

MERIT-Infonomics Research Memorandum series

ICT Revolution, Globalization and Informational Lock-in

Bulat Sanditov

2004-013



*MERIT – Maastricht Economic Research
Institute on Innovation and Technology*

PO Box 616
6200 MD Maastricht
The Netherlands
T: +31 43 3883875
F: +31 43 3884905

<http://www.merit.unimaas.nl>
e-mail: secr-merit@merit.unimaas.nl



International Institute of Infonomics

c/o Maastricht University
PO Box 616
6200 MD Maastricht
The Netherlands
T: +31 43 388 3875
F: +31 45 388 4905

<http://www.infonomics.nl>
e-mail: secr@infonomics.nl

ICT Revolution, Globalization and Informational Lock-in

*Bulat Sanditov**

June 8, 2004

Abstract

We examine a model of social learning in networks following the lines of Bala and Goyal (1998, 2001). As a model of agents' behaviour we have chosen the model of informational cascades of Bikhchandani et al (1992). Similarly to Bala and Goyal we find that the higher the 'degree of integration' within the society is, the more likely it is that conformity of actions will arise. However, unlike their results our model suggests that in the presence of informational externalities globalisation of informational flows, expressed in the increasing density of communication channels in a network, may drive down the expected social welfare.

JEL Codes: D83, O33, F02.

Keywords: Social learning, Cascades, Globalization, Networks.

*MERIT, Maastricht University, P.O.Box 616, 6200MD Maastricht, The Netherlands. Tel: +31 (0)43 3883867. Fax: +31 (0)43 3884905. E-mail: b.sanditov(at)merit.unimaas.nl. I thank Robin Cowan, Saurabh Arora, Saeed Parto, Andreas Reinstaller, Lorenzo Zirulia, participants of the seminars at MERIT(Maastricht), CEFIR(Moscow), WU(Vienna), USE(Utrecht), and NAKE Day (DNB,Amsterdam) for their helpful comments and suggestions. All remaining errors and shortcomings are solely mine.

1 Introduction

Recent major advances in information and communication technologies, sometimes referred as IT (or ICT) revolution, seem to have radically changed our lives, and promise even more change in the years to come. One can ‘travel’ to the places, one never knew before, not leaving a couch in front of one’s TV, or learn an amazing variety of languages, or cooking styles just from the computer on one’s desk. Although the benefits of these innovations appear to be obvious, we may wonder whether this increasing intensity of information flows is always a good thing.

Let us try to project the history into the future. Globalization is not something absolutely new and unique for the age of IT. Indeed, in the magnitude of the effects of globalization perhaps we are yet to reach the time of the Great Geographical Discoveries, which has shaped the world as we know it today. On the one hand, one could say that the mankind as a whole benefited from those discoveries in the sense of bringing new, more productive technologies to the places where they had not been used before, to name only a few - the wheel and the horse to Americas; corn and the potato to Europe. But, on the other hand, it came at a price, and this price was paid mainly by local populations. The new technologies required land and labour, hence the land and labour had been expropriated from locals. Moreover, wherever the local supply of workers was inadequate for large-scale production slaves were imported. Thus, while new technologies were more efficient, the local populations were excluded from the benefits of the increased productivity.

However, one may argue that the things are different now. In contrast with the early days, ‘ICT globalization’ is based not so much on the exchange of the material goods, but rather on the exchange of information. In a ‘knowledge economy’ technologies are no longer embedded in capital goods, but rather in the information. If so, then more productive technologies can be transferred worldwide, and it will not be easy if not impossible to exclude somebody from the profit generated by his knowledge. Would it mean that the ‘information-based’ globalization is going to make us better off, while at the same time letting us avoid the worst consequences?

In a recent series of rigorous and extensive studies Bala and Goyal (referred as BG

henceforth) propose that the answer is likely to be ‘yes’ (Bala and Goyal, 1998, 2001),¹ because ‘better practices’ always win in the long-run (under some conditions, as to be discussed later). Acknowledging the importance of their work, we shall nevertheless argue that the prospects of IT globalization might be not so charming, if one takes into account the possibilities of a ‘lock-in’ emphasized by the students of economics and history of technical change (David, 1985, Cowan, 1990).

Although most models of ‘lock-in’s assume some kind of ‘network effects’, i.e. payoff externalities (Arthur, 1994), there are some which consider only effects of information (Bikhchandani et al., 1992, Banerjee, 1992)². To make the argument we have chosen the model of Bikhchandani, Hirshleifer and Welch (1992) (BHW henceforth). Our main interest is the relationship between the scale of communication, expressed as network density, as well as the structure of the communication network on the one hand, and social welfare on the other.

The rest of the paper is organized as follows. In the next section we will review the basic BHW model of informational cascades, and agents’ decision rule. Then we will formulate and examine a model of informational cascades in a network. Discussion of our results follows in Section 4. The last section concludes.

2 BHW model

There are several models of social learning and imitation besides BHW (e.g., Ellison and Fudenberg, 1995), but BHW has some features which makes it attractive for our purposes:

- (i) The model is based on *individual rationality*. Whether this concept is realistic or not remains beyond the scope of this paper, but at least it exempts us from making too many *ad hoc* assumptions about agents’ motivations, decision processes, role of social interactions and so on.
- (ii) The model captures (although very schematically) both emergence and diffusion of behavioural patterns/“common practices”. It is essential that the model allows for a lock-in to (*ex-post*) suboptimal outcomes.

¹This is my interpretation of their results

²Presence of ‘network effects’ is likely to strengthen our results

- (iii) The model does not require superrationality from our agents. The decision rule is simple, transparent, and intuitive.
- (iv) The model does not require superrationality from us either. It is easy to handle and allows us to generate treatable results with fairly modest means.

On the top of that, the predictions of the model have been proven to fit with the results of experimental studies (Anderson and Holt, 1997). Below we describe the BHW model as it was presented in Bikhchandani et al. (1998).

There are n identical agents each of whom has to choose between the two actions: to switch to a new ‘practice’ (adopt the new technology), or to keep *status quo* (reject the technology). The payoff to adopting the technology, V , is either 1 or -1, with equal probability. The payoff to keeping the *status quo* is 0. BHW assume *sequential* decision making, such that one (and each) period only one agent makes his choice. The order in which agents decide is exogenously given and known to all.

