Innovation Diffusion in Heterogeneous Populations:
Contagion, Social Influence, and Social Learning

H. Peyton Young

University of Oxford
The Brookings Institution
The Santa Fe Institute

Abstract

New ideas, products, and practices take time to diffuse, a fact that is often attributed to some form of heterogeneity among potential adopters. People may realize different benefits and costs from the innovation, have different beliefs about its benefits and costs, hear about it at different times, or delay in acting on their information. This paper analyzes the effect of incorporating heterogeneity into three broad classes of models -- contagion, social influence, and social learning. Each type of model leaves a characteristic 'footprint' on the shape of the adoption curve that amounts to a restriction on the pattern of acceleration under few (and in some cases no) restrictions on the distribution of parameters. These restrictions provide a basis for discriminating empirically between different models, and have potential application to marketing, technological change, fads, and epidemics.

1. Introduction

A basic puzzle posed by innovation diffusion is why there is often a long lag between an innovation's first appearance and the time when a substantial number of people have adopted it. There is an extensive theoretical and empirical literature on this phenomenon and the mechanisms that might give rise to it.[1] A common feature of these explanations is that *heterogeneity* among the agents is the reason that they adopt at different times. However, most of the extant models incorporate heterogeneity in a very restricted fashion, say by considering two homogeneous populations of agents, or by assuming that the heterogeneity is described by a particular family of distributions.[2]

In this paper I show how to incorporate heterogeneity into some of the benchmark models in marketing, sociology, and economics without imposing any parametric restrictions on the distribution of the underlying parameters. The resulting dynamical systems turn out to be surprisingly tractable; indeed, some of them can be solved *explicitly* for any distribution of the parameter values. I then demonstrate that each class of models leaves a distinctive 'footprint'; in particular, they exhibit noticeably different patterns of acceleration, especially in the start-up phase, with few or no assumptions on the distribution of the parameters. The reason is that the models themselves have fundamentally different structures that even large differences in the distributions cannot overcome. It follows that, given sufficient data on the aggregate dynamics of a diffusion process, one could assess the relative plausibility of different mechanisms that might be driving it with little or no prior knowledge about the

---

[1] For general overviews of the subject see Mahajan and Peterson , 1985; Mahajan, Muller, and Bass, 1990; Geroski, 2000; Stoneman, 2002; Rogers, 2003; and Valente , 1995, 1996, 2005.

[2] See among others Jeuland, 1981; Jensen, 1982; Karshenas and Stoneman, 1992; Geroski, 2000.

distribution of parameters. While this type of analysis is not an identification strategy, and is certainly no substitute for having good micro-level data, it could be useful in situations where such data are unavailable.

I shall consider three basic types of innovation diffusion models, each arising from a different account of how innovations spread.

1. *Contagion.* People adopt an innovation when they come in contact with someone who has already adopted.

2. *Social threshold*. People adopt when enough other people in the group have adopted.

3. *Social learning*. People adopt once they see enough evidence among prior adopters to convince them that the innovation is worth adopting.

For each type of model I show how to incorporate heterogeneity of the parameters in considerable generality without losing analytical tractability; moreover this can be done even there are *multiple sources* of heterogeneity. Of the three types of models, social learning is perhaps the most interesting from an economic standpoint, since it is based on the assumption that agents use payoff information from prior adopters in order to make a decision. (The other two classes are based on the notion of exposure rather than on utility maximization, though as we shall see the social threshold model can be reinterpreted in a utility maximization framework.) While there is a substantial theoretical and empirical literature on social learning models,[3] however, surprisingly little prior work has

---

[3] See among others Bikchandani, Hirshleifer, and Welsh, 1992; Banerjee, 1992; Kirman, 1993; Ellison and Fudenberg, 1993, 1995; Kapur, 1995; Bala and Goyal, 1998; Smith and Sorensen, 2000; Chatterjee and Hu, 2004; Banerjee and Fudenberg, 2004; Manski, 2004.

been done on the implications of such models for the *shape* of the adoption curve.[4] One of the main contributions of the paper is to show how to express the aggregate dynamics of such a model in an analytically tractable form even when there are multiple sources of heterogeneity among agents, including different costs of adoption, different prior information, and different amounts of 'connectedness' with the rest of the population. A limitation of the analysis is the use of a mean-field approach in which the population is assumed to be large and encounters purely random. If instead agents interact through a fixed social network, the aggregate dynamics are substantially more complex and depend on the network topology.[5] The extension of the approach to this situation will be considered in future work.

2. Inertia

Before launching into a discussion of the three main models, it will be useful to consider an even simpler reason why innovations might take time to diffuse, namely, people sometimes *delay* in acting on new information.   This hypothesis leads to a particularly tractable model that has been studied in other contexts, notably heterogeneous duration models (see among others Lancaster and Nickell, 1980; Heckman and Singer, 1982; Heckman, Robb, and Walker, 1990).

First consider the situation where there is no heterogeneity among the agents. Let $\lambda > 0$ be the instantaneous rate at which any given non-adopter first adopts. Let $p(t)$ be the proportion of adopters at time t, and let us set the clock so that $p(0) = 0$. The function $p(t)$ is called the *adoption curve*. Assume for simplicity

---

[4] Notable exceptions are Jensen (1982) and Lopez-Pintado and Watts (2006).
[5] For models of diffusion in social networks see among others Blume (1993, 1995), Bala and Goyal (1998), Young (2003), Chatterjee and Hu (2004), Cowan and Jonard (2004), Jackson and Rogers (2006) , Jackson and Yariv (2007), Golub and Jackson (2007), Vega-Redondo (2007).

that once agents have adopted, they do not disadopt within the time frame of the analysis. Then the expected motion is described by the ordinary differential equation $\dot{p}(t) = \lambda(1 - p(t))$, which has the solution $p(t) = 1 - e^{-\lambda t}$ given the initial condition $p(0) = 0$.

Notice that this curve is *concave* throughout; in particular, it is not S-shaped. It is a rather remarkable fact that this remains true when *any amount of heterogeneity* is introduced. To see this, suppose that $v(\lambda)$ is the distribution of λ in the population. Then the expected trajectory of the process is given by

$$p(t) = 1 - \int e^{-\lambda t} dv. \tag{1}$$

Differentiating (1) twice over, we see that $\ddot{p}(t) < 0$ *irrespective of the distribution* $v(\lambda)$.[6] The intuition is straightforward: agents with high values of λ tend to adopt earlier than those with low values of λ. It follows that the average value of λ in the current population of non-adopters is non-increasing, and the number of such individuals is strictly decreasing. Thus the flow of new adopters is strictly decreasing.

This simple example illustrates the kinds of results that hold in more complex situations: the structure of the model has implications for the *shape* of the curve that remain true even when an arbitrary amount of heterogeneity is introduced; indeed we can derive analogous results even with multiple sources of heterogeneity.

---

[6] Furthermore the higher-order derivatives alternate in sign; see Heckman and Singer (1982).

3. Contagion

Contagion refers to a process in which an agent adopts a new product or practice when he comes into contact with someone else who has adopted it. An everyday example would be a new fashion that spreads because people imitate those who have already adopted. The resulting dynamics are similar to those of an epidemic; indeed, some of the models are borrowed more or less directly from the epidemiology literature. In the context of innovation diffusion it is common to use a two-parameter model that allows for contagion from within the group (at one rate) and also from sources outside the group (at a possibly different rate). This is known as the *Bass model* of new product diffusion (Bass, 1969, 1980) and also as the *mixed-influence diffusion model* (Mahajan and Peterson, 1985). In the context of a new fashion in clothing, for example, these two rates would correspond to seeing other people on the street who are wearing it, and seeing ads that are promoting it.

