

, Adam, Manuel Trajtenberg and Michael Fogarty (2000), 'Knowledge spillovers and patent citations: evidence from a survey of inventors', *American Economic Review*, Papers and Proceedings (May), 215–18.

n, Richard C. and Peter C. Reiss (1984), 'Tests of a Schumpeterian model of R&D and market structure', in Z. Griliches (ed.), *R&D, Patents and Productivity*, Chicago, University of Chicago Press for the National Bureau of Economic Research, pp. 175–208.

Office of Technology Assessment and Forecast (OTAF) (1976), *US Department of Commerce, Patent and Trademark Office, Sixth Report*, Washington, DC, US Government Printing Office.

tt, Keith (1988), 'Uses and abuses of patent statistics', in A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*, Amsterdam: Elsevier Science, pp. 509–36.

enberg, Nathan (1976), 'Technological change in the machine tool industry, 1840–1910', in N. Rosenberg (ed.), *Perspectives in Technology*, Cambridge, Cambridge University Press, pp. 9–31.

akibara, Mariko and Lee G. Branstetter (2001), 'Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms', *Rand Journal of Economics*, 32 (1), 77–100.

erman, Brian (2002), 'International Patent Classification – US SIC concordance', <http://www.rotman.utoronto.ca/~silverman/ipcsic/ipcsicfiles.ZIP>

tenberg, Manuel, Adam Jaffe and Rebecca Henderson (1997), 'University versus corporate patents: a window on the basicness of invention', *Economics of Innovation and New Technology*, 5 (1), 19–50.

15. Equilibrium, epidemic and catastrophe: diffusion of innovations with network effects

Luís M.B. Cabral*

1 INTRODUCTION

It seems that important papers are characterised by long publication lags. Maskin's famous mechanism design theorem and Holmstrom's seminal paper on managerial concerns each took about 20 years to get published.¹ Prominent among the list of famous works that remained unpublished for a long time is Paul David's 'Contribution to the theory of diffusion' (David, 1969). In that paper, David develops an equilibrium model of new technology adoption and shows how S-shaped diffusion paths reflect heterogeneity among adopters.

In this chapter, I too focus on the issue of diffusion of innovations, specifically innovations subject to network effects. Like David and others, I start from an equilibrium model of adopter heterogeneity. However, I will argue that, in the presence of strong network effects, the nature of the adoption process is quite different from what was previously characterised. In particular, I show that network effects imply discontinuous adoption paths – mathematically speaking, a catastrophe.

In a previous paper (Cabral, 1990), I noted how network effects may lead to discontinuous adoption paths. This chapter goes beyond Cabral (1990) in two ways. First, I provide a more precise set of conditions under which a catastrophe takes place (section 3). Second, I suggest a possible test to distinguish between alternative theories of new technology adoption (section 5).²

S-shaped diffusion paths, one of the most robust empirical regularities found in the literature, are consistent with a number of theories. I consider two types: (1) equilibrium diffusion theories based on adopter heterogeneity, and (2) epidemic theories based on some form of imperfect information and/or word-of-mouth effects. I start from a model of the first type and add network effects to it. I then compare it to a model of the second type, also allowing for the possibility of network effects.

2 AN EQUILIBRIUM MODEL OF NEW TECHNOLOGY ADOPTION

Consider a new technology available from some time t_0 . The cost of adopting such innovation is c_t per period. That is, upon adoption, a flow cost c_t must be paid. I assume that c_t is decreasing, which reflects gradual, post-invention, technological development, as well as increased competition in supply.

I am particularly interested in the case when the benefit from an innovation can be measured by its use. For example, the benefit from having a telephone is proportional to the use that is made of such telephone (or, if we also want to consider the 'stand-alone' benefit from owning a telephone, then total benefit is a linear function of use). Formally,

Assumption 1 Each adopter's benefit is proportional to use.