The information available to an agent includes both a (private) signal and public information which consists of the history of his predecessors’ decisions. The agent’s signal, v_t ($t = 1..N$), is either *High* or *Low*. If the technology is worth adopting, i.e. $V = +1$, the probability to receive a *High* signal is $p > 0.5$, while if $V = -1$, the probability of *High* is $(1 - p)$. The precision of the signal, p , is the same for all agents. The signals are identically distributed and independent conditional on V . Agents are risk neutral and choose the action which has a higher expected payoff. If an agent is indifferent between the two alternatives we impose a tie-breaking rule somewhat different from the original BHW setting,³ namely we assume that in case of a ‘draw’ he puts slightly higher weight on his private information and follows his signal. This kind of tie-breaking rule was used by Anderson and Holt in their experiments (Anderson and Holt, 1997).

Details for the solution of the agent’s problem can be found in Appendix A (or better in the original work). The agent’s decision rule is as simple as this: if by the time the agent is to take the decision the difference between the number of the agents who adopted (adopters) and the number of the agents who rejected the technology (non-adopters), d ,

³BHW assume that when the expected payoffs are equal, the agent randomises (tosses a coin). Although probabilities to be locked in a cascade differ slightly for the different tie-breaking rules, the results are essentially the same.

is greater or equal to $+2$, then he adopts regardless of the private signal; if the difference is less or equal to -2 , he should reject whatever his signal is; if neither is the case, then he follows the signal, i.e. adopts when the signal is *High* and rejects when it is *Low*.

BHW define an *informational cascade* as a situation where the decision of the agent does not depend on his private information ($d \leq -2$ or $d \geq 2$). Once it happens, further accumulation of public information stops, and conformity of the actions arises. Indeed, after the first agent who decides to discard his private signal and ‘join the herd’, each of the following agents will face exactly the same decision problem and should ‘join the herd’ as well. If the system converges to adopting the technology it is said to be an *UP cascade*, if, in contrast, it converges to rejecting the technology it is said to be a *DOWN cascade*. Without loss of generality in what follows we will assume that the **true value of the technology is 1**. Therefore an UP cascade is the ‘correct’ cascade.

With the signal probabilities and the decision rule one can find some characteristics/statistics of the process which will be useful for the analysis of the informational cascades in a network. These characteristics include total payoff (sum of all payoffs) normalized by the size of population ($W \in [0, 1]$), the share of the agents who end up in an UP cascade, S_{up} , and the share of unlucky ones in DOWN cascades, S_{down} . One can also interpret S_{up} and S_{down} as the probabilities for an agent to find himself locked into an UP or a DOWN cascade respectively. The expressions for W , S_{up} , and S_{down} can be found in Appendix A.

3 Informational cascades in a network

The BHW model of informational cascades offers a compelling explanation of imitation and herding behaviour observed in many circumstances (see Bikhchandani et al. (1998) for a review) based on individual rationality. Not the least, as already has been mentioned above, the predictions of the model are in good terms with the results of experiments. Nevertheless, we cannot use the model for our purposes without some adaptation. Large-scale social or economic phenomena such as diffusion of a new technology or social norms occur in large populations which are often scattered in space. Therefore to model such a process one would have to take into consideration that actors hardly have an opportunity to ob-

serve actions of all other agents in the population. The scope of such information is rather limited to the actions of their close neighbours with whom they interact/communicate on a regular basis. As a result the structure of the interactions/communications might come into play.

There are two ways in which ICTs may affect the structure of communication in the society. The first effect is known as the ‘death of distance’, i.e. decreasing importance of geographical proximity for making and maintaining a contact. The other effect is the increasing density of the communication network, or put it differently, the growing size of information-sharing communities. Although the ICT revolution is likely to increase both the geographical scope and the density of the communication network, these two effects might have different implications for the social welfare. Thus our model should be formulated in such a way that it would allow us examine the effects separately.

3.1 The model

To model informational cascades in a structured population we shall modify the basic BHW model described above. Instead of assuming that agents observe the whole population of the size N , we shall assume that they can observe only a subset of the agents with whom they communicate on the regular basis. The communication structure of the population can be viewed as a graph G : a node i of G represents agent i , a (non-directed) edge ij is the ‘informational channel’ connecting i and j . The degree of the node i , k_i , is the size of the i th neighbourhood.⁴

Following Bala and Goyal (1998) we assume that the rationality of our agents is bounded: an agent does not attempt to infer the decisions of unobserved agents (neighbours of neighbours and so on) from the actions of observed agents (neighbours). One may think that our agents are ignorant of the world outside their local communities. As for the agents’ perception of communication structure within their neighbourhoods we assume that an agent believes that none of his neighbours are neighbours of each other, and therefore he treats observed actions as independent (conditional on V).⁵

⁴The communication structure of the BHW model would be represented by a complete graph of N nodes.

⁵Although it is quite a deviation from the original setting of BHW model, it would not change the

Because the range of all possible network configurations for any interesting number of agents, N , and edges, K , can be extremely large, we limit ourselves to the study of informational cascades in a particular class of networks - networks generated by ‘ β -algorithm’ used by Watts (1999). According to this algorithm a network is constructed from a lattice graph (1-lattice), where each node is connected with k of its closest neighbours, by rewiring each of its edges with some probability β . Therefore a network is characterized by only three parameters: the number of nodes (N), average degree of a node (k), and the rewiring probability (β). An advantage of this approach is that while random (and therefore generic), ‘ β -networks’ have similar topology (in terms of the path lengths, clustering and so on (Watts, 1999)) for the same parameter β . The algorithm allows us to generate a wide spectrum of networks with the given degree of randomness ranging from perfect lattice at $\beta = 0$ to a random graph at $\beta = 1$ (for the same number of edges ($kN/2$)).⁶ Moreover, our experiments with some other types of networks (and with substrates other than 1-lattice) suggest that the results reported here are rather general.

With the ‘ β -algorithm’ we can disentangle the two effects of the globalization of information flows mentioned above. Indeed, since parameter β relates to the share of the distant links, the ‘death of distance’ effect can be modelled by increasing β . Furthermore, for a given number of agents the average degree of a node, k , gives us the average size of information-sharing community, and we can analyse the effect of the density of the communication network on the social welfare by changing k (for fixed N and β).

