different from the mechanism used by Ellison and Fudenberg (1993). In their case only part of population has an option to revise their decisions. Thus, inertia is introduced by imposing agents to stick to their choices for extended periods. In our case every agent repeats the choice during every period. Therefore, in order to take into account the inertia in the system we make sure that product valuations depend positively on the number of consumptions prior to the current choice date.

By allowing the possibility of rumors in case of word-of-mouth communication we allow agents to pass on the information that they have obtained from others, but had no chance to verify. This creates the mechanism that goes beyond the usual assumptions of "must-see-to-adopt" type (Bjonerstedt and Weibull, 1995; Schlag, 1998). Due to this feature the proposed framework can account for the sudden emergence of a practice in neighborhoods with no prior history of similar behavior, which is clearly not possible with "must-see-to-adopt" type assumptions.

The modeling choice that we make involves a sacrifice. In our model the option is as good as perceived by the society. Because there are no objective payoffs to options, we cannot discuss the social optimality of the outcomes that has been the main concern of the literature. However, our choice has two significant advantages over alternatives. One is the fact that it allows us to go one step further than the rest of the literature and derive implications for the organization of behavior. Previous work has obtained results on equilibrium frequency distributions over options (e.g. Ellison and Fudenberg, 1995; Bala and Goyal, 1998). In addition to replicating these results, we are able to discuss the location of agents behaving in certain ways. In particular we are able to show that in some cases the economy will exhibit the clustering in social space, which means that agents interacting with each-other will behave similarly. Unlike Bala and Goyal (2001), who also obtain similar results, we show that clustering can occur under the interaction topology where every agent has the same degree of social embeddedness.

The second advantage of our approach is that it permits for the straightforward extension of the two-option model to the multi-option environment. Modeling diffusion of rumors and inertia at the same time allows us to separate the dynamics of the valuation profile (across the population of agents) of one option from the dynamics of the valuation profile of all the other options. Therefore, extension to the choice among multiple options does not create any particular difficulty, which has been the case with the previous studies (in particular with Bala and Goyal, 2001).

The remaining of the paper is organized as follows. Section 2 presents the model. Section 3 presents main results for the two-option environment. Section 4 presents results in case of multiple options. Section 5 presents the discussion about the implications of modeling rumors. And the section 6 concludes.

2 The model

Consider an economy inhabited by a large, finite number (S) of agents, indexed by s. Each is a single decision-maker faced with the same fixed, finite set of exclusive options, indexed by n. In each period, each agent chooses one option. The decision is based on agent's subjective valuations of every available option. Assume the cost for every option is the same, so we can omit it from consideration.

We define $v^s_{n;t}$ as the valuation agent s ascribes to option n at time period t and $\mathbf{V}^s$ the vector of valuations of all options for agent s at period t. As in Ellison and Fudenberg (1993), agents are using rules-of-thumb in order to choose between the options, given their private valuation vectors. In particular, we assume there exists a function mapping option valuations into choice probabilities. As a consequence we will have $p^s_{n;t}$, the probability that agent s will choose option n at time t.

Intuitively, we assume that $\partial p^s_{n;t}/\partial v^s_{n;t} > 0$, and that $\partial p^s_{n;t}/\partial v^s_{j;t} < 0$, $\forall j \neq n$.

As we asserted in introduction valuations can change over time. In this particular case we are interested in modeling word-of-mouth interaction together with the inertia in the system. Therefore, it is useful to separate the effects of two forces on individual valuations. We assume that valuation is additive in two parts: $v^s_{n;t} = x^s_{n;t} + y^s_{n;t}$, where $x^s_{n;t}$ is determined by own choice history (incorporating inertia), and $y^s_{n;t}$ by the choice history of other members of the society.

To model word-of-mouth interaction among agents we assume that every consumer has a fixed social location and a fixed neighbourhood. A neighbourhood is the set ($\mathcal{H}^s$) of other agents with whom an agent (s) interacts directly. In this context, interaction is tantamount to information exchange. Each information exchange consists of two agents revealing to each other their private evaluations of each of the options. The information revealed is assumed to be "convincing" in the sense that the post-exchange valuations of each of the two agents partially converge. Hence, this exchange process can be expressed simply in terms of the dynamics of beliefs of a single agent, s, following her exchanges with all of her neighbours, i:

$$\Delta y^s_n = \sum_{i \in \mathcal{H}^s} \frac{\mu}{|\mathcal{H}^s|}(v^i_n - v^s_n), \tag{1}$$

where $|\mathcal{H}^s|$ is the cardinality of the set $\mathcal{H}^s$ (number of neighbours of agent s), and μ ($\in [0,1]$) is the intensity of interaction. We assume that all products are substitutes and there are no *ex ante* systematic differences among consumers, so interaction intensity is the same across all the options and agents.

For concreteness, assume that consumers are located on a one-dimensional, regular, periodic lattice such that the distance between any two agents corresponds to the social distance between them, and the distance between immediate neighbours is constant across all the population. In this case we can define the neighbourhood of an agent ($\mathcal{H}^s$) simply by specifying the number of agents (H^s) with whom this consumer interacts on the left and on the right. Then $|\mathcal{H}^s| = 2H^s$.

If we assume neighbourhood size to be equal across the population, that is $H^s = H$ $\forall s$, we can write

$$\Delta y^s_n = \frac{\mu}{2H} \sum_{h=1}^{H} \left[(v^{s+h}_n - v^s_n) + (v^{s-h}_n - v^s_n) \right], \tag{2}$$

where s can be interpreted as a "serial number" of an agent, or her address (consequently, $s+1$ and $s-1$ are her immediate neighbours to the right and left respectively).

Re-arranging, (2) can be rewritten as

$$\Delta y_n^s = \frac{\mu}{2H}\left[\sum_{h=1}^{H}(v_n^{s+h} + v_n^{s-h}) - 2Hv_n^s\right]. \tag{3}$$

Modeling inertia in behavior has usually taken a form of allowing only randomly selected part of agents to make choices at any time period (e.g. Ellison and Fudenberg, 1993). In our case, however, this option is not viable due to mathematical complication it creates. Therefore, we model inertia as an agent forming a habit for an option, as this mechanism ensures that choices are somewhat "sticky." Habits in economics have mostly been understood from a macro prospective. For example, for macroeconomists, habits in consumption mean strong positive autocorrelation in expenditures (e.g. Abel, 1990; Constantinides, 1990). However, in our case we consider forming a habit for making one particular choice, and model it as an increment in valuation of the product that has been consumed. The economic justification for this kind of behavior can range from learning particular new features about the option (think about purchasing a sophisticated consumer electronic product) to the fear of disappointment with the new option (think about the prospect of changing the school your child goes to). In any case, this kind of behavior can be observed in real life (see, for example, Chintagunta et. al., 2001; Arnade et. al., 2008) and in this particular case it gives us the opportunity to include the inertia in our system.

Mechanically, we assume that Δx_n^s is equal to zero for the options that are not chosen in a given period and is equal to some positive value for the chosen option:

$$\Delta x_n^s = \begin{cases} \zeta & \text{if } n \text{ has been chosen} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where ζ (> 0) is a constant.

To summarize the model we can make explicit the sequence of agents' actions. At the start of each period every agent makes a choice and adjusts the valuation for the chosen option accordingly. At the end of the period each agent socializes with all of her neighbours and passes to them all the information (that is, her valuations of all goods) that she possesses. Based on the information communicated to them, all agents adjust their valuations for all options.

The model described above is not solvable in its current form. Therefore, we make two modifications. First, we assume that the habit formation process can be well-approximated (at least in the region of interest) by a linear function. Second, we re-write the model as continuous in time and space.