I am also interested in the case when the innovation is subject to network effects, that is, the case when adoption benefits are increasing in the number of adopters. Specifically, suppose that each potential user derives a benefit from communicating with a set of other users. Such benefit can only be gained if the other users are also hooked up to the network, that is, if the other users have adopted the innovation as well. Suppose, moreover, that the event of being part of the list of desirable links is independent of the user's type. Then the use of (and benefit from) the innovation is a linear function of the number of users.

Assumption 2 Each adopter's willingness to pay is a linear function of the number of users.

I assume that potential adopters are different from each other. Specifically, each potential adopter is characterised by a parameter $\theta \in \Theta$ that measures its willingness to pay for the innovation. Specifically, I assume that

Assumption 3 Each adopter's use of the innovation is proportional to the adopter's type θ .

Assumption 4 $\theta \sim N(\mu, \sigma)$.

The above assumptions imply that, upon adoption (which I assume is irreversible), an adopter of type θ receives a benefit flow given by

$$u_t^\theta = \theta((1 - \lambda) + \lambda n_t), \quad (15.1)$$

where n_t is the measure of adopters at time t . The parameter λ measures the importance of network effects. In the extreme when $\lambda = 0$, benefit is simply given by θ (stand-alone utility), that is, independent of network size. In the opposite extreme, when $\lambda = 1$, stand-alone utility is zero and benefit is proportional to network size.

It is straightforward to show that, if type θ finds it optimal to adopt before time t , then the same is true for type $\theta' > \theta$. It follows that, in equilibrium, the set of adopters at time t is given by all types with θ greater than some critical value. Let $\theta' = \theta'(t)$ be such critical value. The equilibrium conditions are then

$$c_t = (1 - \lambda)\theta' + \lambda\theta'N_t, \\ n_t = \int_{\theta'}^{\infty} f(\theta)d\theta,$$

where f is the density of θ . The first equation guarantees that the marginal adopter (type θ') is just indifferent between adopting and not adopting at time t : the left-hand side is the (flow) cost of adoption, whereas the right-hand side is the benefit from adoption. The second equation is a 'closure' condition: it implies that the network size is the measure of all adopters of type greater than the marginal type.

The above equations can be combined to yield the following equilibrium condition:

$$1 - \lambda F(\theta') = \frac{c_t}{\theta'}, \quad (15.2)$$

where F is the cumulative distribution function of θ .

For each time t , and the corresponding value of c_t , (15.2) can be solved for θ' . Each value of θ' in turn corresponds to a value of n . Therefore, (15.2) induces an equilibrium correspondence $E(t)$ giving the possible equilibrium values n_t for each t .³ Although the graph $E(t)$ is a continuous and smooth manifold (Cabral, 1990), the equilibrium correspondence can, in principle, be multi-valued. In fact, in the presence of network effects this would not be a surprising feature.

3 NETWORK EXTERNALITIES AND CATASTROPHIC ADOPTION PATHS

If network effects are non-existent or mild, then (15.2) induces a single-valued equilibrium correspondence $E(t)$ and a continuous equilibrium adoption path (EAP). This is illustrated in Figure 15.1. The left-hand side

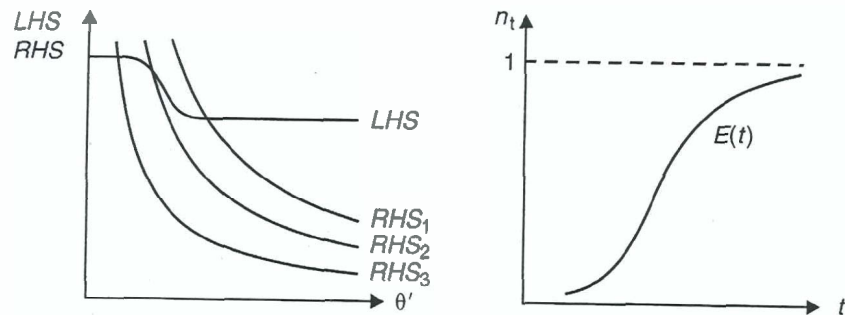


Figure 15.1 Continuous adoption path with 'mild' network externalities ($\mu = 5$, $\sigma = 1$, $\lambda = .25$)