Let us begin by describing the homogeneous version of the model, then we shall introduce heterogeneity. Let λ be the instantaneous rate at which a current non-adopter 'hears about' the innovation from a previous adopter *within* the group, and let γ be the instantaneous rate at which he 'hears about' it from sources *outside* of the group. We shall assume that λ and γ are nonnegative, and that not both are zero. In the absence of heterogeneity, such a process is described by the ordinary differential equation $\dot{p}(t) = (\lambda p(t) + \gamma)(1 - p(t))$, and the solution is

$$p(t) = [1 - \beta\gamma e^{-(\lambda+\gamma)t}]/[1 + \beta\lambda e^{-(\lambda+\gamma)t}], \ \ \beta > 0. \tag{2}$$

When contagion is generated purely from internal sources ($\gamma = 0$) this boils down to the ordinary logistic function, which is of course S-shaped.[7] When innovation is driven solely by an external source ($\gamma > 0$ and $\lambda = 0$), we obtain the pure inertia discussed earlier. When both γ and λ are positive, we can choose β in expression (2) so that $p(0) = 0$; namely, with $\beta = 1/\gamma$ we obtain

$$p(t) = [1 - e^{-(\lambda+\gamma)t}]/[1 + (\lambda/\gamma)e^{-(\lambda+\gamma)t}]. \tag{3}$$

This model has spawned many variants, some of which assume a degree of heterogeneity, such as two groups with different contagion parameters (Karshenas and Stoneman, 1992; Geroski, 2000), while others employ distributions from a specific parametric family, such as gamma distributions (Jeuland, 1981).

In fact, we can formulate a fully heterogeneous version that is analytically tractable and places virtually no restrictions on the joint distribution of the parameters. Specifically, let μ be the *joint distribution* of the contagion parameters λ and γ. Assume for analytical convenience that μ has bounded support, which we may take to be $\Omega = [0,1]^2$. (Rescaling λ and γ by a common factor is equivalent to changing the time scale, so this involves no real loss of generality.) The only substantive restriction that we place on μ is $\bar{\gamma} = \int_\Omega \gamma d\mu > 0$, for otherwise the process cannot get out of the initial state $p(0) = 0$.

---

[7] The logistic model was common in the early work on innovation diffusion; see for example Griliches (1957) and Mansfield (1961).

Let $p_{\lambda,\gamma}(t)$ be the *proportion* of all type-(λ,γ) individuals who have adopted by time t. Then the proportion of all individuals who have adopted by time t is

$$p(t) = \int p_{\lambda,\gamma}(t)d\mu.\tag{4}$$

(Hereafter integration over Ω is understood.) Each subpopulation of adopters $p_{\lambda,\gamma}(t)$ evolves according to the differential equation

$$\dot{p}_{\lambda,\gamma}(t) = (\lambda p(t)+\gamma)(1-p_{\lambda,\gamma}(t)).\tag{5}$$

This defines an infinite system of first-order differential equations coupled through the common term $p(t)$. We can reduce it to a single differential equation by the following device: let $x_{\lambda,\gamma}(t) = \ln(1-p_{\lambda,\gamma}(t))$ and observe that (5) is equivalent to the system $\dot{x}_{\lambda,\gamma}(t) = -(\lambda p(t)+\gamma)$ for all (λ,γ). From this and the initial condition $x_{\lambda,\gamma}(0)=0$ we obtain

$$x_{\lambda,\gamma}(t) = -\int_0^t (\lambda p(s)+\gamma)ds = -\lambda\int_0^t p(s)ds - \gamma t.\tag{6}$$

From the definition of $x_{\lambda,\gamma}(t)$ it follows that

$$p(t) = 1 - \int e^{x_{\lambda,\gamma}(t)}d\mu,\tag{7}$$

that is, $p(t)$ satisfies the integral equation

$$p(t) = 1 - \int e^{-\gamma t-\lambda\int_0^t p(s)ds}d\mu.\tag{8}$$

Differentiating we obtain

9

$$\dot{p}(t) = \int (\lambda p(t) + \gamma) e^{-\gamma t - \lambda \int_0^t p(s)ds} d\mu. \tag{9}$$

Expression (9) can be put in more standard form by defining $y(t) = \int_0^t p(s)ds$. Then $\dot{y}(t) = p(t)$, $\ddot{y}(t) = \dot{p}(t)$, and (9) becomes a second-order differential equation in y, namely, $\ddot{y}(t) = \int (\lambda \dot{y}(t) + \gamma) e^{-\gamma t - \lambda y(t)} d\mu$. The right-hand side is Lipschitz continuous in $t, y,$ and $\dot{y}$, hence on any finite interval $0 \le t \le T$ there exists a unique continuous solution satisfying the initial condition $y(0) = p(0) = 0$. By the Picard-Lindelöf theorem, such a solution can be constructed by successive approximation (Coddington and Levinson, 1955).

It turns out that we can deduce some key dynamic properties of the process without solving it explicitly however. In particular, we claim that $\dot{p}(t) / [p(t)(1 - p(t))]$ is *strictly decreasing*, that is, the hazard rate $\dot{p}(t) / (1 - p(t))$ decreases relative to the adoption level $p(t)$ *irrespective of the joint distribution of* λ *and* μ.

Proposition 1. *Suppose that diffusion is driven by heterogeneous contagion with joint distribution* μ *on the parameters* $(\lambda, \gamma) \in [0,1]^2$ *such that the unconditional mean* $\bar{\gamma}$ *is positive. Then for all* $t > 0$, $\dot{p}(t) / [p(t)(1 - p(t))]$ *is strictly decreasing in* t, *which is equivalent to*

$$\frac{\ddot{p}(t)}{\dot{p}(t)} < \frac{(1 - 2p(t))\dot{p}(t)}{p(t)(1 - p(t))} \text{ for all } t > 0. \tag{10}$$

The latter condition implies that the relative rate of acceleration has an upper bound that goes to zero as $p(t) \to 1/2$; in particular the process cannot accelerate at all beyond $p = 1/2$. As we shall see in subsequent sections, this stands in

contrast to social threshold and social learning models, where $\dot{p}(t)/[p(t)(1-p(t))]$ can be strictly increasing and acceleration can continue well beyond $1/2$. The proof of Proposition 1 is given in the Appendix.

Note that Proposition 1 is a statement about the *curvature* of the adoption function, and is considerably more subtle than saying that it is S-shaped. Indeed there are perfectly reasonable S-shaped curves that are inconsistent with this criterion. Consider a curve of the form $\dot{p}(t) = p^a(t)(1-p(t))$, which was first proposed as a model of innovation diffusion by Easingwood, Mahajan, and Muller (1981, 1983). When $a > 1$, $\dot{p}(t)/[p(t)(1-p(t)] = p^{a-1}(t)$ is strictly increasing, hence Proposition 1 shows that such a process cannot arise from a contagion model with any amount of heterogeneity. Nevertheless, it is possible to generate an S-shaped curve from a contagion model -- in fact from a homogeneous contagion model – whose overall appearance is very similar (see Figure 1). The differences between the two models are only revealed by studying the behavior of the modified hazard rate $\dot{p}(t)/[p(t)(1-p(t))]$.
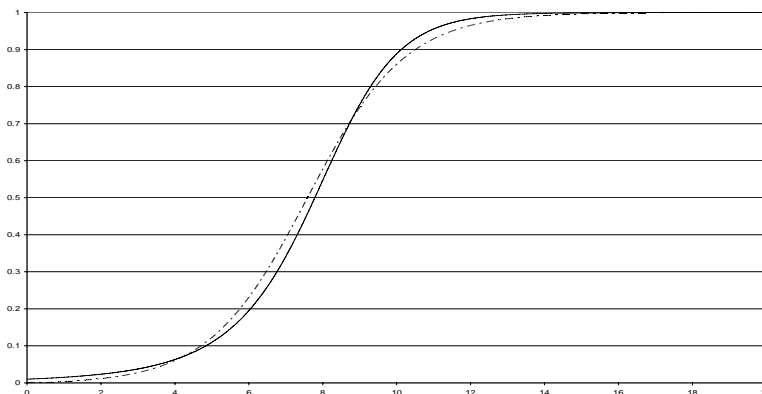

Figure 1. Two adoption curves: the solid line is generated by $\dot{p}(t) = p^{1.2}(1-p)$ and $p(0) = 0.01$, the dashed line by a Bass model with $\lambda = .75$ and $\gamma = .0025$.

4. Social thresholds

The sociological literature on innovation stresses the idea that people have different 'thresholds' that determine when they will adopt as a function of the number (or proportion) of others in the population who have adopted. The dynamics of these models were first studied by Schelling (1971, 1978), Granovetter (1978), and Granovetter and Soong (1988); for more recent work in this vein see Macy (1991), Valente (1995, 1996, 2005), Dodds and Watts (2004, 2005), Centola (2006), and Lopez-Pintado and Watts (2006).