3.2 Results

Figure 1 presents the results of our simulations of informational cascades ($p = 0.75$) in a range of networks with $N = 200$, for $\beta = 0$ (1-ring), and $\beta = 1$ (random graph). For each value of k and β we generated 500 networks, and for each of the networks we drew 50 realizations of signal-agent sequences (signals and the order of their arrival). Values of S_a , S_{up} , S_{down} , and $S_{cascade} \equiv S_{up} + S_{down}$ were estimated for each realization and averages reported.

agents’ decision rule.

⁶Our simulations show no difference between networks with $\beta = 1$ and random graphs with the same number of edges ($kN/2$). Therefore in what follows we will use the terms interchangeably.

Two approximations to our model for which we can easily find exact solution can be instrumental for understanding the results of our simulations. The first is the BHW model discussed above, with n equal to the size of average neighbourhood of G , which is $(k + 1)$. For $k \ll N$ one may think of it as an approximation of G by ‘caveman’ graph, a graph which consists of $N/(k + 1)$ isolated subgraphs (‘caves’), where each subgraph is a complete graph of $(k + 1)$ nodes. Characteristics S_a , S_{up} , S_{down} as functions of k are given by corresponding expressions for the BHW model (Appendix A, $n = k + 1$) and shown at Figure 1 (red circles). At the limit of $k = N - 1$ (complete graph) our simulations for informational cascades in all networks must produce the same results as BHW with $n = N$.

The other approximation uses the ‘mean-field approach’ (Weisbuch et al., 2000). In this approach we approximate actual realizations of random variables by corresponding values averaged over the population and solve dynamics for those average values. Hence the share of positive signals is approximated with the probability of the positive signal, and we would examine the behaviour of a ‘representative agent’ in a ‘representative neighbourhood’ (Appendix B). In doing so we neglect possible fluctuations of the random variables assuming that in the limit of large population the fluctuations may have only minor impact on the dynamics of the system. The mean-field solution for our model is shown by dotted line at Figure 1.

As one could have already guessed, these two approximations lie at the extremes from the point of view of the strength of local effects. In the ‘caveman’ world local cohesion is at maximum: within the local community all agents are neighbours of each other. As a result there is high correlation between neighbours’ choices. In contrast, the mean-field approach denies possible local fluctuations,⁷ thus there is no relationship between neighbours’ choices other than on the level of the whole population.

Note, that even though all networks generated by β -algorithm are connected in the graph-theoretic sense (Watts, 1999), i.e. for any pair of nodes i and j there is a path connecting them, for a small k the society is likely to be “effectively disconnected”: for the action of agent i to affect the choice of agent j , by the time when agent j is to make his decision all agents on one of the paths connecting i and j must have made their choice,

⁷both fluctuations of signal-agent series and of structure of the generated network, G .

which is rather unlikely when k is small (average distance in G is large). For this reason interactions effectively ‘localized’ within a community.

These local effects do make a difference. Due to informational externalities the model has a self-reinforcing (‘positive feedback’) mechanism, and sufficiently strong fluctuations have potential to grow up to the size of the local community. But the strength of those effects depends on the structure of the network. In a network with high intensity of local interactions we may expect that the local effects prevail and cascades will be formed in each of the local communities almost independently, therefore BHW provides a reasonable approximation; while in a network with loose structure of local contacts, local fluctuations do not have significant impact on the overall outcome, and the process will be governed by averages over the population, hence a mean-field approach does better in this case.

It seems that the results of our simulations are consistent with the logic explained above (see Figure 1). Indeed, the 1-lattice ($\beta = 0$) obviously has much more prominent local structure than a random graph ($\beta = 1$), and as one can see the results for 1-lattice can be fairly well approximated by the BHW/‘cavemen’ model. The results for the random graph, on the other hand, go particularly well with the mean-field solution for $k \ll N$. As k increases and comes close to N , the global local effects starts to dominate the picture (the size of the local community approaches the size of the whole population), therefore the results for all networks expectedly converge to the corresponding values of the BHW model ($n = N$).

4 Discussion

From the point of view of the social planner it is essential to know which kind of the communication structure (represented by our graphs) does better in terms of the social welfare. In our model social welfare could be approximated by overall social outcome, the sum of agents’ payoffs normalized by the size of population.

A remarkable feature of Figure 1 is that for the same number of edges a random graph significantly outperforms 1-lattice (except the region of very small k , where threshold effects cannot be neglected). To explore this point more thoroughly we run simulations fixing k and changing the rewiring parameter β . The results of the simulations are shown

in Figure 2 ($N = 200$, $k = 10$). As one can see ‘randomisation’ of the structure by rewiring edges, which creates more of ‘distant’ links at the expense of the ‘local’ ones, increases the social welfare: the overall outcome grows fast until it reaches plateau at $\beta \sim 0.5$. This increase seems to be related to the opposite effect that rewiring has on the shares of UP and DOWN cascades (Figure 2a and 2b). The origin of this effect is rather clear. In a network with high local cohesion, such as 1-lattice, cascades emerge almost independently in each of the local communities and we would see a set of subpopulations some of which are evolving toward UP cascades and others toward DOWN cascades. For large N/k the share of subpopulations in UP cascades must be about the probability of UP cascade in BHW model, and the same for DOWN cascades. If we start to create links between such local communities the agents who get links outside their subpopulation become exposed to ‘practices’ in other communities. Because the share of communities in UP cascades is much higher than the share of agents in DOWN cascades, an agent with ‘external’ links is more likely to be connected to the communities in UP cascade. This would not cardinaly change the agent’s perception of the technology, if the agent already belonged to the community which is going to converge to UP cascade, it would only reinforce the tendency. Consequently, the share of agents locked in UP cascades increases with rewiring. In contrast, if the agent belonged to a community which tends to converge to a DOWN cascade rewiring would have opposite effect, damping or even preventing the growth of the ‘negative’ fluctuation. One can also see it from Figure 1 (diagrams 1a and 1b).

Another striking feature of Figure 1 is that an increase in the density of connections (increase in k) does not always improve social welfare. Indeed, for a network “more connected” than caveman graph the social outcome W is not a monotonic function of k : first, it rapidly increases with k until some k^* , and then starts to fall. The explanation is straightforward. In a dense network ($k \sim N$) the size of local communities is comparable with the size of the population. As we already know, small ‘negative’ fluctuations may spread through the community, which now has size $\sim N$, leading to the outcome suboptimal with respect to social welfare.