Linearization. Above, equation (4) shows habit formation: a consumer forms habits only for the good she consumes, and the effect on her valuation takes place in discrete jumps. This describes a path dependent process. This is problematic, as analysis of the system at any point in time requires analyst's knowledge of the whole history of the system. However, making use of choice heuristics that our agents follow allows us to describe the dynamics of $\Delta x^s_{n;t}$ with a the following Markov process :

$$\Delta x^s_n = \begin{cases} \zeta & \text{with probability } p^s_{n;t} \\ 0 & \text{with probability } 1 - p^s_{n;t}. \end{cases} \tag{5}$$

Then, the expected change in valuation due to habit formation can be written as:

$$E\left(\Delta x^s_n\right) = \zeta p_n(\mathbf{V}^s_t). \tag{6}$$

The choice probability for a product n depends on valuations of all available options. However, it is reasonable to assume that the contribution of changes in valuations of options other than n are of second order significance. This is easy to see if we consider the effects of an increase in the valuation of option n. This will increase its purchase probability by Δp_n. This will also decrease the purchase probabilities of all the other products, each by Δp_j. As probabilities are normalized values it will be the case that $|\Delta p_n| = \sum\limits_{j \neq n} |\Delta p_j|$. If we have relatively large number of products in the economy, it will in general be true that $\Delta p_n \gg \Delta p_j$, $\forall j \neq n$. Thus, a change in the valuation of one option will cause the change in its choice probability. It will also cause the changes in choice probabilities of other options, but the size of each of these changes will be considerably smaller. Therefore, we impose a restriction on our probability function: it has to satisfy the following relation

$$\left|\frac{\partial p_n}{\partial v_n}\right| \gg \left|\frac{\partial p_n}{\partial v_j}\right|, \tag{7}$$

$\forall j \neq n$.

Consider the linearization of function $p_n(\mathbf{V}^s_t)$. If the requirement (7) is satisfied, as a first approximation, we can disregard the effects of lower orders of magnitude and write a linearized function as $p_n(\mathbf{V}_t^s) \approx \gamma v_{n;t}^s$. This permits us to write the expected change in $x_{n;t}^s$ as

$$\Delta x_n^s = \alpha v_n^s, \tag{8}$$

where α $(= \gamma\zeta)$ can be interpreted as the rate of habit formation.[1]

This allows us to write our model as

$$\Delta v_n^s = \alpha v_n^s + \frac{\mu}{2H}\left[\sum_{h=1}^{H}(v_n^{s+h} + v_n^{s-h}) - 2Hv_n^s\right]. \tag{9}$$

From (9) it is clear that the law of motion of valuation for every option for any agent depends on the agent's own valuation of that option, and on the valuations of the agent's neighbours of that same good. Note that in linearized system (9) it does not depend on valuations for other options. This is the characteristic of our approach that allows us to analyze the multi-option environment in section 4.

Before moving to multi-option environment, for the demonstration of main implications we assume that there are only two options in the choice set ($N = 2$), and that each agent has exactly two neighbours ($H = 1$). In this case the model reduces to a system of S pairs of equations of the form

$$\Delta v_1^s = \alpha v_1^s + \frac{\mu}{2}\left(v_1^{s+1} + v_1^{s-1} - 2v_1^s\right) \tag{10}$$

$$\Delta v_2^s = \alpha v_2^s + \frac{\mu}{2}\left(v_2^{s+1} + v_2^{s-1} - 2v_2^s\right), \tag{11}$$

where $s = 1, 2, 3, \ldots, S$.

Continuous time and space. We seek the solution to the system given by (10) - (11). In the two-option system, what drives the dynamics is the difference in the probabilities that each of the options is chosen (by each consumer). We can thus re-write the system in terms of the difference in valuations between two options. Define the valuation difference $z^s = v_1^s - v_2^s$ and rewrite the system (10)-(11) as

$$\Delta z^s = \alpha z^s + \frac{\mu}{2}\left(z^{s+1} + z^{s-1} - 2z^s\right). \tag{12}$$

[1]In what follows we drop the expectation sign, although it should be remembered that all the discussion in this section is about the expected values of the variables.

Assume the population is dense enough on the circle that we can safely use the continuous space approximation. To do this we define a new variable δ which is the distance between two neighbouring agents in social space (on a circle). This permits us to treat the agent index as a variable.

Further, due to the way we have modeled inertia in the system, we can also allow agents to make choices with infinite speed and still be sure that inertia remains. This allows us to rewrite the system in continuous time.

The continuous analog of (12) becomes

$$\frac{\partial z(s)}{\partial t} = \alpha z(s) + \frac{\mu}{2}\left(z(s+\delta) + z(s-\delta) - 2z(s)\right). \tag{13}$$

Take a second order Taylor approximation in space around s for the terms $z(s+\delta)$ and $z(s-\delta)$:

$$z(s+\delta) \approx z(s) + \delta\frac{\partial z(s)}{\partial s} + \frac{\delta^2}{2}\frac{\partial^2 z(s)}{\partial s^2}, \tag{14}$$

and

$$z(s-\delta) \approx z(s) - \delta\frac{\partial z(s)}{\partial s} + \frac{\delta^2}{2}\frac{\partial^2 z(s)}{\partial s^2}. \tag{15}$$

Substituting equations (14) and (15) into equation (13) collapses our system into one partial differential equation

$$\frac{\partial z}{\partial t} = \alpha z + \tilde{\mu}\frac{\partial^2 z}{\partial s^2}, \tag{16}$$

where $\tilde{\mu} = \mu\delta^2/2$.[2]

In the following section we investigate the long run behavior of the dynamic system specified in (16).

3 Organization of behavior

It simplifies the analysis to separate the dynamics of $z(s;t)$ into the dynamics of the average over the population $\bar{z}(t)$; and the dynamics of the deviations from this

[2] Note that making higher order approximations in (14) and (15) will leave only the even number terms in the expression (16). Odd number will always cancel out. Thus, the third order term, the one with the order of significance from the omitted terms, can be safely ignored. Taking into account the fourth or higher order terms is not customary to economics.

average $\tilde{z}(s;t) = z(s;t) - \bar{z}(t)$. Then, we can characterize the long-run behavior of the system by following three lemmas.

Lemma 1. *At any point in time, $\bar{z}(t)$ can be described by*

$$\bar{z}(t) = e^{\alpha t}\bar{z}(0).$$

The proof of Lemma 1 can be found in the appendix. As $\alpha \geq 0$, Lemma 1 implies that the difference in average option valuations increases or decreases exponentially with time. $z(0)$ determines the direction of $z(t)$ dynamics.

Lemma 2. *With time, $\tilde{z}(s;t)$ converges to*

$$\tilde{z}(s;t) = e^{\sigma t}\cos\left(k\frac{2\pi}{l}s\right)\tilde{z}(0;0),$$

where l is the length of the circle on which consumers are placed, while σ is the amplitude growth rate and $k(\in \mathbb{Z}_+)$ is the frequency of the sinusoid $\tilde{z}$.[3]

The comprehensive proof of this proposition can be found in Turing (1952); here we give the basic intuition. The general solution to differential equations of type (16) can be represented as the (possibly infinite) sum of exponential functions of the form Ae^{bt}, where A and b are (possibly complex) coefficients. The real part of each summand in the solution can be represented as the dynamic sinusoid (in our case around the lattice on which consumers are located). These sinusoids are the components of Fourier decomposition of the actual distribution of $\tilde{z}$ on social space. The real part of each b will be the growth rate of the amplitude of the corresponding sinusoid. As a result, as $t \to \infty$ one summand will dominate all the others. This will be the term with the largest real part of b. Consequently the dynamics of the solution will converge to one sinusoid.