For each agent i, suppose that there exists a minimum proportion $r_i \geq 0$ such that i adopts as soon as r_i or more of the group has adopted. (If $r_i > 1$ the agent never adopts.) This is called the *social threshold* of agent i. The precise meaning of these thresholds varies from one context to another; broadly speaking we can think of them as representing different *degrees of responsiveness to social influence*. A concrete example would be the transmission of rumors: some people would need to hear the rumor from many people to pass it on while others might only need to hear it once. A key feature of the model is that the adoption depends on the innovation's current *popularity* rather than on how good or desirable the innovation has proven to be. The latter is the basis of *social learning models*, which we shall take up in the next section.

We wish to model the mean-field dynamics of heterogeneous social threshold models without assuming a parametric form for the distribution of thresholds. To this end, let $F(r)$ be the cumulative distribution function of thresholds $r \geq 0$ in some given population. Following Granovetter (1978), we can then define the discrete-time version of the adoption process as follows. Let $p(t)$ be the proportion of adopters in period $t = 0, 1, 2, \ldots$. The clock starts in period 0 when no one has yet adopted: $p(0) = 0$. In period 1, everyone adopts whose thresholds are zero, that is, $p(1) = F(0)$. These are the *innovators*. We shall assume that

$F(0) > 0$, for otherwise the process cannot get started. In period 2, everyone adopts whose thresholds are at most $F(0)$. Thus at the end of the second period the fraction $p(2) = F(F(0))$ have adopted. Proceeding in this way, we obtain $p(t) = F^{[t]}(0)$, where $F^{[t]}$ is the t-fold composition of F with itself.

A useful generalization is to allow for some *inertia* in the adoption decision. Specifically, let us assume that in each period only a fraction $\alpha \in (0,1)$ of those who are prepared to adopt actually do so. In other words, among those people whose thresholds have been crossed but who have not yet adopted by the end of period t, only α will adopt by the end of the next period.[8] This yields the discrete-time process

$$p(t+1) - p(t) = \alpha[F(p(t)) - p(t)]. \tag{11}$$

The continuous-time analog is

$$\dot{p}(t) = \lambda[F(p(t)) - p(t)], \lambda > 0 .[9] \tag{12}$$

Assume now that $F(0) > 0$ and let b be the *first fixed point*, that is, the smallest number in (0, 1] such that $F(b) = b$. (Such a point always exists.) We then have $F(r) > r$ for all $r \in [0, b)$. Since (12) is a separable ordinary differential equation, we obtain the following explicit solution for the inverse function $t = p^{-1}(x)$:

$$\forall x \in [0, b), \quad t = p^{-1}(x) = (1/\lambda) \int_0^x dr /(F(r) - r) .[10] \tag{13}$$

---

[8] A stochastic version of the model would represent this as a discrete-time process such that each agent whose threshold has been crossed converts with some given probability in each period. Here, and throughout the paper, we shall focus solely on the expected dynamics in a large population setting.

[9] Independently, Lopez-Pintado and Watts (2006) derive the same continuous-time generalization and study its fixed points under various assumptions about F.

Observe that the right-hand side is integrable because $F(r)$ is monotone nondecreasing and $F(r)-r$ is bounded away from zero for all r in the interval $[0, x]$ whenever $x < b$. (The constant of integration is zero because of the initial condition $p(0) = 0$.) As $x \to b$, the right-hand side of (13) goes to infinity, which implies that the adoption curve approaches b asymptotically; in particular, *the adoption process peters out at the first fixed point of F.*

This phenomenon is illustrated in figure 2. Here we assume that the thresholds are normally distributed to the right of the origin, and there is a point mass at the origin corresponding to the subset of innovators -- the people who are willing to adopt even when no one else adopted. Notice that the adoption curve asymptotes to $p = 0.50$, which is the first fixed point of F. There is nothing special about the normal distribution in this regard; similar results hold for any c.d.f. where it first crosses the 45° - line.

In this example the adoption curve is concave, but this is by no means necessary or even typical for this family of models. Figure 3 illustrates an entirely different adoption curve that is generated by a truncated normal with smaller mean and variance. An interesting feature of this curve is that it accelerates very sharply in the early stages; indeed it can be shown that it grows at a faster-than-exponential rate up to $p = 0.10$. We shall now show that whenever the process starts in the left tail of the distribution, these are the only two possible shapes: the curve either decelerates initially, or it accelerates initially at a super-exponential rate.
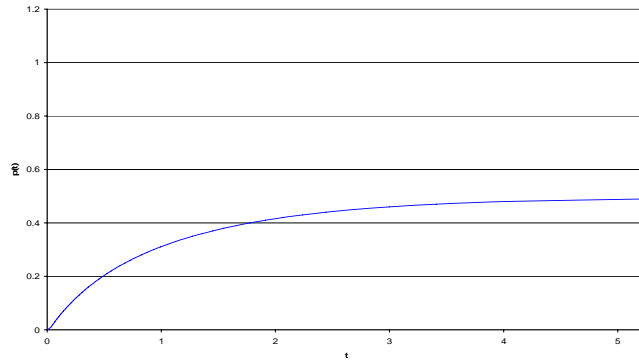
---

[10] The fact that this kind of process has an *explicit* analytic solution for any distribution seems not to have been recognized before.
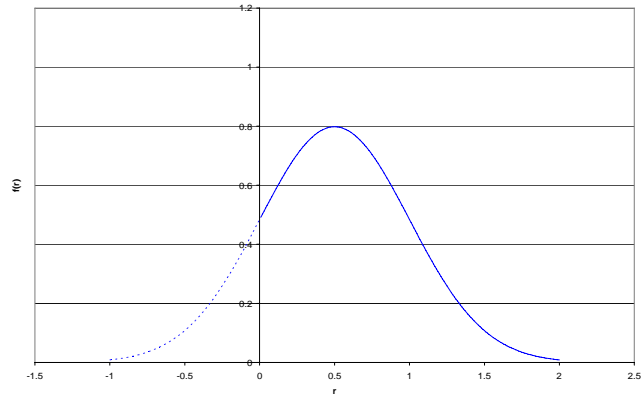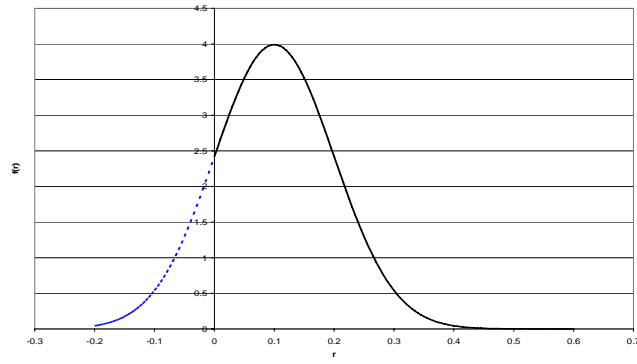
a) truncated normal density of thresholds
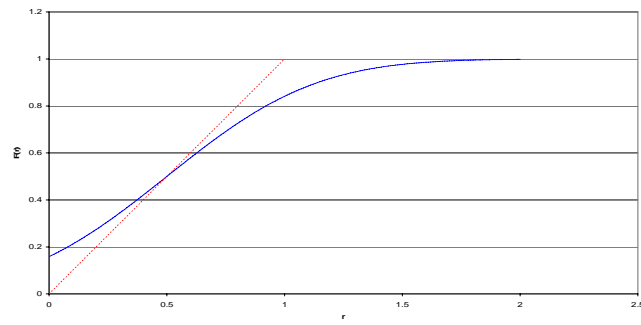
b) c.d.f. of thresholds and 45° - line

c) social threshold adoption curve $p(t)$

Figure 2. Density, c.d.f., and social threshold adoption curve generated by N(.50, .25) and $\lambda = 4$.

a) truncated normal density of thresholds

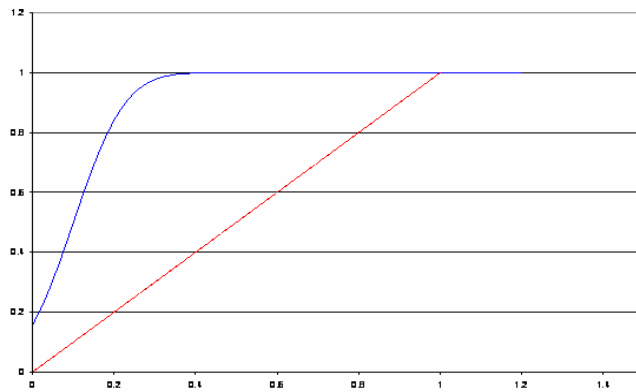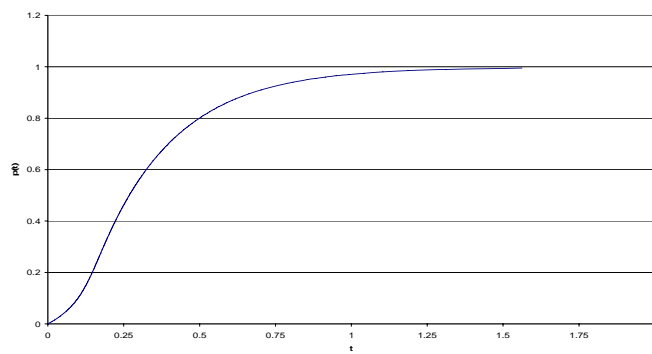b) c.d.f. of thresholds and 45° - line

c) social threshold adoption curve $p(t)$

Figure 3. Density, c.d.f., and social threshold adoption curve generated by N(.10, .01) and $\lambda = 4$.