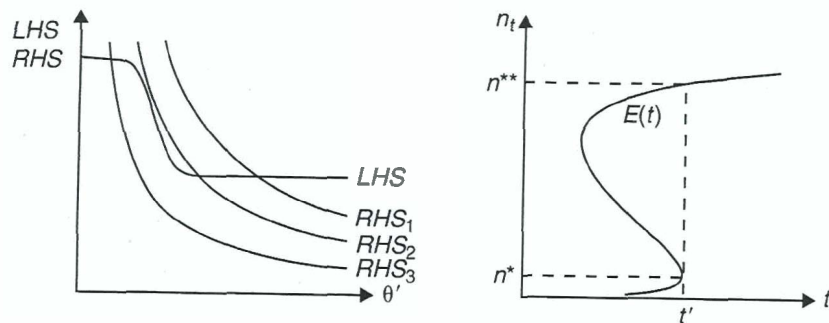


Figure 15.2 Strong network externalities and catastrophe adoption path ($\mu = 5$, $\sigma = 1$, $\lambda = .5$)

depicts the two sides in (15.2). As can be seen, for each value of t (and c_t), there exists a unique solution to the equation. This implies that the equilibrium correspondence $E(t)$ is single valued and there is a unique EAP, namely $n_t = E(t)$.

Consider now the case when network externalities are significant. This case is illustrated in Figure 15.2. The left-hand side of the figure shows that, for values of t slightly greater than the one corresponding to RHS_2 , several solutions exist to equation (15.2) (in this figure, equation (15.2) corresponds to $LHS = RHS_1$). This results in an equilibrium correspondence $E(t)$ that is multi-valued for an interval of values of t . Such equilibrium correspondence is depicted in the right-hand side of the figure. A multi-valued equilibrium correspondence means that there are multiple possible EAPs, in fact a continuum of them.

Despite this multiplicity of equilibria, it can be readily seen that every EAP is discontinuous at least for some t . In fact, the most reasonable EAP consists of following the lower branch of $E(t)$ up to time t' and then jumping from $n_t = n^*$ to $n_t = n^{**}$. In the jargon of topology, the point (t, n^*) is a catastrophe point: although the equilibrium correspondence is continuous, a small increase in t implies a discontinuous change in the value of n_t .

Figures 15.1 and 15.2 suggest that catastrophes are more likely when network externalities are stronger. My first result formalises this intuition:

Proposition 1 For given μ and σ , a catastrophe point exists if and only if $\lambda > \lambda^*$. Conversely, for given μ and $\lambda > 0$, a catastrophe point exists if and only if $\sigma < \sigma^*$.

A formal proof may be found in the Appendix.

In terms of actual behaviour, one would normally not expect to observe a discontinuous EAP like the one suggested above. The shift from n^* to n^{**} would very likely occur over a period of time. In fact, if one assumes that potential adopters make their decisions at time t based on the installed base at time $t - \tau$ (that is, there is an observation lag τ); then it can be shown that, for small τ , the adoption path follows closely $E(t)$ up to time t' and then moves gradually towards the upper portion of $E(t)$ along a concave path. The result of this process is an S-shaped adoption path.⁴

4 ALTERNATIVE THEORIES

As I mentioned before, the economics literature has produced a large number of theoretical explanations consistent with the stylised fact of an S-shaped adoption path. Any claim for the worth of a new explanatory theory has to be confronted with competing claims.

At the risk of over-generalising, we may classify the different theories into two different categories.⁵ First, we have the equilibrium diffusion theories based on adopter heterogeneity. These theories are similar to the model I presented above (or vice versa), except for the inclusion of network effects. In words, these theories explain diffusion as a result of adopter heterogeneity. Specifically, an S-shaped adoption path results from the shape of the cumulative distribution function of the adopters' type. In particular, the steep portion of the adoption path corresponds to a high density of adopters around the relevant valuation parameter.