Lemma 3. *The amplitude growth rate of the dominant sinusoid of system (16) is*

$$\sigma = \alpha - \tilde{\mu}k^2\left(\frac{2\pi}{l}\right)^2.$$

[3]Note that as consumers are located on a periodic lattice, the identity of agent zero is arbitrary, and thus can be placed anywhere on the circle. To write down proposition 2 we have set label 0 such that $s_0 = \underset{x\in\left[0,\frac{l}{k}\right]}{\arg\max}\cos\left(k\frac{2\pi}{l}x\right)$, which effectively means that we label agents such that the sinusoid identified in proposition 2 reaches its maximum at the agent number zero.

Proof of Lemma 3 can be found in the appendix.

Lemmas 1 through 3 fully characterizes the solution to the system (16). In what follows we report on the implications of this solution for organization of choice behavior.

For making interpretations of the results transparent, it is useful to go back to the discrete representation of the model. Thus, we move back to treat s as the serial number of an agent.[4] This makes $\tilde{\mu} = \mu/2$ and $l = S$. In this case we can write the complete solution to our system, by combining Lemmas 1 through 3, as

$$z_t^s = e^{\alpha t}\bar{z}_0 + e^{\sigma t}\cos\left(k\frac{2\pi}{S}s\right)\tilde{z}_0^0. \tag{17}$$

where

$$\sigma = \alpha - 2\mu\frac{\pi^2}{S^2}k^2 \tag{18}$$

Equation (17) determines the value of the difference in valuations (z) for every agent for every $t \gg 0$. The distribution of z along the circle has the form of a wave in space around the average, which points to the fact that in some neighbourhoods $\tilde{z} > \bar{z}$, while in some other neighbourhoods the opposite holds true. Thus, the general result is that the clustering in behavior is an emergent property of our system.

Our concern is whether any observed clustering is persistent over time. Consider the case when $\exists t \geq 0$ such that $\bar{z}_t \neq 0$. That is, at some point in time one of the options is perceived as superior on average.

Proposition 1. *If $\exists t$ such that $\bar{z}_t \neq 0$, then as $t \to \infty$, $v_i^s > v_j^s$ $\forall s$ and thus in the long run everybody will choose one of the options with the higher probability than the other.*

Proof. Consider the situation when $\bar{z}_t > 0$. Define $z^{min} \equiv \min_s (z^s)$ as the valuation difference of an agent with the lowest z.

Case 1: $z^{min} > 0$. This implies that $\forall s$ $z^s > 0$, thus there is one cluster of size S. This is a stable pattern as both forces (interaction and habit formation) work to reinforce it.

Case 2: $z^{min} < 0$. In this case some of the consumers prefer the relatively "inferior" product.

[4] This effectively means that we fix $\delta = 1$. This move does not undermine the results of Lemmas 1 through 3. Moving back to consumer addresses is convenient for relating parameters in the solution to the parameters of the model.

Case 2a: $\sigma < 0$. Lemma 2 tells us that if $\sigma < 0$, with time, the amplitude of the wave goes to zero, which implies that $\forall s\ z^s = \bar{z}$. This, together with proposition 1, results in $z^s > 0\ \forall s$ as $t \to \infty$.
Case 2b: $\sigma > 0$. From lemma 2 we know that the amplitude of the wave around the average increases at rate σ. At the same time, propsition 1 suggests that the average over agents of the valuation-difference rises at the rate α. Therefore z^{min} is rising at the rate $\alpha - \sigma$. Equation (18) establishes that this rate is positive.[5] $\alpha - \sigma > 0$ ensures that as $t \to \infty$, $z^{min} > 0$. $z^{min} > 0$ implies that $\forall s\ z^s > 0$. Thus case 2b with certainty collapses into case 1 at some point in time.

These intuitions hold for the situation when $\bar{z}_t < 0$. □

Notice, that due to the fact that agents use probabilistic choice heuristics there are two relevant spaces for reporting results: the valuation space and the choice space. Of course the choice space is the derivative of the valuation space. The correspondence from one to another is best seen on the example of the proposition 1. What proposition 1 implies is that there exists the solution of the model where the whole economy comprises one cluster in the valuation space. Because the correspondence between the valuation and choice spaces is probabilistic, in general, this will only imply the fact that everybody will choose one of the options with higher probability. We call this pattern in choice space a probabilistic clustering. We also define somewhat stronger notion of absolute clustering, which means that neighbors will consistently choose the same product in the long run. As in our case choices are probabilistic, this will only be the case when the probability of choice of one of the options goes to one in the long run.

Proposition 1 implies that there is a probabilistic clustering in the sustem. In this particular case, however, the system will be characterized by the absolute clustering.

Proposition 2. *If $\exists t$ such that $\bar{z}_t \neq 0$, as $t \to \infty$, $v_i^s - v_j^s \to \infty\ \forall s$, therefore clustering in the economy will be absolute and in the long run global conformism will obtain.*

Proof. Proof of corollary 1 directly implies not only that $v_i^s > v_j^s\ \forall s$ in equilibrium, but also that $v_i^s - v_j^s \to \infty$, which on its own implies that as long as the choice probability function is a positive monotonic mapping of valuations to

[5] Unless $\mu = 0$, which is not a very interesting case as it implies no word-of-mouth communication. In this case the existing choice pattern is reinforced indefinitely.

choice probabilities, the probability of any agent purchasing product i converges to 1. $\square$

Proposition 2 implies that probabilistic clustering converges to absolute clustering in behavior asymptotically. Thus, $\bar{z}_t \neq 0$ is a relatively trivial case, and implies that ultimately only one option survives in the population, no matter the dynamics of the deviations from the average. Similar results on global conformism have been obtained by researches on global (Ellison and Fudenberg, 1995) and local (Bala and Goyal, 2001) interaction, with repetitive (Ellison and Fudenberg, 1993) and sequential (Banerjee, 1992) choices.

Far more interesting is the case in which $\forall t \; \bar{z}_t = 0$, which permits both options to co-exist indefinitely. To analyze this case note that intuitively the stability of the cluster should depend on its size. For example, if one individual constitutes a cluster she is susceptible to influence from both her neighbours, both proponents of the choice contrary to hers. This cluster is less likely to be stable than a larger cluster where most of the members of the cluster (the ones away from its boundaries) receive information that reinforces their choices. Thus, there should be some minimum cluster size for which clustering will be persistent. When $\forall t$ $\bar{z}_t = 0$ we know that behaviour of the system is governed by the pattern sinusoid, which implies that all the clusters are of an equal size in the long run.

Proposition 3. *In system (16), if $\forall t \; \bar{z}_t = 0$, clustering in behavior is stable if and only if the pattern wave of the system results in the clusters of size $c \geq \underline{c} = \frac{\pi}{\sqrt{2}}\sqrt{\frac{\mu}{\alpha}}$.*

Proof. From equation (17) it can be readily seen that, when $\bar{z}_t = 0 \; \forall t$, temporal stability of clustering depends on the sign of σ. If $\sigma < 0$, as $t \to \infty$, $z^s \to 0 \; \forall s$, which implies that $v_1^s \to v_2^s \; \forall s$. This means that valuations of products converge, so in the case of probabilistic purchases every agent decides on her purchase by tossing a (fair) coin. This, clearly, will result in no clustering pattern.