To see why this is so, consider the basic dynamic equation in (12). Assume that $F(0) > 0$ and that $F(r)$ has a continuous density $f(r)$ defined for all $r > 0$. (Note that the density is not defined at the origin because there is a point mass there.) Differentiating (12) with respect to t and dividing through by $\dot{p}(t)$, which by assumption is positive, we obtain

$$\ddot{p}(t) / \dot{p}(t) = \lambda[f(p(t)) - 1]. \tag{14}$$

In other words, the relative acceleration rate traces out a positive linear transformation of the underlying density. It follows that the process accelerates initially if and only if the initial density is large enough, that is, $f(r) > 1$ in a neighborhood of the origin. Suppose further that $f(r)$ is *increasing* in a neighborhood of the origin, that is, the process starts in the left tail of the distribution of thresholds. Then (14) shows that the relative acceleration rate $\ddot{p}(t) / \dot{p}(t)$ is also strictly increasing, which means that the adoption curve exhibits *super-exponential growth*.[11] This phenomenon results from the compounding of two effects. First, as more and more people adopt, the amount of information available to the remainder of the population increases. Second, the number of people persuaded by each additional bit of information increases as the process moves up the left tail of the distribution.

These conclusions continue to hold when λ is heterogeneously distributed according to some distribution function $v(\lambda)$. Since λ is a scaling parameter, there is no real loss of generality in assuming that the support of v lies in (0, 1]. Let $F_{\lambda}(r)$ be the cumulative distribution of thresholds conditional on λ, and

---

[11] Growth is *exponential* if $\ddot{p}(t) / \dot{p}(t)$ is constant, *super-exponential* if $\ddot{p}(t) / \dot{p}(t)$ is strictly increasing, and *sub-exponential* if $\ddot{p}(t) / \dot{p}(t)$ is strictly decreasing.

assume that the conditional density $f_\lambda(r)$ exists for every $r > 0$. (We shall assume that $F_\lambda(0) > 0$, so $f_\lambda(0)$ is undefined.)

Proposition 2. *Suppose that diffusion is driven by a heterogeneous social threshold model such that, for each level of inertia $\lambda \in (0,1]$, the conditional distribution of thresholds satisfies $F_\lambda(0) > 0$, has a continuous density $f_\lambda(r)$ for all $r > 0$, and the densities $f_\lambda(r)$ are strictly increasing in a common open neighborhood of the origin. Then initially the relative acceleration rate $\ddot{p}(t)/\dot{p}(t)$ is strictly increasing, which implies that the process either decelerates or accelerates at a super-exponential rate.*

Proof.   The equations of motion are

$$\dot{p}_\lambda(t) = \lambda[F_\lambda(p(t)) - p_\lambda(t)], \tag{15}$$

where the initial conditions are $p_\lambda(0) = 0$.   Since $F_\lambda(0) > 0$, $\dot{p}_\lambda(t) > 0$ for all sufficiently small $t > 0$ and

$$\ddot{p}_\lambda(t)/\dot{p}_\lambda(t) = \lambda[f_\lambda(p(t)) - 1] . \tag{16}$$

By hypothesis the functions $f_\lambda(p(t))$ are strictly increasing on some common interval $0 < p(t) < \bar{p}$.   It follows that, for every λ and all t in a suitable interval $(0,T]$,

$$\dddot{p}_\lambda(t)/\dot{p}_\lambda(t) - (\ddot{p}_\lambda(t)/\dot{p}_\lambda(t))^2 > 0, \tag{17}$$

that is,

$$\sqrt{\dddot{p}_\lambda(t)\dot{p}_\lambda(t)} > \ddot{p}_\lambda(t). \tag{18}$$

Hence

$$\int \sqrt{\ddot{p}_\lambda(t)\dot{p}_\lambda(t)}\,dv > \int \ddot{p}_\lambda(t)\,dv = \ddot{p}(t). \qquad (19)$$

By Schwarz's inequality,

$$\sqrt{\ddot{p}(t)\dot{p}(t)} = (\int \ddot{p}_\lambda(t)\,dv \int \dot{p}_\lambda(t)\,dv)^{1/2} \geq \int \sqrt{\ddot{p}_\lambda(t)\dot{p}_\lambda(t)}\,dv. \qquad (20)$$

Combining this with (19) we conclude that $\sqrt{\ddot{p}(t)\dot{p}(t)} > \ddot{p}(t)$, which implies that $\dfrac{d[\ddot{p}(t)/\dot{p}(t)]}{dt} > 0$, that is, $\ddot{p}(t)/\dot{p}(t)$ is strictly increasing on $(0,T]$. This concludes the proof of Proposition 2.

5. Social learning

A difficulty with both of the preceding models is that they provide no clear reason *why* an agent would adopt an innovation given that others have adopted it. In this section we consider a class of models in which the adoption decision flows directly from expected utility maximization. Specifically, an agent adopts if he has reason to believe the innovation is better than what he is doing now, where the evidence comes from directly observing the outcomes among prior adopters. For example, when a new product becomes available -- e.g., a new medication (aspirin), communication technology (cellphone), or agricultural practice (no till) – many people will want to see how it works for others *over a period of time* before trying it themselves. These are variously known as *social learning models* or *social learning models based on direct observation*.

There is a sizable theoretical literature on social learning, but it is difficult to summarize due to the great diversity in behavioral and informational assumptions that different authors use. Some assume that payoff outcomes among prior adopters are fully observable, while others assume only that the act

of adoption is observable (the latter are usually called herding models). Some assume that agents can recognize others' types, while others assume that types cannot be identified. There are also significant differences in the relevant *characteristics* that authors choose to focus on, including heterogeneity in risk aversion, discount rates, and amount of information.[12] There is considerable empirical evidence, however, that learning from the experience of others does in fact occur.[13]

Here I shall make a number of simplifying assumptions in order to get a handle on an issue that has not received much previous attention in this literature, namely, *what do the short-run aggregate dynamics of a heterogeneous learning model look like, and do they differ qualitatively from the dynamics generated by other classes of models?* To make some progress on this question let us make the following assumptions: i) payoffs are observable; ii) agents are risk-neutral and myopic; iii) there is no idiosyncratic component to payoffs due to differences in agents' types, but agents may have different costs (not necessarily observable); iv) there are differences in agents' prior beliefs about how good the innovation is relative to the status quo; and v) there are differences in the number of people they observe and hence in the amount of information they have. Many other complicating factors could be introduced, such as discounting, one-time switching costs, risk aversion, and imperfect observability (to name but a few), but these would distract from the main point, which is to identify the dynamic characteristics of a fairly general class of learning models without attempting to formulate the most general such model.

---

[12] See among others Bikchandani, Hirshleifer, and Welsh, 1992; Banerjee, 1992; Kirman, 1993; Ellison and Fudenberg, 1993, 1995; Kapur, 1995; Bala and Goyal, 1998; Smith and Sorensen, 2000; Chatterjee and Hu, 2004; Banerjee and Fudenberg, 2004; Manski, 2004.
[13] Foster and Rosenzweig (1995), Munshi (2004), Conley and Udry (2005).