The epidemic theories, which are based on some form of imperfect information, provide an alternative explanation for S-shaped diffusion paths. In its simplest form, the epidemic theory assumes that potential adopters become aware of the *existence* of the innovation by word of mouth. Word-of-mouth dynamics are known to have the dynamics of medical epidemics, where the rate of change is proportional to the product of the number of infected and not-infected agents. This results in an S-shaped path very similar to the empirically observed paths.⁶

To summarize: an S-shaped adoption path does not require a catastrophe. In fact, it does not even require that there are network externalities at all or that adopters are heterogeneous, so long as there is imperfect information of some sort. Therefore, the simple observation of the aggregate rate of adoption is not sufficient to validate any theory in particular. In the next section, I focus on empirical implications that separate the different theories.

5 TESTING BETWEEN THEORIES

As I have argued in the previous sections, there exist many theories that are consistent with an S-shaped adoption path. Different theories must then be distinguished by observables other than the diffusion path. My second result implies one such test for the case when the intensity of use can be easily measured.

As I mentioned before, one natural motivation for the utility function (15.1) is the distinction between stand-alone and network-related benefit. Network benefit is proportional to total use, which in turn is proportional to θn_t . Based on this observation, different theories have different implications with respect to the time path of average use, a_t , given by

$$a_t = \frac{\int_{\theta}^{\infty} (1 - \lambda + \lambda n_t) \theta g_t(\theta) d(\theta)}{\int_{\theta}^{\infty} g_t(\theta) d\theta},$$

where $g_t(\theta)$ is the density of θ types who adopt by time t . Specifically, we have the following results:

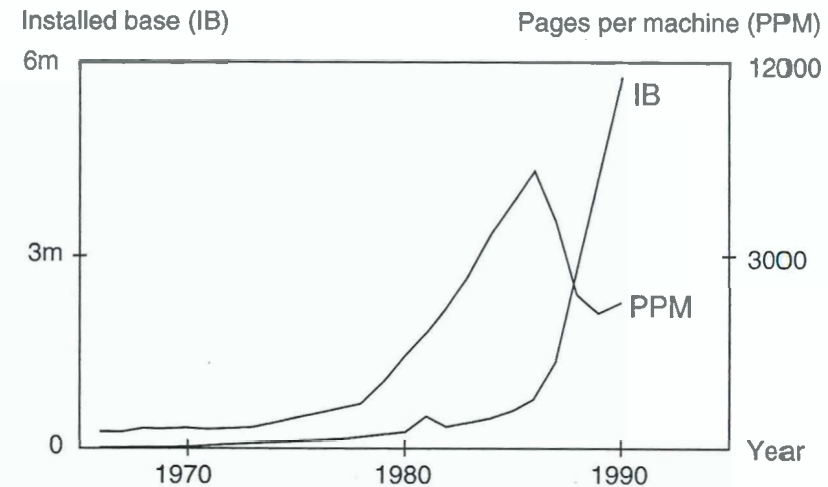
Proposition 2 Under equilibrium adoption with heterogeneous adopters, $\lambda = 0$ implies that a_t is decreasing for all t ; $\lambda = 1$ implies that a_t is increasing for low t and decreasing for high t .

Proposition 3 Under epidemic diffusion, $\lambda = 0$ implies that a_t is constant for all t ; $\lambda = 1$ implies that a_t is increasing for all t .

Table 15.1 summarises Propositions 2 and 3.

Table 15.1 Summary of Propositions 2 and 3

Theory	No net effects	Net effects
Heterogeneous adopters	$\rightarrow$	$\nearrow \searrow$
Epidemic	$\rightarrow$	$\nearrow$



Source: Farrell and Shapiro (1992); see also Economides and Himmelberg (1995).

Figure 15.3 Fax machines in the US: installed base and intensity of use

6 AN EXAMPLE: FAX MACHINES

In order to test the applicability of my theoretical results, I consider data on the diffusion of fax machines in the USA. Figure 15.3 plots the value of the installed base of fax machines as well as the average use per machine (pages per machine), from the mid-1960s to 1990.