However, if $\sigma > 0$ the amplitude of the pattern wave increases exponentially with time, thus clustering becomes more and more pronounced. If $\sigma = 0$, the amplitude of the wave does not change with time, and clustering is still stable.

Given the parameters of the model, the sign of σ depends on the frequency of the wave in the initial condition. We can pin down the critical frequency of the pattern wave (k), for which clustering will be stable, by simply solving $\alpha - \mu k^2 \frac{2\pi^2}{S^2} = 0$, for k. This results in $\bar{k} = \frac{S}{\pi}\sqrt{\frac{\alpha}{2\mu}}$. And $k \leq \bar{k}$ ensures that $\sigma \geq 0$. The inverse of the frequency is the wave length, and the size of the cluster is half of the wave length. Since the size of the economy is S, the size of the cluster(s) is $S/(2k)$. Thus, given $\bar{k}$, we can find the size of the smallest cluster that will persist over time: $\underline{c} = \frac{\pi}{\sqrt{2}}\sqrt{\frac{\mu}{\alpha}}$. Any pattern wave exhibiting clusters larger than $\underline{c}$, would ensure $\sigma \geq 0$, and thus will result in stable clustering. $\square$

The important property of the minimum stable cluster size is that it does not depend on the size of the economy. However, as σ depends on S, a larger economy (*ceteris paribus*) increases the likelihood that the pattern wave of the system will support clusters of any given size c, thus it also increases the likelihood of clustering. We also point out that the minimum stable cluster size depends on the ratio of two parameters, habit formation and information transmission: μ/α.

There are three distinct behavioral clustering patterns identified in the proof of proposition 3. These are implied by followinf three scenarios: $\sigma = 0$ (this is the same as $c = \underline{c}$), $\sigma > 0$ ($c > \underline{c}$) and $\sigma < 0$ ($c < \underline{c}$).

$\sigma = 0$: In this situation the valuation distribution converges to a static sinusoid. Consequently, the long run valuations are constant. This implies that $v_i^s - v_j^s$ is bounded $\forall s$. Therefore, the in case of $\sigma = 0$ the is only a probabilistic clustering in behavior.

$\sigma > 0$: In this case valuation distribution is governed by the sinusoid with ever increasing amplitude. Therefore, the behavior in social space is organized as alternating neighborhoods of agents with $v_i^s - v_j^s \to \infty$ and $v_i^s - v_j^s \to -\infty$. In this case polarization among clusters reaches extreme values and the organization converges to absolute clustering in behavior.

$\sigma < 0$: This is the case when there is no clustering in behavior, no particular pattern of organization. Here valuations for the options converge to each other for every agent. Therefore, every decision maker's probability of choosing one of them converges to 0.5. In this case information coming through word-of-mouth is so strong[6] about each of the products, that it confuses the agent, who ultimately decides to randomly choose between the products.

This result is somewhat similar to the result of "confounded learning" by Smith and Sorensen (2000). In a sequential choice model with interactions they find a

[6]From equation 18 one can easily see that negative σ is a result of higher rate of communication μ.

scenario where learning process consistently maintains the balance between the options in a sense that information gathered from other decision-makers carries no value for the decision process of an agent.

The analysis so far has assumed that each agent has two neighbours ($H = 1$ on either side). It is interesting how results of the model change if we consider larger neighborhoods.

Proposition 4. *In the case of arbitrary neighbourhood size $2H$, where agents interact with H closest neighbors from each side, the minimum sustainable cluster size is*

$$\underline{c}_H = \frac{\pi}{2\sqrt{3}}\sqrt{2H^2 + 3H + 1}\sqrt{\frac{\mu}{\alpha}}.$$

The proof of this proposition can be found in appendix.

Proposition 4 implies that as neighbourhoods grow in size so does the minimum sustainable cluster. The intuition is that a larger neighbourhood facilitates the information diffusion process: each agent receives information from relatively distant agents. This works to homogenize the information structure across the population, and so works against small clusters.

There are few relevant findings in literature that we can draw parallels with. For example, Ellison and Fudenberg (1995) find that less communication increases the likelihood of conformism. In our case we can decompose the "amount" of communication into intensity of communication (controlled by μ) and the scope of communication (controlled by H). In our model the outcome of global conformism does not depend on any model parameters (proposition 1). However, any type of clustering is conformism and if clustering is local, so is conformism. In our model, once global conformism is ruled out, the likelihood of local conformism is inversely related with both μ and H (see equation (25) in the proof of proposition 4). In Ellison and Fudenberg (1995) slow information exchange ensures multiplicity of trials before the equilibrium is reached and thus increases the likelihood of the society learning about the true best option. In our model slow information exchange gives the chance to groups of agents to "develop the taste" for one particular option.

Related finding has been reported by Bala and Goyal (2001). They concentrate directly on local conformism as the long run outcome. They characterize the social network by the degree of integration of decision-makers and find that lower degrees of integration increase the neighborhood of clustering. In our model H

can also be viewed as the degree of integration: higher H means that every agent interacts with larger number of other agents. This directly implies higher level of integration. Then, our results are in line with the findings of Bala and Goyal (2001): lower level of integration increases the likelihood of amplitude growth rate of the dominant sinusoid being positive. Positive σ is a sufficient requirement for local conformism.

On more general level, the existing literature has examined the effect of the interaction scope. In general the contrast is made between local and global interactions. Local interactions imply the limited (and usually fixed) subset of other agents that any given agent interacts with, while global interactions assume that information stream from every agent can directly reach any other agent in the economy. Contrasting these two interaction schemes researchers find that global interactions usually result in more ordered systems, while local interaction usually produces richer and more complex dynamics (e.g. Glaeser and Scheinkman; 2000; Gonzalez-Avella et al., 2006). This issue can be addressed in our model by looking at its behaviour as neighbourhoods become very large ($H \to S/2$). According to proposition 4, increasing the neighbourhood size (H) puts an upward pressure on minimum stable cluster size $\underline{c}$ and for larger region of parameter space pushes it above the threshold ($\underline{c} > S/2$) beyond which clustering is unstable in the long run (in case when the differences between average valuations are zero).[7] Thus, in line with previous research, our model demonstrates that local interactions result in richer and more complex dynamics than do global interactions.

Based on proposition 4, we can analyze how minimum sustainable cluster size changes with enlargement of the interaction neighbourhood. It is obvious from proposition 4 that $\underline{c}_{H+1} - \underline{c}_H$ is increasing with H. Moreover, it turns out that

$$\lim_{H \to \infty} \left(\underline{c}_{H+1} - \underline{c}_H \right) = \frac{\pi}{\sqrt{6}} \sqrt{\frac{\mu}{\alpha}}. \tag{19}$$

Equation 19 implies that for any value of μ/α, minimum sustainable cluster size increases linearly with the size of the neighbourhood, as long as H is sufficiently large.

[7]For example, in the small economy that we have simulated ($S = 100$), $H = 49$ implies that the speed of habituation, α, must be roughly 80 times higher than the influence of neighbours, μ, in order the system to be stable for the largest possible cluster ($\underline{c} = S/2$)

4 Multi-option environment

As asserted in the introduction to the paper, one of the advantages of present approach is that it is straightforward to extend the analysis to multi-option environment. In fact the core of the model has been written down in this environment and two-option setup has been chosen only for the demonstration of major findings in previous sections.

Proposition 3 describes this relation between the parameters of the model and the average cluster size in the long-run of two-option case. In this section we analyze the same relationship for the multi-option environment.

Linearization of inertia in the system allows us to separate the valuation distribution dynamics for different options from each other in a way that they are no more interdependent. We exploit this property in order to analyze the implications of word-of-mouth interaction for organization of behavior in multi-option environment.