Do our results differ from those of BG? To see that let us first briefly restate the results of their work relevant to our study. BG consider ‘boundedly rational’ agents repeatedly

experimenting with technologies producing stochastic outcomes. The expected outcome of a technology is determined by the state of nature. Agents employ Bayesian learning in updating their beliefs about the state of the nature (the quality of the technology) with the experience of their own and the other agents with whom they have links (neighbours). In the settings of Bala and Goyal (1998) (more general than Bala and Goyal (2001)) BG proved that in a connected society of homogeneous agents the better technology always drives out inferior ones (complete social learning), if the following conditions on beliefs and the structure of the relationships are satisfied: (a) agents' priors are dispersed enough (else agents may be stuck with the inferior technology from the beginning);⁸ (b) there is no 'royal family' (or 'nearly royal families'), which is a finite subset of the population such that all agents from this subset are 'visible' to all other agents in a network, or the size of the 'royal family' is sufficiently small with respect to the size of agents' neighbourhood (to ensure that a 'bad' experience of the 'royal family' will not overwrite locally emerging knowledge). In addition, in the case of agents with heterogeneous preferences lock-in to an inferior technology may happen if the society is not 'group-wise' connected, i.e. when an agent (or several agents) of one type is surrounded by the agents of the other type effectively separating this agent from others of his type (Bala and Goyal, 2001).

Let us hypothesize what may happen in the world of BG agents, if we start to increase the number of informational channels (add edges into G). First, and obvious, at some point a disconnected society (such as represented by a 'caveman graph') will become connected, that increases the chances that the better technology will prevail. Second, one may think that the heterogeneity of priors is 'effectively' increasing as we connect previously disconnected groups, hence decreasing the probability of an 'inferior lock-in'. Third, if we add more new links, the effect of the 'royal family' (if one existed in the initial network) would be weakened because each agent's 'reference group' is growing. And finally, increasing the density of the communication network in the society of heterogeneous agents would lead to higher chances of 'group-wise' connectivity.

Although we cannot compare different models, it is almost obvious that implication of the models for the issue we are interested in might be different. Intensification of information exchange expressed in the increasing density of the network links would un-

⁸Condition H in Bala and Goyal (2001).

ambiguously benefit the social welfare in the BG model, while in our model with the BHW model of agents' behaviour it may lead to a 'global lock-in' to inferior practices (although at the beginning of the process, i.e. with small k , it significantly improves social well-being, see Figure 1c).

It is not a secret for the reader, that the divergence of our results from the results of BG could have been predicted already from the differences in the settings of their model and the BHW model used by us (the order of decisions, the scope of publicly available information and so on). Hence a would-be productive discussion should shift to the discussion of the realism of the assumptions of the BHW model. However, for such discussion one is better to refer to the original work of BHW (Bikhchandani et al., 1992, 1998)⁹. Here we would limit ourselves only to several remarks on the difference between the assumptions of the BG model and the BHW settings.

First, in many circumstances private information is unlikely to be shared, either because it has commercial value, and transactions costs of buying information from scattered sources might be high, or simply because of its 'tacit' nature. Furthermore not everybody "practices what he preaches", and not every piece of information which one is able to collect can be trusted ("actions speak louder than words"). Second, the payoffs to adopting might be realized only in a distant future (in our model, only after the whole process is already over). We could recall an example of tobacco, which was introduced into Europe, and spread as a cure for headache. Third, not always decisions about adoption of certain technology are simultaneous as in BG model (instead of *sequential* order of the BHW model). In addition, switching from one technology to another may be costly, as a result once one has invested in a technology he could be reluctant to replace it with another technology any soon. Should one argue that in 'true' long-run the true state of nature (quality of the technology) will be revealed (as in the example with tobacco) due to some information leakages, and switching costs are spread through the time and not so relevant, we shall remind that, first, in the long-run 'we are all dead'- by the time the society should figure out the true quality of the technology, this technology may be substituted by another (perhaps inferior as well). Consider also that the consequences of today's deci-

⁹One can also find the annotated bibliography on informational cascades/herding on the website devoted to informational cascades <http://welch.som.yale.edu/cascades/>.

sions may turn to be irreversible.¹⁰ Not the least to mention, humans tend to rationalize their choices ex-post, and, arguing over today's practices, which have 'survived' in the evolution, one would have to resort to counterfactual reasoning not very welcomed in the modern economic theory.¹¹

It seems that the assumption most crucial for our results concerns with the scope of information available to an outside observer. In the 'previous-signal-observable' (PSO) regime of the BHW model, where not only actions chosen by agents, but also their signals are publicly observable, the true state of nature will be revealed relatively fast therefore socially optimal outcome will be reached, and no cascade will emerge. We may expect similar results in a network. In addition, increasing density of the communication network is likely to speed up the process of learning. Indeed, the results of simulations for PSO regime in our model presented in Figure 3 show that social welfare monotonically increasing with the density of the network.

5 Conclusions

Now we are ready to answer the question raised in the beginning of this paper: is globalisation of information flows always good for human society? From (fairly limited) prospective of our model we can say that all depends on how this process affect the structure of communications in our society. Namely, if the ICT revolution is about making the Big World small through enhancing exchange information between distant communities (which corresponds to rewiring in terms of our model), the answer would be '*yes*'. Exposing ourselves to experiences from different parts of the world might prevent a lock-in to inferior practices. In this point we share the conclusions of BG. In addition, our model suggests that we can expect the effect of such globalisation to be higher at the beginning (when β is relatively small) than at the end of the process (interim as well) as one can see from Figure 2.¹² This is the 'bright side' of globalisation.

¹⁰Not too much we can do about global warming today, which is likely to be related to air pollution in the last century. Even less could be done about some animal species which disappeared from the surface of the Earth once and forever, directly or indirectly due to human activities.

¹¹see Cowan and Foray (2002) for a discussion.

¹²Compare with BG's 'degree of integration' η .