I claim that, under the above assumptions, the dynamical system has a surprisingly simple structure. In particular, one can reduce the various types of heterogeneity to a *composite index* that measures the *probability of a given agent adopting conditional on the amount of information that has been generated so far*. The equation of motion of such a process turns out to be analogous to the one obtained for the social threshold model, except that here the relevant state variable at time t turns out to be the *integral of the adoption curve* up through t, $\int_0^t p(s)ds$, rather than the *level of adoption* $p(t)$. The reason is that $\int_0^t p(s)ds$ measures the cumulative information generated by all prior adopters *from the time they first adopted*, which is the relevant variable in the learning context.

To illustrate how such a model works, let us walk through a particular example using a standard normal-normal updating framework. This is chosen mainly for its computational transparency; similar results hold under alternative assumptions.[14] Consider a population of n individuals, where n is large, and let us maintain the five assumptions mentioned above. Let us also assume that the payoff from the innovation is a normally distributed random variable X with mean $\mu > 0$ and variance σ^2, which is i.i.d. among agents and time periods. We shall interpret μ as the *mean payoff gain per period from using the innovation as compared to the status quo technology*. Each agent is assumed to have an idiosyncratic variable cost c_i of using the innovation, so he adopts if and only if he believes that the mean payoff per period is at least c_i.

If everyone knew the true value of μ from the outset, then everyone would adopt for whom $c_i < \mu$. This is the *efficient outcome*. Ex ante, however, people do

---

[14] Jensen (1982) and Lopez-Pintado and Watts (2006) study the case where the outcome variable is binomial (payoffs are "high" or "low"). However, they assume that agents pay attention only to current outcomes, not the cumulative amount of information generated from earlier periods.

not know the true value of μ; furthermore, they may start with substantially different beliefs (based on their prior private information) about what the true value is. As more information comes in, they update their beliefs. If the information is sufficiently favorable, more people will adopt, which creates a still-larger base of information, which causes even more people to adopt, and so on. This is the essential logic driving the learning dynamics.

To continue with our example, suppose that each agent i has a prior belief about the unknown mean μ and unknown precision $\rho = 1/\sigma^2$ such that: i) the marginal of ρ is gamma-distributed, and ii) for each value of ρ the conditional distribution of μ is normal with mean μ_{i0} and precision $\rho\tau_i$. (This is a standard normal-normal updating model; see for example De Groot (1970).) Low values of τ_i reflect *flexibility* in beliefs, that is, relatively little evidence is needed to shift $i's$ belief about the mean by a given amount. Low values of μ_{i0} reflect *pessimism* about the payoffs from the innovation. In particular, if i initially believes that the mean is less than his costs, $\mu_{i0} < c_i$, he will not want to adopt. As more information comes in, however, his posterior estimate of the mean, μ_{it}, may increase sufficiently that he changes his mind. The point at which this happens depends (among other things) on how much information i collects and how flexible his beliefs are.

For the moment we shall assume a discrete-time process, then consider the continuous-time analog. Let X_{it} be the payoff realization to adopter i in time period t. By the end of period t, a total of N_t independent realizations of X will have been generated in the population, where

$$N_t = n[(t-1)p(1) + (t-2)[p(2) - p(1)] + ... + [p(t) - p(t-1)]] = n\sum_{s=1}^{t} p(s). \quad (21)$$

Let $r(t) = \sum_{s=1}^{t} p(s)$. A given agent can only be expected to know a few of these outcomes given the random nature of information flows and the fact that each agent has a limited number of social contacts. To model the information transmission process, assume that, in each period s, agent i hears about some of the payoff outcomes according to a Poisson arrival process with mean $\beta_i p(s)$. Thus $\beta_i p(s)$ is the expected number of realizations of X that i hears about in period s, and $E[n_{it}] = \beta_i r(t)$ is the expected number that i has heard about through period t. The parameter β_i is a measure of $i's$ *information*, or the extent to which i "gets around."[15]

Let $\bar{x}_{it}$ denote the realized mean among the n_{it} observations of agent i at time t. Given our assumptions, $\bar{x}_{it}$ is normal with mean μ and standard devation $\sigma / \sqrt{n_{it}}$. In the present framework, $i's$ Bayesian posterior estimate of the mean, μ_{it}, can be expressed very simply as a convex combination of μ_{i0} and $\bar{x}_{it}$, namely,

$$\mu_{it} = \frac{n_{it} \bar{x}_{it} + \tau_i \mu_{i0}}{n_{it} + \tau_i} \ . \tag{22}$$

In other words, the posterior estimate is just a weighted average of the prior and the observed mean, where the weight on the mean is the number of independent observations that produced it.

---

[15] If agents were embedded in a fixed social network, the analogous parameter would be the number of other agents with whom a given agent is connected. In this case, however, the aggregate amount of information $r(t)$ will generally not be sufficient to describe the state of the system; the dynamics of the process will depend on the specific network topology.

Since i is myopic, she is prepared to adopt once μ_{it} is at least equal to her discounted switching cost c_i, which by (22) is equivalent to

$$(\bar{x}_{it} - c_i)n_{it} \geq \tau_i(c_i - \mu_{i0}) \ . \tag{23}$$

This expression becomes more transparent when we focus on the subpopulation of agents for whom adoption is worthwhile: $P^0 = \{i : \mu > c_i\}$. In the interest of computational transparency let us also (temporarily) substitute the expected value $E[n_{it}] = \beta_i r(t)$ into (23). After rearranging terms we obtain

$$r(t) \geq \frac{\tau_i(c_i - \mu_{i0})}{\beta_i(\mu - c_i)} - \frac{\sigma\sqrt{r(t)}z_{it}}{(\mu - c_i)\sqrt{\beta_i}} \ , \tag{24}$$

where z_{it} is $N(0,1)$. Define i's *resistance level* (or *information threshold*) to be the expected value of the right-hand side of this inequality, namely,

$$r_i = \frac{\tau_i(c_i - \mu_{i0})}{\beta_i(\mu - c_i)} \ . \tag{25}$$

From (24) and (25) we conclude that an agent with characteristics $(c_i, \tau_i, \beta_i, \mu_{i0})$ is increasingly likely to adopt as $r(t)$ passes the threshold r_i. Moreover, this threshold has a natural interpretation: agents with high r_i are those who are initially *pessimistic* that the innovation will cover their costs ($c_i - \mu_{i0}$ is large), *inflexible* in their initial beliefs (τ_i is large), *marginally profitable* ($\mu - c_i$ is low), and *relatively uninformed* (β_i is small).

The precise form of expressions (24) and (25) is not particularly important for our purpose however: the key point is that each agent i has a *response function $\phi_i(r)$, which represents the probability that i believes the innovation is worth adopting, given that the total amount of information generated by the prior adopters equals r.*[16]

Let us now abstract from this particular example and take the notion of a response function as a primitive of the model. Suppose that each agent in the population has a "type" i that is characterized by a response function $\phi_i : R_+ \to [0,1]$, where $\phi_i(r)$ is the probability that $i's$ information threshold has been crossed when the total amount of information generated by the prior adopters is r. For ease of interpretation we shall assume that the functions $\phi_i(r)$ are monotone nondecreasing, though this is not actually necessary for some of the results to follow. Notice that a given individual will typically know only a small fraction of the prior outcomes, that is, r is a state variable that represents a common pool of information but it is not common knowledge.