Consider the setup where decision-makers have to choose between N options. Assume again that consumers interact with only two of their neighbors ($H = 1$). Then, the dynamics of our model is represented by N equations of form (9). We can randomly choose one of the options as a numeraire (say product N) and subtract the value of its valuation from every other product for each consumer $z_n = v_n - v_N$, $\forall i \neq N$. After rewriting the system in continuous time and space and applying Taylor approximation to appropriate terms, N-option system will be described by $N - 1$ equations of type

$$\frac{\partial z_n}{\partial t} = \alpha z_n + \tilde{\mu} \frac{\partial^2 z_n}{\partial s^2}. \tag{20}$$

Consequence of separability property is that the dynamics of z_n does not depend on the dynamics of z_i, $i \neq n$.

Every equation in the system (20) has the same form as equation (16). Therefore, similar to the two-option environment, in this case we can also have two different outcomes. One where there is global conformism, the other where several choices appear with the positive weight in the frequency distribution of the long-run behavioral practices. Which of these scenarios will prevail depends on initial conditions.

As lemma 1 applies to all $N - 1$ equations for identifying the pattern of choice organization we have to compare the growth patterns of average valuation differences. As α has the same value across all $N-1$ equations, the growth rate of average valuation differences across every option is the same (according to lemma 1, this rate is equal to α). Then, what becomes important is the initial value of the average valuation for each of the option. It can be shown that the difference between two variables that grow with equal exponential rate goes to positive or negative infinity depending on the sign of the initial value difference. Therefore, we can formulate the following remark.

Remark 1. *In multi-option environment an option with the highest initial average valuation will be the only choice for every agent in the long run. Therefore, there will be absolute clustering and global conformism.*

If two or more options have the exact same highest initial average valuation these will be the only surviving options in the long run. Therefore, for the sake of analysis of the long-run behavior, we can safely drop all inferior practices and concentrate on the ones surviving in the long run. In this case the system can be reformulated, reindexed as the system with several options with equal average initial valuations. In what follows we analyze the behavior of this kind of system. For the sake of reducing unnecessary complication in notations we assume there are N products having equal initial average valuations.

As it can be readily seen from equation (20) each of $N-1$ equations has the same form, and the same parameter values as the unique equation (16) in two-option case. Therefore, the following remark is true.

Remark 2. *In the case of a multi-option environment, minimum cluster size implied by the dominant sinusoid of each $N-1$ valuation difference distribution is unchanged and is equal to $\underline{c}_N = \underline{c} = \frac{\pi}{\sqrt{2}}\sqrt{\frac{\mu}{\alpha}}$.*

Although the solution to the system is very similar to two-option case, its implications for the organization of behavior is considerably harder to analyze. The reason for this is multiplicity of dominant sinusoids that are present in the system. However, one important finding that we can directly point out is that there is clustering in valuations of every product separately. Therefore, the multi-product system should also result in clustering in behavior (probabilistic or absolute). The only exception to this will be the case when amplitude growth rates of all $N-1$ dominant sinusoids are negative. In this case each product will have equal chance of being chosen by any consumer in the long run. Furthermore, number of products increases the chance of clustering as each sinusoid becomes less important.

As long as there is one sinusoid with $\sigma \geq 0$, clustering will be present in the economy. Increase in number of products decreases the likelihood of a coincidence where all σs are negative.

In order to predict clustering patterns we have to go deeper in comparison of amplitude growth rates of dominant sinusoids. Recall that according to equation (18) $\sigma = \alpha - 2\mu\frac{\pi^2}{S^2}k^2$, where $k \in \mathbb{Z}_+$ is the frequency of the sinusoid. As lower k implies higher σ, as long as initial conditions permit, the fastest growing sinusoid will be the one corresponding to $k = 1$. The role of initial conditions requires additional clarification. Recall the outline of the proof of lemma 2. If we have S decision-makers, with Fourier transforms initial distribution of choice valuations over social space can be represented as the sum of waves with $k = 1, 2, \ldots, S/2$, each with corresponding initial amplitude and its growth rate. As $\partial\sigma/\partial k < 0$ we know that out of all the Fourier components the highest chance for becoming the dominant wave has the longest wave ($k = 1$). The only case when $k = 1$ will not emerge as dominant sinusoid is if its initial amplitude is equal to zero. In this case the amplitude will not change over time. The next most probable nominee for the domination will be the wave corresponding to $k = 2$ and so on. This is true for the valuation distribution of every option.

In multi-option case what becomes important is not only the dominant sinusoid for each valuation distribution, but also the competition among the dominant waves across all the options. We know that most of the dominant sinusoids have the same amplitude growth rate $\sigma = \alpha - 2\mu\frac{\pi^2}{S^2}$. The rest of the sinusoids have lower amplitude growth rates. Thus, what becomes important for identifying the winner, the champion wave, is the initial amplitude. Because the difference between the amplitudes of two sinusoids with the same (positive) growth rates and different initial values goes to infinity in the long-run, we can formulate following remark.

Remark 3. *Consider the economy with equal initial average valuations for all N options. If in this economy there is at least one option characterized by the dominant sinusoid with the positive growth rate, in the long run there will be an option that will be consistently chosen by (exactly) half of the population. The number of clusters across which this half of the population will be distributed depends on the frequency of the champion wave.*

To see why this remark is true consider space where social location is measured on abscissa (each point on abscissa corresponds to a unique decision-maker) and valuation difference (to the numeraire option) is measured on ordinate. Starting from any initial condition, after the lapse of some time, there will be $N-1$ sinusoids across abscissa all around the same horizontal - valuation difference is equal to zero - line. For demonstration purposes assume the numeraire option is not the one we claim will be the champion. It is a feature of the sinusoid that half of its mass lies in positive values, while the other half lies in negative portion of the ordinate. As we have argued earlier, there will be an option such that the difference between the amplitude of this wave and any other $N-2$ waves will go to infinity in the long run. This will ensure that positive half of this sinusoid dominates half of the social space and leaves the remaining half for other options to share.

Number of clusters across which all these $S/2$ agents will be distributed depends on the wave length of the champion wave. We know that waves with higher σs will always dominate waves with lower amplitude growth rates.

Remark 4. *As the option set expands, the probability of one of the clusters in the economy covering exactly half of the social space in the long run converges to one.*

This is true because large number of options increases the likelihood of at least one option having the dominant sinusoid with wave length $k=1$, in which case this wave is guaranteed to be the champion.

Thus, we have established that half of the social space will be organized in few clusters (most probably in just one). In order to understand how the other half will be organized note again that we are dealing with sinusoids. The remaining half of the social space has to be split by the options other than the champion. In determining how this space is distributed not only the initial amplitude and amplitude growth rate, but also the location of the sinusoid becomes important.

Despite the fact that the average valuations for all the products are equal, there will be differences in their popularity.