On the other hand, decreasing costs of communication and bringing new technologies for data handling¹³ improvements in ICT increase not only the geographical reach of information flows, but also the size of ‘information sharing’ communities which makes the whole world one local community (k increases and approaches to N), which would lead to convergence to a uniform culture (perceptions, norms, values etc.) and elimination of diversity. This should not be cause for concern, if the ‘global culture’ were to select only the best practices from the variety of those existing today, eliminating inferior ones. The bad news is that there is no such a warranty. It might well happen that the historical (path-dependent) process would lead us to an inferior state of affairs in so far as social well-being is concerned.

There are several ways to deal with the ‘global informational lock-in problem’ which follow directly from the model (if we exclude radical, but unfeasible solutions such as restrictions of public access to ICT). The first one coming from the studies of technological lock-ins is to prevent ‘early standardization’ (early emergence of informational cascades) and encourage diversity of opinions, for instance support views alternative to the mainstream. Another recipe is to promote wide discussion and engage different communities into dialog, so to stimulate agents to reveal their private information (PSO regime of BHW). As we have seen under PSO regime as in the model of BG increasing density of information flows unambiguously enhances social welfare.

References

- ANDERSON, L. AND C. HOLT (1997), “Information Cascades in the Laboratory,” *American Economic Review*, **87**, 847-62.
- ARTHUR, W.B. (1994) *Increasing returns and path dependence in the economy* (Ann Arbor: University of Michigan Press).
- BALA V., and S. GOYAL (1998), “Learning from Neighbours,” *Review of Economic Studies*, **65**, 595-621.

¹³And institutional changes related to it (Petersen and Rajan, 2002, e.g.).

- BALA V., and S. GOYAL (2001), "Conformism and diversity under social learning," *Economic Theory*, **17**, 101-120.
- BANERJEE, A. (1992), "A Simple Model of Herd Behaviour," *Quarterly Journal of Economics*, **107**, 797-817.
- BIKHCHANDANI S., D. HIRSHLEIFER and I. WELCH (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, **100**, 992-1026.
- BIKHCHANDANI S., D. HIRSHLEIFER and I. WELCH (1998), "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *Journal of Economic Perspectives*, **12**, 992-1026.
- COWAN, R. (1990), "Nuclear Power Reactors: A Study in Technological Lock-in," *Journal of Economic History*, **50**, 541-67.
- COWAN, R., and D. FORAY (2002), "Evolutionary Economics and the Counterfactual Threat: On the Nature and Role of the Counterfactual History as an Empirical Tool in Economics," *Journal of Evolutionary Economics*, **12**, 539-62.
- DAVID, P.A. (1985), "Clio and the Economics of QWERTY," *American Economic Review*, **75**, 332-37.
- ELLISON, G., and D. FUDENBERG (1995), "Rules of thumb for social learning," *Journal of Political Economy*, **110**, 93-126.
- PETERSEN, M., and R. RAJAN (2002), "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, **57**, 2533-70.
- WATTS, D.J. (1999) *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton: Princeton University Press).
- WEISBUCH, G., KIRMAN, A., and D. HERREINER (2000), "Market Organisation and Trading Relationships," *Economic Journal*, **110**, 411-36.

Appendix A: BHW model

Decision rule. Consider the decision problem of an agent who is to take his decision, $\Delta_i \in \{Adopt, Reject\}$, at time t . Information available for him, I_t , consists of his private signal $v_t \in \{High, Low\}$, and the history of agents' actions prior to t , H_t . Expected payoff to adopting is

$$E[u_t] = 1 \cdot \Pr(V = 1|I_t) + (-1) \cdot \Pr(V = -1|I_t).$$

He adopts if $E[u_t] > 0$, rejects if $E[u_t] < 0$. In case of a 'draw', $E[u_t] = 0$, he follows his signal, i.e. adopts provided that $v_t = High$, otherwise he rejects.

Applying Bayes' rule and taking into account that signals are conditionally independent (therefore $\Pr(I_t|V) = \Pr(v_t|V)\Pr(H_t|V)$) we can find the condition for an UP cascade, the situation where he adopts regardless to the signal (priors $\Pr(V = 1) = \Pr(V = -1) = 1/2$)

$$(1 - p) \Pr(H_t|V = 1) > p \Pr(H_t|V = -1).$$

Similarly for a DOWN cascade

$$p \Pr(H_t|V = 1) < (1 - p) \Pr(H_t|V = -1).$$

Suppose that by time $(t - 1)$ no cascade has happened, hence agents' actions unambiguously reveal their signals. Let n_+ be the number of adopters, and n_- be the number of non-adopters. Then the condition for UP cascade can be rewritten as

$$p^{n_+}(1 - p)^{n_-+1} > (1 - p)^{n_+}p^{n_-+1} \Leftrightarrow \left(\frac{p}{1 - p}\right)^{n_+ - n_- - 1} > 1.$$

By assumption $p > 1/2$, therefore condition for UP cascade is

$$n_+ - n_- - 1 \geq 1 \Leftrightarrow n_+ - n_- \geq 2.$$

Similarly for a DOWN cascade

$$p^{n_++1}(1 - p)^{n_-} > (1 - p)^{n_++1}p^{n_-} \Leftrightarrow \left(\frac{p}{1 - p}\right)^{n_+ - n_- + 1} < 1.$$

or

$$n_+ - n_- \leq -2.$$

Thus the decision rule is: if by the time when the agent is to take the decision the difference between the number of the adopters and the number of the non-adopters, $d \equiv n_+ - n_-$, is greater or equal to +2, then he adopts regardless of the private signal; if the difference is less or equal to -2, the agent rejects whatever is the private signal; if neither is the case, he chooses to follow the signal, i.e. to adopt if the signal is *High* and reject when it is *Low*.

Shares of cascades, total outcome. Without loss of generality let us assume that $V = 1$. The decision rule described above implies that an informational cascade may emerge only when t is odd. Suppose that there is no cascade by date $t = 2(l - 1)$. An UP cascade will emerge at $t = 2l + 1$, if both agents at $t = 2l - 1$, and $t = 2l$ receive positive signals. Therefore, the probability of UP cascade is p^2 . Similarly, the probability of a DOWN cascade is $(1 - p)^2$, and the probability of no cascade is $q \equiv 2p(1 - p)$.