Let p_i be the *proportion* of i-types in the population, which we shall assume is infinitely large. When the total information generated by prior adopters equals r, the proportion of the population whose thresholds have been crossed is given by the function

$$F(r) = \sum_i p_i \phi_i(r) \ . \tag{26}$$

---

[16] In the present case, the response function has the following explicit representation. Given that $r(t) = r$, n_{it} is Poisson-distributed with mean $\beta_i r$. Given a realization $n_{it} = k > 0$, the mean observed payoff, $\overline{x}_{it}$, is normal with mean μ and variance $k\sigma^2$. Let Φ denote the standard normal c.d.f. Then the probability that $i's$ posterior estimate exceeds $i's$ costs, as given by expression (23), is

$$\phi_i(r) = \sum_{k=1}^{\infty} \frac{(\beta_i r)^k e^{-\beta_i r}}{k!} \Phi[(\frac{(\mu - c_i)\sqrt{k}}{\sigma} - \frac{\tau_i(c_i - \mu_{i0})}{\sigma\sqrt{k}}].$$

$F(r)$ is a monotone nondecreasing function which we can interpret as a *notional distribution function* of agents' information thresholds. We allow for the possibility that some agents have an infinite threshold, hence $\lim_{r \to \infty} F(r)$ may be less than 1. [17]

In a discrete-time framework, the total information generated by the end of period t is $r = r(t) = \sum_{s=1}^{t} p(s)$. If agents adopt as soon as their information thresholds have been crossed, the *discrete-time dynamics* is described by the equation $p(t+1) = F(\sum_{s=1}^{t} p(s))$. More generally, when agents act with probabilistic delay $\alpha \in (0,1]$, we have

$$p(t+1) - p(t) = \alpha[F(\sum_{s=1}^{t} p(s)) - p(t))],\qquad(27)$$

In a continuous-time setting the cumulative information generated by time t is $r(t) = \int_{0}^{t} p(s)ds$, and the aggregate dynamics is described by the differential equation

$$\dot{p}(t) = \lambda[F(\int_{0}^{t} p(s)ds) - p(t)], \lambda > 0.\qquad(28)$$

---

[17] The closest model to this one in the literature is due to Dodds and Watts (2004, 2005). They consider a cumulative-dose model of infection with heterogeneity in the thresholds at which agents become infected (including social as well as biological interpretations of "infection"), though their analysis of the dynamics emphasizes different features from the ones considered here.

When λ and r are jointly distributed, the dynamics are described by the system of differential equations

$$\forall \lambda \in [0,1], \quad \dot{p}_\lambda(t) = \lambda[F_\lambda(r(t)) - p_\lambda(t)], \text{ where } r(t) = \int_0^t \int_0^1 p_\lambda(s) dv ds, \qquad (29)$$

where $v(\lambda)$ is the distribution of λ, whose support is assumed to lie in the interval $(0,1]$. Given the initial condition $p(0) = 0$, for any finite time $T \geq 0$ there exists a unique continuous solution on $t \in [0,T]$ provided that, for each value of λ, $F_\lambda(0) > 0$ and $F_\lambda(r)$ is Lipschitz continuous.

Observe that (28) is similar to the dynamical equation (12) defining a social threshold model, except that in the present case the argument of F is the *integral of the adoption curve* rather than the adoption curve itself. This arises because agents use all past information generated by previous adopters rather than just the most recent information.[18]

The cumulative feature of the social learning model has some important implications for the shape of the adoption curve. Indeed such a model *always decelerates initially irrespective of the distribution generating it.* First we shall show why this is so assuming a homogeneous inertia rate λ; it will then be clear how to generalize the argument to the case of heterogeneous λ's. Fix some $\lambda \in (0,1)$

---

[18] This suggests the following generalization. If agents discount past information at some rate $\delta \geq 0$, then the dynamical equation takes the form $\dot{p}(t) = \lambda[F(\int_0^t e^{-\delta(t-s)} p(s) ds) - p(t)]$. When $\delta = 0$ we obtain the undiscounted learning model discussed in the text, whereas when δ is large the process is similar to the social thresholds model. The specific context will determine which of these seems most appropriate.

and let $F(r) = F_\lambda(r)$ be the distribution function of the resistance levels. Assume that $F(r)$ has a differentiable density $f(r)$ such that $f'(r)$ is *continuous and bounded* for $0 < r \leq 1$. Differentiating the defining equation (28) with respect to t, and recalling that $p(t) = \dot{r}(t)$, we obtain

$$(1/\lambda)\ddot{p}(t) = p(t)f(r(t)) - \dot{p}(t) . \qquad (30)$$

The solution $p(t)$ is continuous and $\lim_{t \to 0^+} p(t) = p(0)$. Assume that initially $p(0) = 0$. (In fact the results below continue to hold if $p(0)$ is positive and sufficiently small.) Then $f(r(t))p(t)$ is close to zero when t is close to zero. We also know from (28) that $\dot{p}(0) = \lambda F(0) > 0$. It follows from this and (30) that

$$\lim_{t \to 0^+} \ddot{p}(t) = -\lambda^2 F(0) + f(p(0))p(0) < 0 . \qquad (31)$$

Figure 4 illustrates this phenomenon for the same two densities that were used to generate the social thresholds adoption curve in Figures 2-3. The logic is that the initial block of optimists $F(0)$ exerts a decelerative drag on the process: they contribute at a decreasing rate as their numbers diminish, while the information generated by the new adopters gathers steam slowly because there are so few of them to begin with. These arguments continue to hold when there is heterogeneity in λ, as the reader may verify.[19]

---

[19]Initial deceleration does not necessarily occur, however, if innovations are *bunched* at particular points in time. An example would be an agricultural innovation (e.g., a new type of crop) that is tried once in each growing season, and whose outcomes farmers observe at the end of the season. For example, Griliches's classic study of the diffusion of hybrid corn found very strong acceleration in the early phases of adoption; this is consistent with a learning model with 'bunched' observations (Griliches, 1957; Young, 2005).
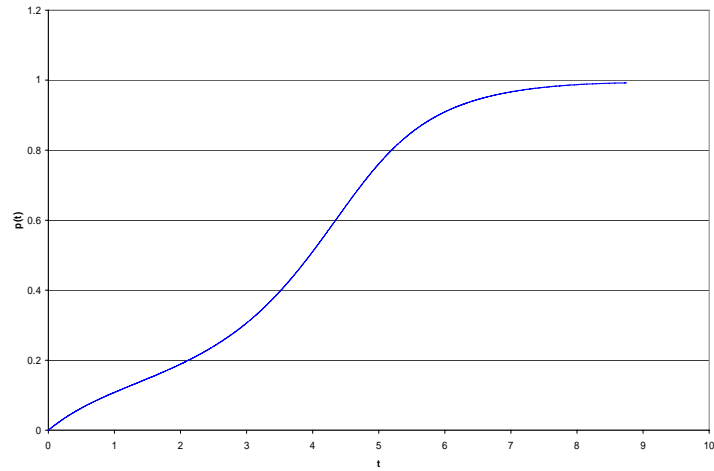
a) curve generated by N(.5, .25) and $\lambda = 4$.
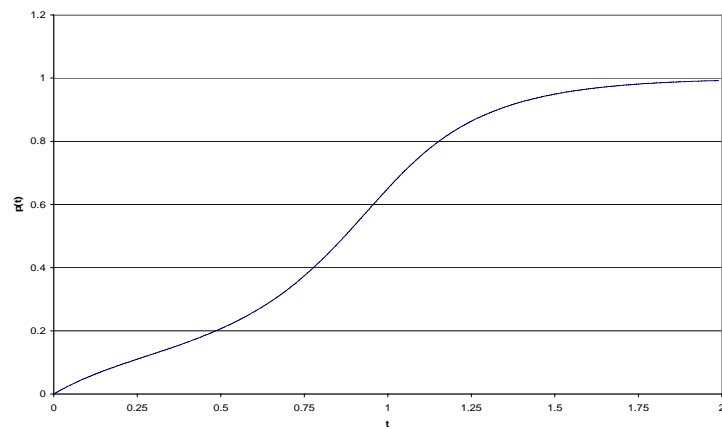
b) curve generated by N(.10, .01) and $\lambda = 4$.

Figure 4. Adoption curves generated by social learning and a normal distribution of information thresholds.

Next it will be shown that the *relative acceleration rate is strictly increasing in a neighborhood of the origin* provided that $\overline{f}(0) = \lim_{r \to 0^+} f(r) > 0$. Let

$$\phi(t) = (1/\lambda)\ddot{p}(t)/\dot{p}(t). \tag{32}$$

From (30) we deduce that

$$\phi(t) = f(r(t))p(t)/\dot{p}(t) - 1. \tag{33}$$

Differentiating (33) we obtain

$$\dot{\phi}(t) = f'(r(t))p^2(t)/\dot{p}(t) + f(r(t)) - f(r(t))\ddot{p}(t)p(t)/\dot{p}^2(t). \tag{34}$$

As $t \to 0^+$ the first term in (34) goes to zero, because by assumption f' is bounded, $p(t) \to 0$, and $\dot{p}(0) > 0$. The third term also goes to zero. However, $f(r(t)) \to \overline{f}(0) > 0$, so the second term is positive in the limit. (Recall that the density is not defined at the origin.) It follows from continuity that $\dot{\phi}(t)$ is strictly positive on some initial interval $0 \le t \le T$. In the region near the origin where $\ddot{p}(t) < 0$, this says that the relative acceleration rate is becoming less negative. Suppose that at some time t_0 the process begins to accelerate. Inspection of (34) shows that if $f'(r(t_0)) > 0$, then $\dot{\phi}(t)$ is positive for some interval of time after t_0, that is, the process undergoes a period of super-exponential growth after it begins to accelerate. The reader may verify that similar arguments hold in the heterogeneous case.

