Herd Behaviour as an Incentive Scheme*

Nicolas Melissas

Appendix two with Marc Pauly

Institut d'Anàlisi Econòmica (CSIC)

June 2002

Abstract

We introduce herding in a model subject to public-good problems. We show how herding, by reducing free-rider problems, may increase efficiency.

JEL classification: D83,D82,D62

Keywords: free-rider problem, public good, information externality, herd behaviour, social learning

---

*Correspondence can be sent to Nicolas Melissas (nicolas.melissas@uab.es), IAE, Campus UAB, 08193 Bellaterra, Barcelona. I am very grateful to my thesis advisor M.Dewatripont for his many helpful suggestions. This paper also greatly benefited from comments made by P.Bolton and P.Legros. I thank seminar participants at ECARE, DELTA, IAE, GREMAQ and WZB. I also benefited from comments made by A.Banerjee, M.Castanheira, J.Gyntelberg, G.Roland, M.Ruckes, X.Vives and J.Zwiebel. Financial support from the TMR network "The evolution of market structure in network industries" and from the Inter university Poles of Attraction Program, Contract PAI P4/28 is gratefully acknowledged. Finally, I am also very grateful to M.Castanheira for his many encouragements at the start of my research work.

1 Introduction

People often infer information out of the actions of other people. For example, when making their purchasing decisions consumers often choose the most popular brand because they think that its popularity indicates a better price/quality combination[1]. People do not go and eat in an empty restaurant, because they believe that the food quality is low. When arguing with someone, try to strengthen your argumentation by claiming that everybody agrees with you on that point. This trick (even if it is not true) often succeeds in convincing more reluctant minds that you are right.

This tendency to base decisions largely on the observed decisions of other agents has recently been modeled as information externalities. Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) (henceforth BHW) made the first models which stressed the inefficiencies of these information externalities in a context of social learning[2]. Both models consider a population of individuals each endowed with a private, costless and imperfect signal concerning the desirability of a course of action. People decide sequentially whether to adopt or reject a given course of action. People observe which actions were taken by the persons who moved before them, but they do not observe their signals. If enough individuals have adopted the same behaviour, then each subsequent individual neglects her private signal and adopts the same action as the one undertaken by her predecessor, because the informativeness of their combined actions is higher than the informativeness of any one signal. More interestingly they also showed that herding is quite likely to cause a "bad outcome", i.e. an outcome where all (or the vast majority of all) players adopt an action which ex post turns out to be suboptimal.

Subsequently a number of other papers appeared which also stressed different inefficiencies present in different herding environments (=an environment in which players may herd on one another's actions). For example,[3] Chamley and Gale (1994) consider a set-up similar to BHW except that all players have the possibil-

[1] Caminal and Vives (1996) analyse a game where firms engage in price competition in order to become more popular and benefit from these information externalities.

[2] With a "context of social learning" we mean a context where one person must choose an action out of a prespecified action set and which has, prior to her decision, the opportunity to learn out of the choices made by her predecessors.

[3] The models we briefly discuss in this paragraph are among the most famous ones, but the list is certainly not exhaustive.

ity to wait in order to observe how many players invest in the current period and to make their investment decisions in the next period based on superior information. They showed how in their context bad outcomes and inefficient waiting may occur in equilibrium. Avery and Zemsky (1998) also consider a set-up similar to BHW except that they add a competitive market maker in the picture who sets the price of an investment asset on the basis of all available information. They show that due to herding short-run price bubbles can occur provided that traders are uncertain about the precision of the other traders' signals. Vives (1993) shows that in the presence of a continuous action space in the long run bad outcomes do not arise, i.e. eventually everyone will adopt the right action. However in the presence of noise (when observing the actions of the other players) the rate of convergence towards the right action is slow.

To summarise, so far the literature on social learning mainly stressed different inefficiencies (short-run price bubbles, bad outcomes, inefficient waiting, slow learning, ...) present in herding environments. In this paper we add a public-good problem in a variant of the standard herding models and we argue that herding[4], by reducing free-rider problems, also possesses some efficiency-increasing properties.

We study the following set-up. We assume that a fixed number of firms enter sequentially in an emerging market. Each firm is run by one manager. Upon entrance in the market managers must choose a technology. Managers can choose to adopt an "old" technology. Adopting this "old" technology is easy: it doesn't require effort from the manager. Managers can also choose to exert effort to check the existence of a new (and more profitable) technology. With a probability p a new technology exists, which will always be invented if effort is provided. Following Banerjee (1992) and BHW (1992) we work under an "exogenous queue" assumption: managers only choose their effort level at their time of entry. The new technology is a public good: if a manager invents it, all his predecessors and subsequent movers can freely copy it and enjoy a higher payoff. Managers act strategically in the sense that they all may exert no effort and adopt the old technology in the hope to free-ride on the effort of another manager inventing a better technology.