If an UP cascade emerges at $t = 2l + 1$, then the share of agents in the UP cascade is $(n - 2l)/n$. Therefore the share of the agents to be locked in an UP cascade is

$$S_{up} = \sum_{l=1}^K \frac{2K - 2l}{2K} p^2 q^l = \frac{p^2}{1 - q} \left(1 - \frac{q}{(1 - q)K} + \frac{q^{K+1}}{(1 - q)K} \right),$$

where $K = n/2$. Note, that the factor $\frac{p^2}{1 - q}$ is the probability that the system eventually ends up in an UP cascade in the limit of $n = \infty$, P_{up} . It is rather natural to expect: the system converges to one of the cascades very fast (exponentially) therefore the share of agents which follow their signals is decreasing ($\sim n^{-1}$) and S_{up} must approach P_{up} . Similarly, for DOWN cascades

$$S_{down} = \sum_{l=1}^K \frac{2K - 2l}{2K} (1 - p)^2 q^l = \frac{(1 - p)^2}{1 - q} \left(1 - \frac{q}{(1 - q)K} + \frac{q^{K+1}}{(1 - q)K} \right).$$

The probability of DOWN cascade in infinite population P_{down} is the asymptotic value for S_{down} for large n .

If an UP cascade emerges at $t = 2l + 1$, then the total outcome normalized by n is $(n - l + 1)/n$, and if at this time a DOWN cascades emerges then the total outcome is $(l - 1)/n$. Therefore

$$W = \sum_{l=0}^{K-1} \frac{(2K - l)p^2 + l(1 - p)^2}{2K} q^l = \frac{1}{2} + \frac{2p - 1}{2(1 - q)} \left(1 - \frac{q}{(1 - q)K} + \frac{q^{K+1}}{(1 - q)K} \right).$$

Shares of cascades, S_{up} and S_{down} , total outcome, W , as functions of the size of population ($n = k + 1$) are shown at Figure 1 ($p = 0.75$).

Appendix B: Mean-field solution

Consider an agent i who is to make his choice, $\Delta_i \in \{Adopt, Reject\}$, at time t . His decision depends on three (random) variables: the number of adopters, ξ_i , and the number of non-adopters (those who rejected the technology), ζ_i , in i th neighbourhood Γ_i , and the private signal $\sigma_i \in \{High, Low\}$, in the following way

$$\Delta_i = \begin{cases} Reject, & \text{if } (\xi_i - \zeta_i) \leq -2, \\ Reject, & \text{if } |\xi_i - \zeta_i| < 2 \text{ and } \sigma_i = Low, \\ Adopt, & \text{if } |\xi_i - \zeta_i| < 2 \text{ and } \sigma_i = High, \\ Adopt, & \text{if } (\xi_i - \zeta_i) \geq 2. \end{cases} \quad (1)$$

Private signals σ^t , $t = 1..N$ are drawn at random with probabilities $\Pr(\sigma^t = High|V = 1) = p > 1/2$ and $\Pr(\sigma^t = High|V = -1) = 1 - p$. Values of ξ_i and ζ_i are determined by the history of signals $\{\sigma^1, \sigma^2, \dots, \sigma^{t-1}\}$ and the structure of the network G .

The state of the system is described by vector $S^t = (s_1^t, s_2^t, \dots, s_N^t)$, where s_i^t is the state of agent i : s_i^t is equal to Δ_i , if he already made his decision by time t , otherwise $s_i^t = 0$. The evolution of S^t is determined by individual decisions according to (1) and the structure of G . To solve the dynamics means to find the distribution of S^t from the distribution of the realizations of the signals $\{\sigma^1, \sigma^2, \dots, \sigma^{t-1}\}$ and the distribution of the order in which the agents make their decisions. The problem already seems to be complicated, but to make the bad things worse we have to take into consideration that for $\beta > 0$ the small-worlds algorithm generates G randomly and to answer the questions we are interested in, we should not only find the distribution of statistics of our interest (S_{up}, S_{down} , and W) at the terminal time $t = N$, but also obtain the distribution of those statistics over the 'ensamble' of the small-world networks generated for certain value of rewiring parameter β .

Although the problem is rather complicated, we can relatively easily find an approximation of the dynamics using mean-field approach. First, we approximate the random variables by their averages, hence before the emergence of a cascade the share of adopters and non-adopters are exactly p and $(1-p)$ respectively, i.e. we neglect fluctuations in realizations of the private signals. Second (and more brutal), instead of analysing what happens in agents' i (idiosyncratic) neighbourhood we are going to analyze a representative neighbourhood $\bar{\Gamma}$ of a representative agent. The term 'representative' in this context stands for 'averaged over the population', i.e. this assumption rules out all possible local fluctuations. This is a very rough approximation of the network G , justified only if local fluctuations do not grow. Nevertheless for networks such as trees where no i 's neighbours are neighbours of each other (so that informational cascades cannot spread), it may produce a reasonable approximation.

Let N_a^t be the number of adopters by time t , and $N_{na}^t = t - N_a^t$ the number of non-adopters (once we leveled down the description of the structure of G to representative neighbourhood, the state of the system can be fully described by these two numbers). Then the number of adopters in $\bar{\Gamma}$ is $k \frac{N_a^t}{N}$, where k is the average degree of a node in G (approximation for ξ_i). Similarly, $k \frac{t - N_a^t}{N}$ is the number of non-adopters in $\bar{\Gamma}$ (approximation for ζ_i). The difference between the number of adopters and non-adopters in $\bar{\Gamma}$ is

$$\bar{d}^t = k \frac{N_a^t}{N} - k \frac{t - N_a^t}{N} = \frac{k}{N} (2N_a^t - t). \quad (2)$$

Therefore, mean-field approximation for evolution of our system can be written as (compare with (1))

$$N_a^t = \begin{cases} N_a^{t-1}, & \text{if } \bar{d}^t \leq -2, \\ pt, & \text{if } |\bar{d}^t| < 2, \\ N_a^{t-1} + 1, & \text{if } \bar{d}^t \geq 2; \end{cases} \quad \text{and } N_{na}^t = \begin{cases} N_{na}^{t-1} + 1, & \text{if } \bar{d}^t \leq -2, \\ (1-p)t, & \text{if } |\bar{d}^t| < 2, \\ N_{na}^{t-1}, & \text{if } \bar{d}^t \geq 2. \end{cases} \quad (3)$$

By assumption $p > 1/2$, hence there no DOWN cascades in the mean-field solution. An UP cascade emerges if N is not too small. More precisely, UP cascade emerges at t^* as N_a reaches the 'critical value', N_a^* , such that $\bar{d}^t = 2$. From (2) and (1) we can find the time when UP cascades starts (if it does)

$$\begin{cases} \frac{k}{N} (2N_a^* - t^*) = 2, \\ N_a^* = pt^*; \end{cases}$$

or

$$t^* = \frac{2N}{k(2p-1)}, \quad \text{if } k(2p-1) \geq 2. \quad (4)$$

The condition in (4) insures that $t^* \leq N$.