Proposition 3. *Suppose that diffusion is driven by social learning, where for each level of inertia $\lambda \in (0,1]$ the conditional distribution of resistances satisfies $F_\lambda(0) > 0$, $\overline{f}_\lambda(0) = \lim_{r \to 0^+} f_\lambda(r) > 0$, and the derivatives $f'_\lambda(r)$ are continuous and uniformly bounded above for all λ in a neighborhood of $t = 0$. Then initially the process strictly decelerates whereas the relative acceleration rate strictly increases; moreover if the densities are strictly increasing when the process begins to accelerate (if it does so at all), then the process exhibits super-exponential growth.*

6. Summary

In this paper I have studied several models of innovation diffusion, and shown how to characterize their dynamical behavior with few (in some cases no) restrictions on the distribution of agents' characteristics. Below I summarize some of the points that follow from the preceding discussion and Propositions 1-3.

1. Acceleration over any part of the trajectory is inconsistent with a pure inertia model.

2. An accelerative phase, possibly at super-exponential rates, can easily occur in social threshold and social learning models; however, acceleration at the very start of the process is inconsistent with the class of social learning models considered here.

3. Acceleration in a contagion model cannot occur beyond the 50% penetration level, and the relative acceleration rate cannot be too large before that level, but these restrictions do not apply to social threshold or social learning models.

While these features are certainly not sufficient to *identify* one family of models to the exclusion of all others, they do provide a way of assessing the relative plausibility of different types of explanation. This could be useful in situations where aggregate adoption curves are available but micro-level adoption data are not.

I hasten to point out that the trio of models discussed here does not cover all of the numerous and varied models in the literature. One important family that has not yet been mentioned is the class of moving equilibrium models (David, 1966, 1969, 1975, 2003; David and Olsen, 1984, 1986; Stoneman, 2002). These proceed on the assumption that adoption is driven by changes in some exogenous variable, such as price: if agents have different net benefits from adopting, for example, then as the price falls more and more agents will adopt. More precisely, suppose that each agent i adopts once the price is less than i's *reservation value* v_i. Let $F(v)$ be the distribution of reservation values in the population, and suppose that prices decline according to some function $\pi(t)$. Then the adoption curve is given by $p(t) = 1 - F(\pi(t))$. In this case, $p(t)$ is simply the composition of two monotone functions, so not much can be said about the shape of the curve without knowing more about the distribution F and the driving function $\pi(t)$. The models that we have considered in this paper are fundamentally different because the dynamics are driven from within. It is this feature that places nontrivial restrictions on the shape of the curve.

References

Bala, Venkatesh, and Sanjeev Goyal. 1998. "Learning from neighbours." *Review of Economic Studies*, 65: 595-621.

Banerjee, Abhijit. 1992. "A simple model of herd behavior." *Quarterly Journal of Economics*, 107: 797-817.

Banerjee, Abhijit, and Drew Fudenberg. 2004. "Word-of-mouth learning." *Games and Economic Behavior*, 46: 1-22.

Bass, Frank. 1969. "A new product growth model for consumer durables." *Management Science*, 15: 215-227.

Bass, Frank. 1980. "The relationship between diffusion rates, experience curves and demand elasticities for consumer durables and technological innovations." *Journal of Business*, 53: 551-567.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy*, 100: 992-1026.

Blume, Lawrence E. 1993. "The statistical mechanics of strategic interaction." *Games and Economic Behavior*, 4: 387-424.

Blume, Lawrence E. 1995. "The statistical mechanics of best-response strategy revision." *Games and Economic Behavior*, 11: 111-145.

Centola, Damon. 2006. "Free riding and the weakness of strong incentives." Cornell University Department of Sociology Working Paper.

Chatterjee, Kalyan, and Susan Hu. 2004. "Technology diffusion by learning from neighbors." *Advances in Applied Probability*, 36: 355-376.

Coddington, E.A., and N. Levinson. 1955. *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.

Conley, Timothy G., and Christopher R. Udry. 2005. "Learning about a new technology: pineapple in Ghana." *Proceedings of the Federal Reserve Bank of San Francisco*.

Cowan, Robin, and Nicolas Jonard. 2004. "Network structure and the diffusion of knowledge." *Journal of Economic Dynamics and Control*, 28: 1557-1575.

David, Paul A. 1966. "The mechanization of reaping in the ante-bellum Midwest." In *Industrialization in Two Systems: Essays in Honor of Alexander Gershenkron*, ed. Henry Rosovksy. New York: Wiley.

David, Paul A. 1969. "A contribution to the theory of diffusion." Research Center in Economic Growth Memorandum No. 71, Stanford University.

David, Paul A. 1975. *Technical Change, Innovation, and Economic Growth*. Cambridge UK: Cambridge University Press.

David, Paul A. 2003. "Zvi Griliches on diffusion, lags, and productivity growth…connecting the dots." Stanford University Department of Economics Working Paper.

David, Paul A., and Trond E. Olsen. 1984. "Anticipated automation: a rational expectations model of technological diffusion," Technological Innovation Project Working Papers No. 2, Stanford University Center for Economic Policy Research.

David, Paul A. and Trond E. Olsen. 1986. "Equilibrium dynamics of diffusion when incremental technological innovations are foreseen." *Ricerche Economiche*, *(Special Issue on Innovation Diffusion)*, 40: 738-770.

De Groot, Morris. 1970. *Optimal Statistical Decisions*. New York: McGraw Hill.

Dodds, Peter Sheridan, and Duncan J. Watts. 2004. "Universal behavior in a generalized model of contagion." *Physical Review Letters,* 92 (21): 218701-1 to 218701-4.

Dodds, Peter Sheridan, and Duncan J. Watts. 2005. "A generalized model of social and biological contagion." *Journal of Theoretical Biology*, 232: 587-604.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller. 1981. "A non-symmetric responding logistic model for technological substitution." *Technological Forecasting and Social Change*, 20: 199-213.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller. 1983. "A nonuniform influence innovation diffusion model of new product acceptance." *Marketing Science*, 2: 273-296.

Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of thumb for social learning." *Journal of Political Economy*, 101: 612-643.

Ellison, Glenn, and Drew Fudenberg. 1995. "Word of mouth communication and social learning." *Quarterly Journal of Economics*, 110: 93-125.

Foster, Andrew and Mark Rosenzweig. 1995. "Learning by doing and learning from others: human capital and technical change in agriculture." *Journal of Political Economy*, 103: 1176-1209.

Geroski, Paul A. 2000. "Models of technology diffusion." *Research Policy*, 29: 603-625.

Golub, Benjamin, and Matthew O. Jackson. 2007. "Naive learning in social networks: convergence, influence, and the wisdom of crowds." Stanford University Department of Economics Working Paper.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American Journal of Sociology*, 83: 1420-1443.

Granovetter, Mark, and Roland Soong. 1988. "Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence." *Sociological Methodology*, 18: 69-104.

Griliches, Zvi. 1957. "Hybrid corn: an exploration of the economics of technological change." *Econometrica*, 25: 501-522.

Heckman, James J., Richard Robb, and James R. Walker. 1990. *Journal of the American Statistical Association*, 85: 582-589.

Heckman, James J., and Burt Singer. 1982. "The identification problem in econometric models for duration data." in *Advances in Econometrics,* ed. Werner Hildenbrand. Cambridge UK: Cambridge University Press.

Jackson, Matthew O., and Brian W. Rogers. 2006. "Relating network structure to diffusion properties through stochastic dominance." Working Paper, Stanford University.

Jackson, Matthew O., and Leeat Yariv. 2007. "Diffusion of behavior and equilibrium properties in network games." *American Economic Review Papers and Proceedings,* 97: 92-98.

Jensen, Richard. 1982. "Adoption and diffusion of an innovation of uncertain profitability." *Journal of Economic Theory*, 27: 182-193.

Jeuland, Abel P. 1981. "Incorporating heterogeneity into parsimonious models of diffusion of innovation." University of Chicago Center for Research on Marketing Working Paper No. 45.