The data seem roughly consistent with the theory of diffusion with heterogeneous adopters and strong network externalities. First, around 1987 there was a sharp increase in the installed base, which suggests a catastrophe point in the diffusion of fax machines. Moreover, the time path of usage per machine seems consistent with the prediction of Proposition 2 for the case $\lambda = 1$: the value of a_t is initially increasing and then decreasing.

Several qualifications are in order, however. First, the time series in Figure 15.3 is a bit too short to uncover a clear pattern in the evolution of a_t . I am currently working on trying to extend this series, in the hope of finding stronger results. One problem with extending the series to the 1990s is that serious consideration must be given to the emergence of email as an alternative to fax (including the emergence of electronic faxing).

Second, it should be noted that the value of pages per minute (PPM) is not necessarily the best measure of a_t . In fact, it is not uncommon for fax machines to be shared among several users. For this reason, using PPM as a measure of average use implicitly amounts to assuming that the number of users per machine remained constant throughout the sample period. However, anecdotal evidence suggests that, as the price of fax machines dropped over time, so did the number of users per machine. Figure 15.3 is therefore consistent with the epidemic-theory-cum-network-effects story. In other words, the time path of PPM is consistent with an ever increasing path of a_t .

7 CONCLUDING REMARKS

Referring to the plethora of theories of S-shaped adoption paths, David (1969: II/13) argues that

It would be possible to find *some* pair of specifications [of $F(\theta)$ and c_t] that would give rise to a diffusion path of the appropriate shape. Hence, meaningful efforts to distinguish between and verify alternative models of diffusion ought to involve some attempt at direct empirical validation of the component specifications, including the postulated characteristics of the distribution $f(X)$.

The analysis in this chapter suggests that the problem is deeper than that: not only there are multiple functional forms consistent with a given time path; there are multiple theories that would produce identical outcomes. On the positive side, Propositions 2 and 3 suggest that, even with aggregate data only, there are ways of distinguishing between competing theories.

To conclude, I should acknowledge that the model in this chapter is based on a somewhat narrow class of innovations, namely communication technologies, where benefits are derived from actual links between potential users. However, as in much of the networks literature, results from the direct network effects model can be extended to the case of indirect effects as well. The crucial point is that Assumptions 1 and 2 (or a variation thereof) hold.

APPENDIX

Proof of Proposition 1 In a continuous equilibrium path (no catastrophe), all values $\theta > 0$ will correspond to the marginal adopter at some time t . The condition determining the marginal adopter is

$$1 - \lambda F(\theta') = \frac{c_t}{\theta'} \quad (15.3)$$

A necessary and sufficient condition for the EAP to be continuous is that the RHS cut the LHS from above for every t , that is,

$$\left| \frac{\partial}{\partial \theta} (1 - \lambda F(\theta)) \right| \leq \left| \frac{\partial}{\partial \theta} \left(\frac{c_t}{\theta} \right) \right|,$$

or simply

$$\lambda f(\theta) \leq \frac{c_t}{\theta^2},$$

for all $\theta > 0$. Substituting (15.3) for c_t and simplifying, we get

$$\lambda(F(\theta) + \theta f(\theta)) \leq 1. \quad (15.4)$$

The next step is to show that $F + \theta f$ is greater than 1 for some value of θ .⁷ The derivative of $F + \theta f$ with respect to θ is

$$\frac{\partial}{\partial \theta} (F(\theta) + \theta f(\theta)) = f(\theta) + f(\theta) + \theta \left(-\frac{\theta - \mu}{\sigma^2} \right).$$

It follows that, for θ sufficiently large,

$$\frac{\partial}{\partial \theta} (F(\theta) + \theta f(\theta)) < 0.$$

Since, in addition

$$\lim_{\theta \rightarrow \infty} (F(\theta) + \theta f(\theta)) = 1,$$

it follows that there exists a θ such that $F(\theta) + \theta f(\theta) > 1$. It follows that if λ is sufficiently large, then the condition (15.4) is violated for some θ .

The second part of the proposition is quite straightforward. For $\theta = \mu$, (15.4) reduces to

$$\lambda \left(\frac{1}{2} + \mu \frac{1}{\sigma \sqrt{2\pi}} \right) \leq 1.$$

Clearly, for given μ and $\lambda > 0$ this condition is violated for σ sufficiently close to zero.