From (4) the share of agents who are locked in an UP cascades, S_{up} , is

$$S_{up} = 1 - \begin{cases} 0, & \text{if } k(2p-1) < 2, \\ 1 - \frac{2}{(2p-1)k}, & \text{otherwise.} \end{cases} \quad (5)$$

Prior to t^* the number of non-adopters grows at rate of $(1-p)$: $N_{na}^t = (1-p)t$; at t^* the growth ceases and N_{na}^t remains the same till the end of the process. Thus the number of non-adopters at the terminal date $t = N$ is

$$N_{na}^N = (1-p) \cdot \min(t^*, N).$$

Finally, the total payoff normalized to the size of the population, N , is

$$W = (N_a^N)/N = \begin{cases} p, & \text{if } k(2p-1) < 2, \\ 1 - \frac{2(1-p)}{(2p-1)k}, & \text{otherwise.} \end{cases} \quad (6)$$

Share of UP cascades, S_{up} , ($S_{down} = 0$ as have been mentioned above) and total outcome, W , as functions of k are shown at Figure 1 ($p = 0.75$).

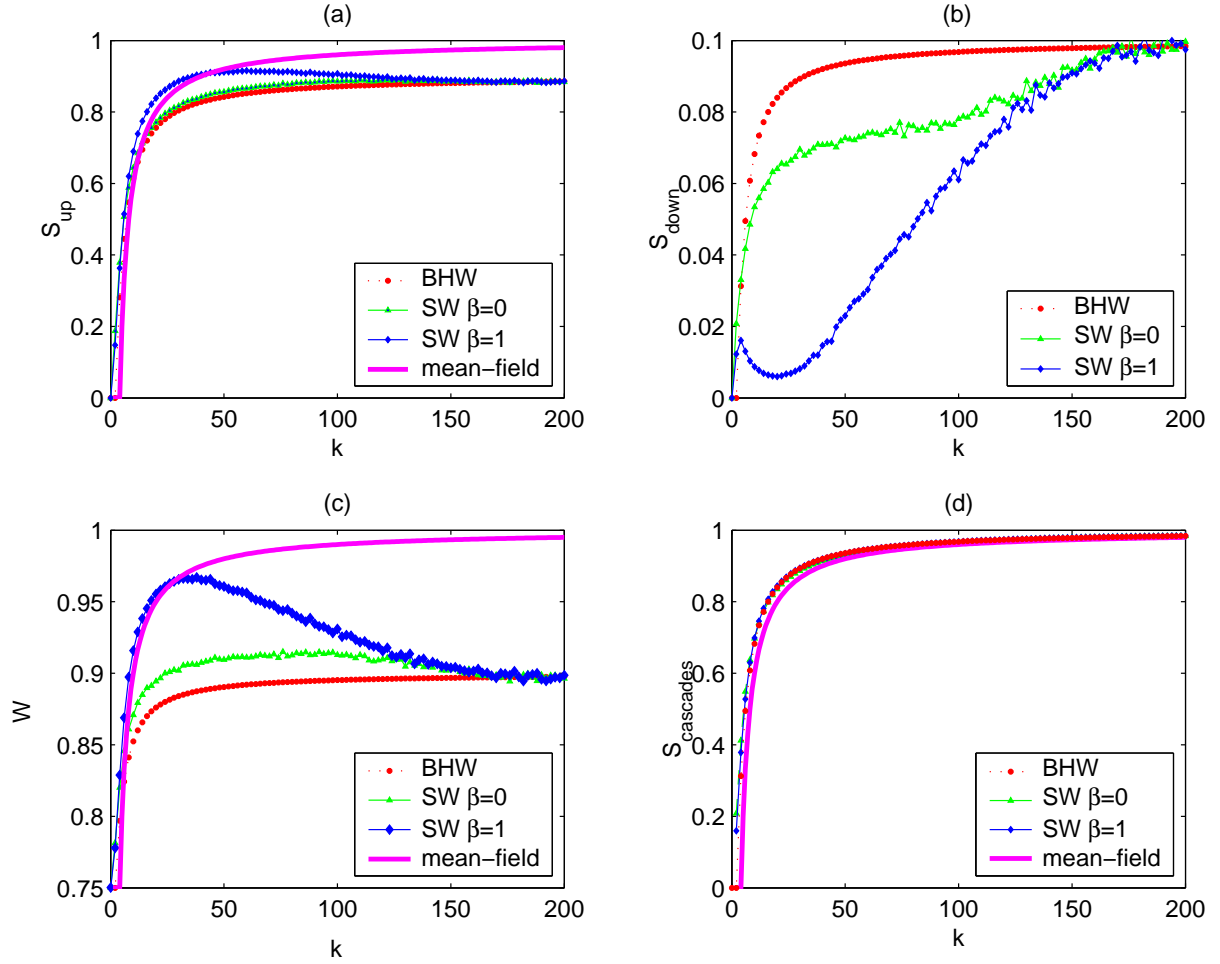


Figure 1: Effect of the network density, the number of edges $kN/2$ ($N = 200$, $p = 0.75$): (a) Share of agents in UP cascade, S_{up} ; (b) Share of agents in DOWN cascade, S_{down} ; (c) Social welfare (total payoff normalized by the size of population), W ; (d) Share of agents locked in cascades, $S_{cascades}$. β -networks: random graph ($\beta = 1$)- diamonds, 1-lattice/ring ($\beta = 0$) - triangle. The BHW/‘caveman’ model - circles. Mean-field solution - solid line (there are no DOWN cascades in mean-field solution (see Appendix B), i.e. $S_{down} = 0$.)