[4]With "herding" we mean a behaviour where one person observes the action(s) of her predecessor(s), updates her prior beliefs and then has *more incentives* to imitate her predecessor(s). This definition allows us to classify Vives' (1993) paper also in the herding literature and it permits us to better explain the originality of this paper.

We first show that, for efficiency reasons, in our model the first manager (and only the first manager) should exert effort. This is because we assume that it is more efficient to adopt the new technology from scratch rather than first adopting the old technology and then switch to the new one. Inefficient switching costs can only be avoided if the first manager exerts effort.

Next we compute the equilibrium outcome if all players observe their predecessors' effort levels and technologies. In this case in the unique subgame perfect equilibrium only the last manager who enters the market exerts effort for sure. The intuition goes as follows: the last manager observes all his predecessors adopting the old technology. However he also observes that none of his predecessors exerted effort. Therefore, he knows that the new technology still exists with probability p. Hence he has no reason to herd on his predecessors' technologies[5] (i.e. to "blindly" imitate his predecessors' technologies), and instead exerts effort to try to invent the superior technology. This equilibrium performs badly in terms of efficiency because if the last manager successfully innovates, all the other managers incur an inefficient switching cost.

Finally, we assume that all managers only observe their predecessors' technologies (and not their effort levels). We show that in this environment late movers may herd on the technologies adopted by the early movers. To see this, consider the following two-manager example. Assume manager two observes manager one adopting the old technology. Depending on manager two's beliefs (concerning manager one's effort level), he may infer the inexistence of the new technology out of the first manager's action. This hampers manager two's incentives to innovate (i.e. manager two herds on manager one's technology[6]). Manager one correctly anticipates that manager two is going to "blindly imitate" his technology. Therefore, manager one knows that he cannot free-ride on manager two's effort level. Therefore, it's optimal for manager one to exert effort. Therefore, the new technology may be invented at time one (which is the most efficient outcome).

Our model is most related to other ones which analysed waiting games in the presence of information externalities. If a manager "waits" (by not exerting effort and adopting the old technology) he gives bad news to his successors. Therefore we

[5]Therefore, we term this environment as a non-herding one.

[6]Therefore, we term this environment as a herding one.

first explain how this paper relates to Chamley and Gale (1994), Chamley (1997), Hendricks and Kovenock (1989) and Zhang (1997).

Chamley (1997) analyses a set-up similar to Chamley and Gale (1994)[7]: all players receive a signal concerning the realised state of the world, investment is only profitable in the good state. All players have the possibility to wait, to observe how many other players invested in the previous period and to make an investment decision on the basis of more information. Chamley shows that - depending on the values of the parameters - this waiting game may be characterised by two symmetric and stable equilibria. In the so-called high-activity equilibrium, the author shows how the occurence of an informational cascade increases a player's incentives to invest early. The argument goes as follows: assume the world is only populated by optimists and pessimists. Due to a low investment cost, pessimists also face a positive gain of investing. Assume everyone anticipates that all players (optimists as well as pessimists) invest at time one. By definition, this represents an informational cascade: all players - irrespective of their types - undertake the same action at time one. Given this anticipation, no player wants to delay her investment action because by doing so she only faces a discounting cost wthout receiving any informational gain. Therefore, if players anticipate an informational cascade, this adversely affects their gain of waiting and increases their incentives to also invest at time one. This "incentive-mechanism" is actually similar to the one operating in this paper: future herding reduces a manager's gain of not exerting effort, thereby increasing her incentives to exert effort. A similar mechanism is also at work in Zhang (1997) and in Cripps, Keller and Rady (2000)[8].

However, in our paper this "incentive-mechanism" may increase efficiency (as compared to the outcome one gets in a non-herding environment). To understand what drives our result we must first answer the following questions: what is the non-herding environment in Chamley and Gale (1994), Chamley (1997) and Zhang (1997)? Which outcome do we then get? In those papers the non-herding environment corresponds to the no-private-information case. For example, if all players would possess the same signal (or would truthfully exchange their signals), the act of investing would not entail an information externality and their models would

[7] In the latter model pessimists do not possess an investment option and can therefore not invest. In Chamley (1997) all players, irrespective of their types, can invest.