Proposition 5. *Consider the environment with a large enough option set, where each of the options is equally valued by the society. In this society the expected*

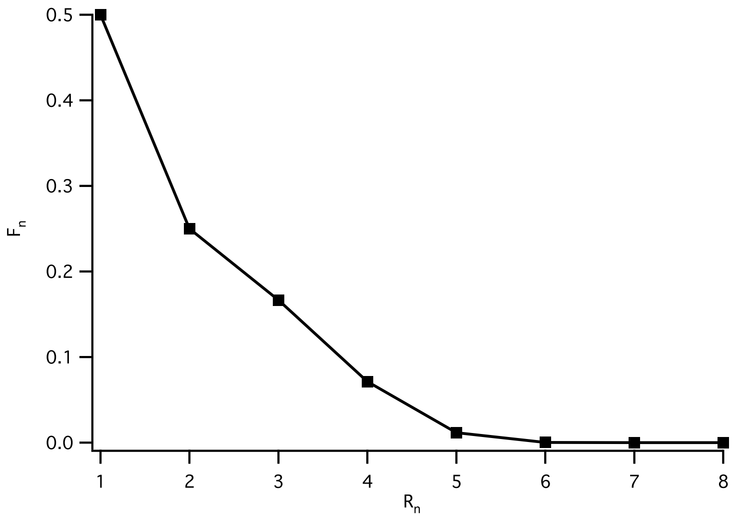

Figure 1: Expected long-run option popularity distribution.

long-run distribution of option popularity is described by

$$F_n = \begin{cases} \frac{1}{2} & \text{if } R_n = 1 \\ \frac{1-\sum_{i=1}^{R_n-1} F_i}{1+2\left(1-\sum_{i=1}^{R_n-1} F_i\right)} & \text{otherwise.} \end{cases}$$

where F_n represents the long-run share of an option n and $R_n(\in \mathbb{Z}_+)$ is option's rank in popularity ranking.

Proof of proposition 5 can be found in appendix. The plot of function F_n is presented on figure 1. The long run distribution of practice popularity does not depend on parameters of the model. It is the universal popularity distribution in the economy we are analyzing. Note that as in the economy described in proposition 5 choice popularity distribution is the same as the cluster size distribution. This is due to the fact that large number of options assures the frequency of dominant sinusoid of each option appearing in long run distribution with the positive weight is equal to one. The expected long-run option popularity distribution is extremely skewed. Few options dominate large portion of social space. Precisely, the three most popular practices cover over 11/12 of social space.

Based on the proof of proposition 5 we can formulate one more remark.

Remark 5. *It may very well be that some of the options are never chosen in the long-run despite the fact that all the options are equally valued by the society.*

This is true even if the economy consists of infinitely many practices in equilibrium and infinite number of agents. However, if the number of decision-makers is finite, due to integer problem (the cluster cannot be of size less then one agent), there will always be the finite number of clusters in the economy. Number of clusters will increase with the size of the economy. In fact every additional option in choice set requires considerably larger economy. For example, if sustaining six practices in the long run needs $1/F_6 = 3614$ decision makers, sustaining seven practices requires $1/F_7 > 6.5 \times 10^6$ agents, sustaining eight – $1/F_8 > 2.1 \times 10^{13}$ and so on. Therefore, no mater how large is the option set, in the economy with less than 3614 agents, there is room for only five of them in the long-run choice set.

5 Effect of Rumors

The main reason why we termed the phenomena that we model "real" word-of-mouth communication is that it incorporates rumors. That is our agents are transmitting the information about the option that they have received from other agents even if they had no chance of verifying it by trying out the option. This is in fact what happens in real life. The word can easily spread about "badness" or "goodness" of the option even if agents forming the informational path have not actually experienced the option in order to have quality information.

In this section we demonstrate the main effect that rumors have on the organization of behavior in our model.

Consider the conventional way interaction has been modeled up to now (e.g. Ellison and Fudenberg, 1995). In these mechanisms agents observe either actions and payoffs of other agents (e.g. Banerjee and Fudenberg, 2004), or some kind of frequency distribution of actions (without payoffs) (e.g. Banerjee, 1993). In both of cases interaction is modeled based on revealed preference. Meaning that information transmitted from agent to agent is about options that decision-makers have actually tried out. This usually implies (implicitly or explicitly) the assumption of "must see to adopt." Therefore, in case of local interaction, the diffusion of actual practices in social space has to be spatially gradual. It is not possible that agents receive information and decide to choose the option unless their neighbors do it. Revealed preferences give us information only about the actual practice and completely neglect the story that is going on in the background. By this we refer to the fact that agents do have preferences over, and information about the options they do not in fact choose.

In reality this information can still propagate beyond actual practices. In fact word-of-mouth interaction, socialization, makes if quite probable that agents reveal this information to each other. Without acknowledging the importance of those "unexpressed" preferences it is difficult to understand a sudden change in practices which is not simply imitating neighbours. Consider the following simple example. Agent $s-1$ ranks option 1 first and option 3 last; agent $s+1$ ranks option 3 first and option 1 last. Both agents, though, rank option 2 second. It is clear that agent s, based on the information communicated to her, could easily rank option 2 before either 1 or 3. If the high rankings of option 2 by $s-1$ and $s+1$ have emerged (due to information received by their neighbours) at roughly the same time, agent s can then switch to option 2, regardless of what he was doing in the past. Maintaining the practice for longer period and passing negative information about option 1 to agent $s-1$ and about option 3 to agent $s+1$, it is also likely that agent s will induce both agents to abandon their choices and switch to option 2.

To demonstrate that this kind of behavior is possible (and in fact quite probable in the early stages of industry development) we perform a small numerical exercise. We analyze the development of the system in a typical run and report on the development of clustering patterns in most preferred options. The setup is as follows. We set the number of goods to $N = 10$; and the population size to $S = 100$. The population is located on a one-dimensional periodic lattice, so the neighbours of agent 1 are agents 2 and 100. We choose the values of model parameters to be $\alpha = 0.001$ and $\mu = 0.01$. Finally, each agent has one neighbour on either side, $H = 1$. To read the figure below, agents are arrayed along the abscissa, remembering that the axis is a circle, so the right-most and left-most agent are neighbours. Time is read on the ordinate, from the initial period, $t = 0$ to the final period, $t = 2000$. Each option is assigned a different shade of gray. The ordering of options, and therefore the shades of gray, is arbitrary. At each point in time the most valued option for any agent is showed by the corresponding color.

Figure 2 presents the development of the system during the first 2000 periods in a typical run of the model. Each shade of grey is randomly assigned to one product. Seeing this shade at certain time period for certain agent means that

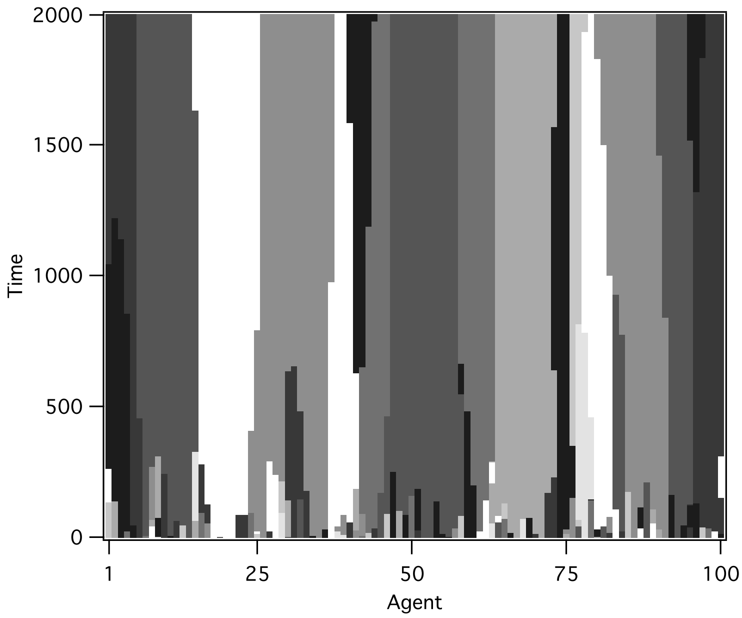

Figure 2: Cluster emergence in most preferred options.

at that moment that is the most valued option for that agent. Assume that probabilistic choice function is highly nonlinear in a way that it puts large weight on the highest valued alternative. Then valuation patterns reported here can well approximate behavioral patterns. Consider agent 40 with its neighborhood. After some initial experimentation the agent sets to value one of the options highest for extended period. She happens to be at the edge of her cluster. Therefore, we know that she is receiving strong positive signals about the option that is preferred by the cluster immediately to her right. Around time period 650 agent 40 changes her most valued option. But she does not change for the option appreciated by her neighbors from the right. Rather for the option completely novel to the neighborhood. With time this new practice becomes popular in the neighborhood. Similar example appears at later stage of development in the run, when agent 96 decides to experiment at $t \approx 1400$.