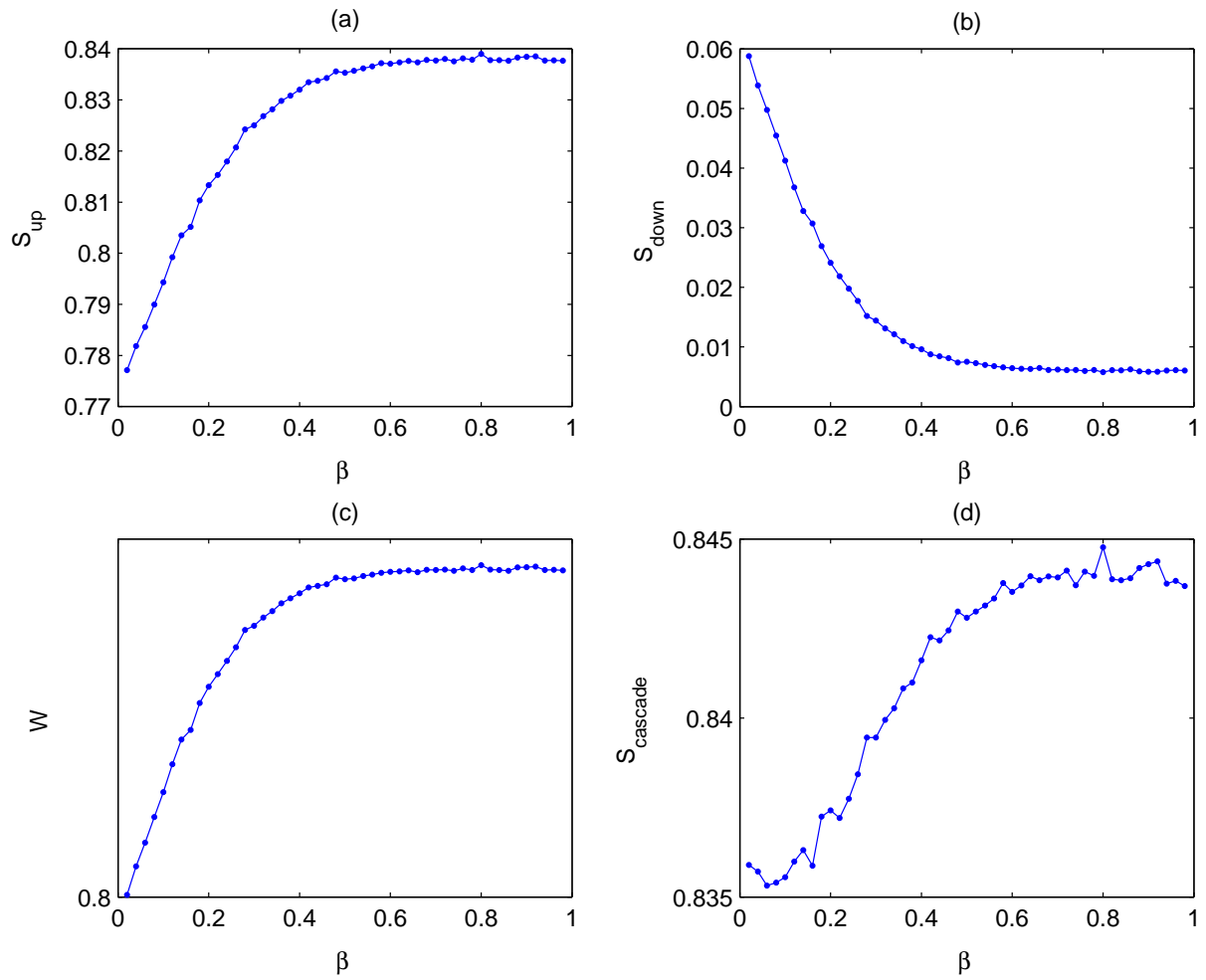


Figure 2: Effects of rewiring ($N = 200$, $k = 10$, $p = 0.75$). (a) Share of agents in UP cascade; (b) Share of agents in DOWN cascade; (c) Social welfare (total payoff normalized by the size of population); (d) Share of agents locked in cascades. Randomness increases with β .

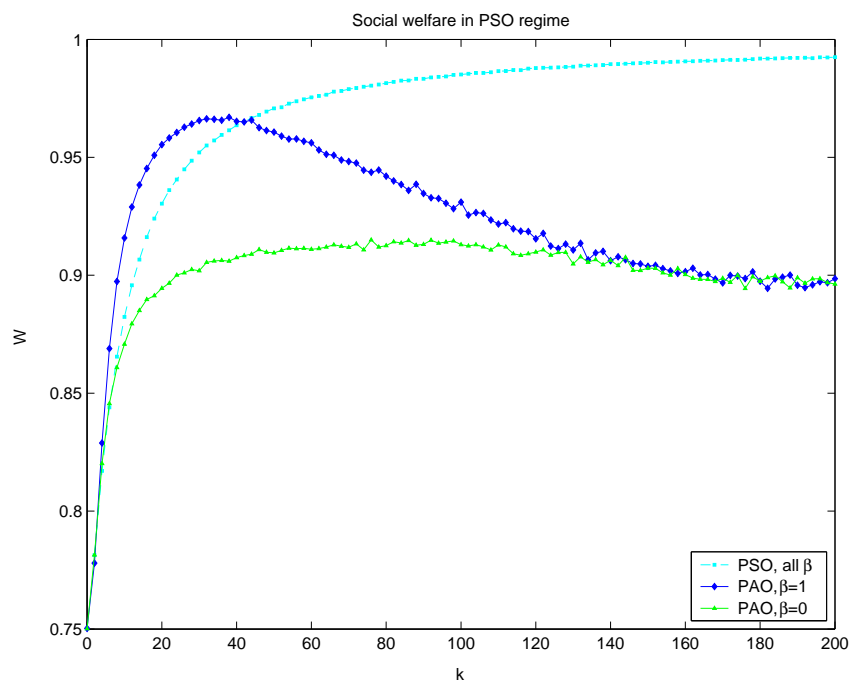


Figure 3: Social welfare as a function of network density under previous-signal-observable (PSO) regime, and under previous-action-observable (PAO) regime. Under PSO the results for networks with all β are the same.