Kapur, Sandeep. 1995. "Technological diffusion with social learning." *Journal of Industrial Economics*, 43: 173-195.

Karshenas, M., and Paul Stoneman. 1992. "A flexible model of technological diffusion incorporating economic factors with an application to the spread of colour television ownership in the UK." *Journal of Forecasting*, 11: 577-601.

Kirman, Alan. 1993. "Ants, rationality, and recruitment." *Quarterly Journal of Economics*, 108: 137-156.

Lancaster, T., and Stephen Nickell. 1980. "The analysis of reemployment probabilities for the unemployed." *Journal of the Royal Statistical Society, Series A*, 144: 141-156.

Lopez-Pintado, Dunia, and Duncan J. Watts. 2006. "Social influence, binary decisions, and collective dynamics." Columbia University, Institute for Social and Economic Research and Policy Working Paper.

Macy, Michael W.. 1991. "Chains of cooperation: threshold effects in collective action." *American Sociological Review*, 56: 730-747.

Mahajan, Vijay, and Robert A. Peterson. 1985. *Models for Innovation Diffusion*. Beverly Hills: Sage Publications.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass. 1990. "New product diffusion models in marketing: a review and directions for further research." *Journal of Marketing*, 54: 1-26.

Mansfield, Edwin. 1961. "Technical change and the rate of innovation." *Econometrica*, 29: 741-766.

Manski, Charles F. 2004. "Social learning from private experiences: the dynamics of the selection problem." *Review of Economic Studies*, 71: 443-458.

Munshi, Kaivan. 2004. "Social learning in a heterogeneous population: technology diffusion in the Indian green revolution." *Journal of Development Economics*, 73: 185-215.

Rogers, Everett M. 2003. *Diffusion of Innovations*, 5th edition. New York: Free Press.

Schelling, Thomas C. 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology*, 1: 143-186.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: Norton.

Smith, Lones, and Peter Sorensen. 2000. "Pathological outcomes of observational learning. " *Econometrica*, 68: 371-398.

Stoneman, Paul. 2002. *The Economics of Technological Diffusion*. Oxford: Blackwell.

Valente, Thomas W. 1995. *Network Models of the Diffusion of Innovations*. Cresskill NJ: Hampton Press.

Valente, Thomas W. 1996. "Social network thresholds in the diffusion of innovations." *Social Networks* 18: 69-89.

Valente, Thomas W. 2005. "Network models and methods for studying the diffusion of innovations." In *Models and Methods in Social Network Analysis*, ed. Peter J. Carrington, John Scott, and Stanley Wasserman. Cambridge UK: Cambridge University Press.

Vega-Redondo, Fernando. 2007. *Complex Social Networks.* Economic Society Monograph Series. Cambridge and New York: Cambridge University Press.

Young, H. Peyton. 2003. "The diffusion of innovations in social networks." In *The Economy as a Complex Evolving System, vol. III,* ed. Lawrence E. Blume and Steven N. Durlauf. Oxford: Oxford University Press.

Young, H. Peyton. 2005. "The spread of innovations through social learning." Brookings Institution Center for Social and Economic Dynamics, Working Paper No. 43.

Appendix: Proof of Proposition 1.

Define the function $H(t) = h(t)/p(t) = \dot{p}(t)/[p(t)(1-p(t))]$; this is well-defined for all $t > 0$ because by assumption $\dot{p}(0) = \bar{\gamma} > 0$, hence $p(t) > 0$ when $t > 0$. We need to show that $\dot{H}(t) < 0$.

For each parameter pair (λ, γ) let $q_{\lambda,\gamma}(t) = (1 - p_{\lambda,\gamma}(t))$ denote the *proportion* of the (λ, γ)-population *that has not yet adopted by* time t. The proportion of the total population that has not adopted by t is therefore

$$q(t) = \int q_{\lambda,\gamma}(t)d\mu . \tag{A1}$$

For each (λ, γ) we have

$$\dot{p}_{\lambda,\gamma}(t) = (\lambda p(t) + \gamma)q_{\lambda,\gamma}(t) . \tag{A2}$$

Integration with respect to μ yields

$$\dot{p}(t) = [\lambda(t)p(t) + \gamma(t)]q(t), \tag{A3}$$

where

$$\lambda(t) = q^{-1}(t)\int \lambda q_{\lambda,\gamma}(t)d\mu \text{ and } \gamma(t) = q^{-1}(t)\int \gamma q_{\lambda,\gamma}(t)d\mu . \tag{A4}$$

Note that $\lambda(t)$ and $\gamma(t)$ are the *expected* values of λ and γ in the population of non-adopters at time t. It follows that

$$H(t) = \dot{p}(t)/[p(t)q(t)] = \lambda(t) + \gamma(t)/p(t) . \tag{A5}$$

Claim: *For every $t > 0$, $\dot{\lambda}(t)p(t) + \dot{\gamma}(t) \leq 0$.* (A6)

Proof of claim. For every $t > 0$ we have

$$\dot{\lambda}(t) = \frac{\int \lambda \dot{q}_{\lambda,\gamma}(t)d\mu}{\int q_{\lambda,\gamma}(t)d\mu} - \frac{[\int \lambda q_{\lambda,\gamma}(t)d\mu][\int \dot{q}_{\lambda,\gamma}(t)d\mu]}{[\int q_{\lambda,\gamma}(t)d\mu]^2}, \quad \text{(A7)}$$

and

$$\dot{\gamma}(t) = \frac{\int \gamma \dot{q}_{\lambda,\gamma}(t)d\mu}{\int q_{\lambda,\gamma}(t)d\mu} - \frac{[\int \gamma q_{\lambda,\gamma}(t)d\mu][\int \dot{q}_{\lambda,\gamma}(t)d\mu]}{[\int q_{\lambda,\gamma}(t)d\mu]^2}. \quad \text{(A8)}$$

To show that $\dot{\lambda}(t)p(t) + \dot{\gamma}(t) \leq 0$, multiply (A7) by $p(t)$ and add it to (A8); after simplifying we obtain the equivalent condition

$$[\int (\lambda p(t) + \gamma)\dot{q}_{\lambda,\gamma}(t)d\mu][\int q_{\lambda,\gamma}(t)d\mu] - [\int (\lambda p(t) + \gamma)q_{\lambda,\gamma}(t)d\mu][\int \dot{q}_{\lambda,\gamma}(t)d\mu] \leq 0. \quad \text{(A9)}$$

We need to show that (A9) holds for every $t > 0$. (Notice that t does not vary in this expression; t is fixed and integration is taken with respect to λ and γ.) We know from (A2) that $\dot{q}_{\lambda,\gamma}(t) = -(\lambda p(t) + \gamma)q_{\lambda,\gamma}(t)$ for every λ, γ, and t. Substituting this into (A9) we obtain

$$[\int (\lambda p(t) + \gamma)^2 q_{\lambda,\gamma}(t)d\mu][\int q_{\lambda,\gamma}(t)d\mu] \geq [\int (\lambda p(t) + \gamma)q_{\lambda,\gamma}(t)d\mu]^2. \quad \text{(A10)}$$

Fix $t > 0$ and define the random variables

$$X = (\lambda p(t) + \gamma)\sqrt{q_{\lambda,\gamma}(t)} \text{ and } Y = \sqrt{q_{\lambda,\gamma}(t)}. \quad \text{(A11)}$$

The realizations of X and Y are determined by random draws from μ. Thus (A10) follows directly from Schwarz's inequality: $E[X^2]E[Y^2] \geq (E[XY])^2$. This establishes the claim.

We can now apply this result to show that $H(t)$ is strictly decreasing in t for all $t > 0$. Differentiating $H(t)$ we obtain

$$\dot{H}(t) = \dot{\lambda}(t) + \dot{\gamma}(t)/p(t) - \gamma(t)\dot{p}(t)/p^2(t). \qquad (A12)$$

By the above claim, $\dot{\lambda}(t)p(t) + \dot{\gamma}(t) \leq 0$, so division by $p(t) > 0$ yields $\dot{\lambda}(t) + \dot{\gamma}(t)/p(t) \leq 0$. Thus the sum of the first two terms on the right-hand side of (A12) is nonpositive. But the last term is strictly negative, because $\gamma(t) > 0$ for all $t > 0$ given the initial condition $\bar{\gamma} = \gamma(0) > 0$. Hence $\dot{H}(t) < 0$ as was to be shown.