As we have demonstrated in our model behavioral clusters can emerge in social space. This is due to the rumors that can aggregate to a signal powerful enough to induce agents with narrow margins at the top of their rankings to switch to the unexplored alternative. Thus, our model is consistent not only with shrinking and disapearance of smaller clusters, but also with the emergence and growth of new ones.

6 Conclusion

This paper has studied a model of word-of-mouth communication. We extended the previous work to include the diffusion of rumors; that is the information that agent has received from her social network but has not verified. This allowed us to closely analyze the organization of economic behavior in multi-option environment. The model is rich enough to accommodate major findings by the previous literature, such as, conformism, diversity and confounded learning. It is also flexible enough to go further than existing literature in deriving the distribution of economic activity over social space in two-, as well as in multi-option environment.

It turns out that for two-option setup the main value deciding on to which type of organizational structure is the society going to converge is the ratio of communication and habit formation rates. However, initial conditions are also important. Higher values of communication intensity make small clusters unstable. Therefore, they push the economy either to no regular behavioral pattern or to larger clusters. In which of these two states is the economy to end up depends on initial conditions.

In case of multi-option environment we have shown that it the population of decision-makers is finite, then the options surviving in the long-run will also be finite. Even if the number of agents is infinite, not all options survive in the long run despite the fact that all of them are equally valued by the society. However, the number of long-run practices can be infinite. We were also able to derive distributions of cluster size and practice frequency. It turns out that these distributions are extremely skewed, with few options dominating large portion of social space.

Appendix

Proof of lemma 1

Proof. In the continuous case the average over space can be defined as $\bar{z} = (1/S)\int_0^S z ds$. This implies that

$$\frac{\partial \bar{z}}{\partial t} = \frac{1}{S}\int_0^S \frac{\partial z}{\partial t} ds.$$

Then, using equation (16) we can write

$$\frac{\partial \bar{z}}{\partial t} = \alpha \frac{1}{S}\int_0^S z ds + \tilde{\mu}\frac{1}{S}\int_0^S \frac{\partial^2 z}{\partial s^2} ds. \tag{21}$$

As space in our system is a periodic lattice the second summand in equation (21) is zero.[8] Then, using the definition of average again we can write equation (21) as

$$\frac{\partial \bar{z}}{\partial t} = \alpha \bar{z}. \tag{22}$$

This is an ordinary differential equation with the solution described in the lemma. □

Proof of lemma 3

Proof. From lemmas 1 and 2, we know that

$$z(s;t) = e^{\alpha t}\bar{z}(0) + e^{\sigma t}\cos\left(k\frac{2\pi}{l}s\right)\tilde{z}(0;0).$$

Substituting this into equation (16) and noticing that

$$\partial^2 \cos(\beta x)/\partial x^2 = -\beta^2 \cos(\beta x),$$

allows us to solve for σ. □

Proof of proposition 4

Proof. Consider the case of arbitrary neighbourhood size of $2H$. In this case after assuming that the distance between two neighbouring consumers is δ and

[8]To see more easily why the second summand is zero, one can discuss the discrete case and thus use equation (12) instead of equation (16). In the discrete case the second summand is $\sum_s \left((z^{s+1} - z^s) - (z^s - z^{s-1})\right)$. As consumers are indexed by s around a circle, it is obvious that this sum is zero.

considering the two-good case, continuous version of equation (9) can be rewritten as

$$\frac{\partial z(s)}{\partial t} = \alpha z(s) + \frac{\mu}{2H}\left[\int_{-H}^{H} z(s+\delta h)dh - 2Hz(s)\right]. \tag{23}$$

Using second order taylor approximation we can rewrite the part of (23) under the integral as

$$\int_{-H}^{H} z(s)dh + \int_{-H}^{H} \delta h\frac{\partial z(s)}{\partial s}dh + \int_{-H}^{H} \frac{\delta^2 h^2}{2}\frac{\partial^2 z(s)}{\partial s^2}dh.$$

Which, after integration of first two summands, is equal to

$$2Hz(s) + 0 + \frac{\delta^2}{2}\frac{\partial^2 z(s)}{\partial s^2}\int_{-H}^{H} h^2 dh.$$

To obtain more accurate values for smaller neighbourhood size, we go back to discrete space and replace the integral in expression above with the sum of squares of integer values.

Substituting this result back to (23) yields

$$\frac{\partial z(s)}{\partial t} = \alpha z(s) + \frac{\mu\delta^2}{4H}\sum_{h=-H}^{H} h^2\frac{\partial^2 z(s)}{\partial s^2}.$$

Thus, it follows that the only modification that this generalization brings to the system can be captured by the definition of $\tilde{\mu}$ in the text being changed to

$$\tilde{\mu} = \frac{\mu\delta^2}{4H}\sum_{h=-H}^{H} h^2. \tag{24}$$

Going back to consumer addresses ($\delta = 1$), using new definition of $\tilde{\mu}$, and the identity $\sum_{n=1}^{x} n^2 = \frac{x^3}{3} + \frac{x^2}{2} + \frac{x}{6}$ we can rewrite equation (18) as

$$\sigma_H = \alpha - 2\mu\left(k\frac{\pi}{l}\right)^2\left(\frac{H^2}{3} + \frac{H}{2} + \frac{1}{6}\right), \tag{25}$$

which results in

$$\bar{k}_H = \frac{S}{\pi}\sqrt{\alpha/\left(2\mu\left(\frac{H^2}{3} + \frac{H}{2} + \frac{1}{6}\right)\right)}, \tag{26}$$

and further in

$$\underline{c}_H = \frac{\pi}{2\sqrt{3}}\sqrt{2H^2 + 3H + 1}\sqrt{\frac{\mu}{\alpha}}. \tag{27}$$

□

Proof of proposition 5

Proof. In order to derive the distribution of popularity it is useful to split the popularity rankings in three parts: $R_n = 1$, $R_n = 2$ and $R_n \geq 3$. We consider each of these cases separately.

$R_n = 1$**:** The fact that $F_1 = 1/2$ is demonstrated by remark 3.

$R_n = 2$**:** Consider the effect of large number of options. We know that highest σ guarantees the championship of the wave. However, as each equation in system (20) has the same parameters, we know that there will be many waves with the same values of σ. Consider the grouping the waves in subsets, where waves in each subset have the same value of σ. Then we can rank these subsets starting from the highest to lowest. We also know that for winning the championship in case of equal σs what matters is the initial amplitude. Then in each subset we can rank waves in decreasing order of their initial amplitude values. Now we have a unique ranking of all the waves. We call this a preliminary ranking as some of the waves might get dropped from the top places due to the subsequent refinement. We will demonstrate that not every option will appear in the long run frequency distribution. As higher waves in ranking have higher chances for ending up in the frequency distribution we assume that the set of options is so large that all the ultimate practices will be selected from the highest ranked subgroup. Therefore, we simply disregard lower ranked subgroups. Thus, large number of options ensures that every option present in the long run frequency distribution with a non-zero weight has the wave length of $k = 1$.