Proof of Proposition 2 Under heterogeneous adopter diffusion, per capita use is given by

$$a_t = \frac{\int_0^\infty ((1-\lambda) + \lambda(1-F(\theta')))\theta f(\theta)d\theta}{\int_0^\infty f(\theta)d\theta}$$

Given that θ is normally distributed, we have

$$a_t = ((1-\lambda) + \lambda(1-F(\theta')))\theta\sigma^2 \frac{f(\theta')}{1-F(\theta')}.$$

If $\lambda = 0$, then $a_t = \sigma^2 f(\theta') / (1 - F(\theta'))$, which is increasing in θ . Since θ' is decreasing in t , it follows that a_t is decreasing in t . If $\lambda = 1$, then $a_t = \sigma^2 f(\theta')$. Since σ is decreasing in t , the value of a_t follows the value of $f(\theta')$: increasing for low values of t , decreasing for high values of t .

Proof of Proposition 3 Under epidemic diffusion, the population of adopters at time t is a representative sample from the population of potential adopters. Moreover, type θ 's use at time t is given by $(1-\lambda)\theta + \lambda\theta n_t$. Together, these facts imply that

$$a_t = (1-\lambda)\mu + \lambda\mu n_t.$$

If $\lambda = 0$, then $a_t = \mu$, which is constant over time. If $\lambda = 1$, then $a_t = \mu n_t$, which is increasing over time.

NOTES

* My first work in this area dates back to a second-year student paper at Stanford (in 1986) that eventually led to Cabral (1990). I am grateful to Paul David, Brian Arthur, and many others who encouraged me on this line of research. Regretfully, I alone remain responsible for all the shortcomings of this and previous related papers.

1. Both appeared in a recent issue of the *Review of Economic Studies*.
2. Cowan (2005), also included in these proceedings, considers a model of cycles in art appreciation and prices. Although I do not consider the possibility of cycles (I assume that technology adoption is irreversible), our models share the prediction that, over time, consumers will shift between equilibria.
3. Specifically, $E(t)$ is obtained by solving

$$1 - \lambda(1 - n_t) = \frac{c_t}{G(1 - n_t)},$$

where $G(\cdot)$ is the inverse of $F(\cdot)$.

4. Cf Cabral (1990). Notice I do not call this an *equilibrium* adoption path since, for a period of time, the system is in disequilibrium, gradually moving from a low-adoption to a high-adoption static equilibrium.
5. See Geroski (1999) for a recent survey.
6. Jensen (1982) proposes an interesting variant of the epidemic theory based on imperfect information about the *value* of the innovation. Specifically, he assumes that adopters differ with respect to their prior beliefs that the innovation is profitable (other than this, he assumes adopters are identical). He shows that, starting from a uniform distribution of prior beliefs, an S-shaped equilibrium adoption path is obtained.
7. One side-result of the above condition is that, for given μ and σ , the EAP is continuous if λ is sufficiently small.

REFERENCES

- Cabral, Luís (1990), 'On the adoption of innovations with "network" externalities', *Mathematical Social Sciences*, 19, 299–308.
- Cowan, Robin (2005), 'Waves and cycles: explorations in the pure theory of price for fine art', chapter 7 in this volume.
- David, Paul A. (1969), 'A contribution to the theory of diffusion', Research Center in Economic Growth Memorandum No. 71, Stanford University.
- Economides, Nicholas and Charles Himmelberg (1995), 'Critical mass and network size with application to the US FAX market', NYU Stern School of Business Working Paper No. EC-95-11, August.
- Farrell, Joseph and Carl Shapiro (1992), 'Standard setting in high-definition television', *Brookings Papers on Economic Activity* (Microeconomics), 1–77.
- Geroski, Paul (2000), 'Models of technological diffusion', *Research Policy*, 29, 603–25.
- Jensen, Richard (1982), 'Adoption and diffusion of an innovation of uncertain profitability', *Journal of Economic Theory*, 27, 182–93.