We know that the most popular product has half of the market size. As large number of products ensures that the champion wave has the wave length of $k = 1$, and thus, according to remark 4, the most popular product has one cluster (of size $S/2$) in the social space. As our social space is circular we can reindex the agents without loss of generality. Assume the champion sinusoid starts at agent $s = 0$. This would mean, that the cluster of the champion practice comprises the social space between $s = 0$ and $s = S/2$. Now, what becomes important for identifying the size of the second largest cluster is the offset of the second ranked wave from the champion. Offset if the difference in social space between the sinusoid under discussion and the champion sinusoid. As we normalized the champion to start at $s = 0$, the offset of any wave will simply be equal to the location s where they start. To identify which option is going to be the second most popular in the long run we go down the preliminary ranking. If the second ranked option in the preliminary ranking has offset exactly equal to zero this means that this sinusoid is positive in space $(0; S/2)$ and negative in $(S/2; 1)$. But so is the champion wave. And we know that champion dominates any other wave completely in the space $(0; S/2)$. Therefore, the wave with offset zero will never show up in the long run frequency distribution with the positive weight. Thus, we can discard the wave and remove it from the rankings.

Then we go down to the rankings until we find the wave with offset $s_i > 0$. Consider how the share of social space dominated by this option depends on s_i. If $s_i < S/2$ we know that this wave will be positive on $(s_i; S/2+s_i)$ and negative on $(0; s_i) \cup (S/2+s_i; 1)$. However, on$(s_i; S/2)$ if will be dominated by the champion wave, therefore this option will only acquire $S/2 + s_i - S/2 = s_i$ part of the social space. In case when $s_i > S/2$ the wave is positive on $(s_i; 1) \cup (0; S/2-1+s_i)$. But it is dominated by the champion on fraction $(0; S/2-1+s_i)$, and thus, it obtains the section $1 - s_i$. It can be easily seen that as s_i goes from $s_i = 0$ to $s_i = S/2$, the part dominated by the second ranked wave also increases linearly from zero to $S/2$. As s_i continues move to the right after passing $S/2$, the part dominated by the wave decreases linearly from $S/2$ to zero (when $s_i = 1$).

Now, as initial conditions are random and agents are distributed uniformly over the social space, the probability of choice of any s_i is constant. Therefore, we can calculate that the average market share of the second ranked practice in the long run $F_2 = 1/4$. In order to build the case for $R_n > 2$ notice that there are two actual waves corresponding to the market share of $1/4$. These are $s_i = S/4$ and and $s_i = 3S/4$. It does not matter for the further calculations which of them we choose to be present while considering $R_n > 2$ options. Because of the circularity of social space, there will always be two waves corresponding to each share distribution. Without loss of generality we always choose to consider that the wave with the smaller s_i is at place. Thus, for later options there will always be some space $(0; W > S/2)$ that we be occupied by stronger waves and the space $(W; 1)$ left to be distributed among the weaker waves.

$R_n = m > 2$: As pointed out in case $R_n = 2$, by now the social space $(0; W)$ is already distributed. Then $W = \sum_{j=1}^{m-1} F_j$. Denote the size of the remaining social space $w = 1 - W$. Then, w is the size of the not-yet-distributed portion. In this case, the we know that the weakest wave already assigned its long-run share is the wave with the positive part on $(S/2-w; 1-w)$. Therefore, any wave to be placed next on the social space has to have the offset more than $s_i > S/2 - w$. This is because the waves with less offset will always be dominated by the already the most popular $m - 1$ waves. Therefore, while going down the preliminary ranking we through out all the waves with offset less then $S/2 - w$, and concentrate only on offsets with higher offsets.

Consider how long-run market share depends on s_i in this case. With s_i increasing from $S/2 - w$ till $S/2$ the share increases linearly from zero to w. In the section where $s_i \in (S/2; 1-w)$ the share is constant at w. Once s_i passes $1 - w$ the share decreases linearly and reaches zero at $s_i = 1$. In this case taking the average long run market size and converting it to shares results in

$$F_m = \frac{w}{1+2w}.$$

It is easy to check that $m = 2$ also obeys this formula (although the calculation of F_2 was slightly different, it was in fact the specific case of these calculations). $\square$

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching up with the Joneses," *The American Economic Review*, 80(2), 38–42.

Ahn, I., and M. Suominen (2001): "Word-of-Mouth Communication and Community Enforcement," *International Economic Review*, 42(2), 399–415.

Arnade, C., M. Gopinath, and D. Pick (2008): "Brand Inertia in U.S. Household Cheese Consumption," *American Journal of Agricultural Economics*, 90(3), 813–826.

Bala, V., and S. Goyal (1998): "Learning from Neighbours," *The Review of Economic Studies*, 65(3), 595–621.

——— (2001): "Conformism and diversity under social learning," *Economic Theory*, 17(1), 101–120.

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

——— (1993): "The Economics of Rumours," *The Review of Economic Studies*, 60(2), 309–327.

Banerjee, A. V., and D. Fudenberg (2004): "Word-of-mouth learning," *Games and Economic Behavior*, 46(1), 1–22.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *The Journal of Political Economy*, 100(5), 992–1026.

Bjonerstedt, J., and J. Weibull (1995): "Nash Equilibrium and Evolution by Imitation," in *The Rational Foundations of Economic Behavior*, ed. by K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt. McMillan, New York.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001): "Panel data analysis of household brand choices," *Journal of Econometrics*, 103(1-2), 111 – 153.

Constantinides, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *The Journal of Political Economy*, 98(3), 519–543.

Cowan, R., W. Cowan, and G. P. Swann (1997): "A Model of Demand with Interaction among Consumers," *International Journal of Industrial Organization*, 15, 711–732.

Ellison, G., and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *The Journal of Political Economy*, 101(4), 612–643.

——— (1995): "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*, 110(1), 93–125.

GLAESER, E. L., AND J. A. SCHEINKMAN (2000): "Non-Market Interactions," working paper 8053, NBER.

GONZALEZ-AVELLA, J. C., V. M. EGUILUZ, M. G. CONSENZA, K. KLEMM, J. L. HERRERA, AND M. SAN-MIGUEL (2006): "Local versus Global Interaction in Nonequilibrium Transitions: A Model of Social Dynamics," *Physical Review E*, 73, 1–7.

GRABOWSKI, H. G. (1970): "Demand Shifting, Optimal Firm Growth, and Rule-of-Thumb Decision Making," *The Quarterly Journal of Economics*, 84(2), 217–235.

HANSEN, F. (1972): *Consumer Choice Behavior: A Cognitive Theory.* The Free Press, New York.

PARK, J.-W., AND R. S. WYER (1993): "The Cognitive Organization of Product Information: Effects of Attribute Category Set Size on Information Recall," *Journal of Consumer Psychology*, 2, 329–357.

ROB, R., AND A. FISHMAN (2005): "Is Bigger Better? Customer Base Expansion through Word-of-Mouth Reputation," *The Journal of Political Economy*, 113(5), 1146–1162.

SCHLAG, K. (1998): "Why immitate and if so, how? A boundedly rational approach to multiarmed bandits," *The Journal of Economic Theory*, (130–156).

SMALLWOOD, D. E., AND J. CONLISK (1979): "Product Quality in Markets Where Consumers are Imperfectly Informed," *The Quarterly Journal of Economics*, 93(1), 1–23.

SMITH, L., AND P. SORENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68(2), 371–398.

WYER, R. S., AND T. K. SRULL (1989): *Memory and Cognition in its Social Context.* Lawrence Erlbaum Associates: New Jersey.