[8] The latter paper is, however, only remotely related to ours. Their model is primarily intended to show how the use of non-stationary Markov-strategies alleviate free-rider problems, whereas we show how free-rider problems can be alleviated by herding.

then be void of any informational cascades or herding. In those models, if all players would truthfully exchange their signals, first best applies i.e. everyone would take the correct ex ante investment decision without delaying (and this is, ex ante, the most efficient outcome). Why is it that in our model the herding environment may be more efficient than the non-herding one and not in theirs? In our model the non-herding environment is prone to a public-good problem. Chamley and Gale (1994), Chamley (1997) and Zhang (1997) also possess a public good in the sense that if a player invests, this releases some (public) information to the other players. However, this public good is only present in the herding environment. If players were to exchange their private information, their model would be void of any public good. This is the crucial difference between those models and ours.

In Hendricks and Kovenock (1989) two oil firms possess a private imperfect signal concerning the profitability of drilling an exploratory oil well. If one firm drills, then the other firm costlessly observes whether there is oil or not and (in case of success) makes a riskless investment decision in the next period. In contrast to Chamley and Gale (1994), Chamley (1997) and Zhang (1997) their non-herding environment is also prone to public-good problems. To see this, assume both firms truthfully exchange their signals. In that case, both firms agree that with some probability p there is oil underneath the sea. If firm one expects firm two to drill at time one, then she strictly prefers to wait because she then costlessly and perfectly learns whether any oil is present underneath the sea or not. So in their model, the public good in the non-herding environment follows from their assumption that the true state of the world falls in the public domain once a firm drilled a well. Their model was primarily intended to show how information externalities may induce both firms to let their leases expire without drilling any well. The authors therefore only analysed the equilibrium outcome in the herding environment. Hence, their model does not show how information externalities, by affecting incentives to drill, may improve efficiency.

This paper is organised as follows. In section two we explain the basic assumptions of our model. In section three we introduce a social planner in our model and compute the most efficient strategy profile. We then analyse the workings of our model under the assumption of observable effort levels and observable technologies (section 4). In section 5 we work under the assumption that players do not observe one another's effort levels. First we illustrate how herding may increase efficiency by focusing on a simple equilibrium in which only one manager exerts effort. In section 5.2 we analyse the case where N managers exert effort (with a certain probability) in equilibrium. In section 6 we discuss some properties and the robustness of the main result of our model. Final comments are summarised in section 7.

2 The Model

2.1 The general framework.

We consider a simple model of dynamic entry in an emerging market. Our model counts two phases: a first one called the *adoption phase* and a second one called the *production phase*. Throughout the paper time is discrete and is denoted by $t = 1, ..., \overline{K}$ $(2 < \overline{K} < \infty)$. The adoption phase starts at time $t = 1$ and stops at time $t = K$ $(1 < K < \overline{K})$. The production phase starts at time $t = K + 1$ and lasts until time $\overline{K}$. In the adoption phase we assume that in each different period one new firm enters in the emerging market. Each firm is run by one risk neutral manager/entrepreneur. Henceforth we call manager t, the manager who enters in the emerging market at time t. In the adoption phase is decided which technology will be used by all our managers in the production phase. In the production phase all managers receive their payoffs by producing and selling their goods in a finite market.

Manager t must adopt a technology at his time of entry.[9] Manager t's initial[10] technology depends on the available technology at time t (denoted by $T_t \in \{o, n\}$), and on manager t's effort level (denoted by $E_t \in \{0, e\}$ $(e > 0)$). o represents an "old", not very profitable, technology. n represents a new, more profitable, technology. $T_1 = o$. $E_t = 0$ (e) means that manager t doesn't exert (exerts) effort. If $T_t = o$, then manager t can adopt the old technology without exerting effort. Similarly, if $T_t = n$, then manager t can adopt the new technology without exerting effort. As will become shortly clear, in this model a manager only wants to exert effort to try to invent the new technology. Therefore, without loss of generality, in our model we do not allow manager t to exert effort when the new technology is freely available at his time of entry (i.e. by assumption $E_t = 0$ when $T_t = n$).

[9] In other words, manager t is not allowed to adopt his first technology at time $t' > t$ $(t' = t + 1, ..., \overline{K})$. This assumption could best be defended by allowing all firms to produce and sell their goods from their time of entry on. However, this would imply that a manager's payoff would depend on his time of entry, an unnecessary complication.

[10] And with "initial" we mean the technology that manager t adopts at his time of entry.

Manager t must only decide whether to exert effort or not whenever $T_t = o$.

It is assumed that the new technology exists with a probability p $\in (0, 1)$ and will always be invented if it exists and if a manager exerts effort e. Assume $T_t = o$. If manager t doesn't exert effort, then he adopts o as his initial technology and $T_{t+1} = o$ $(\forall t < K)$. If manager t exerts effort, there are two possibilities: (i) with probability p he invents n and (ii) with probability $(1 - p)$ he doesn't invent n. In case (i), manager t adopts n as his initial technology (and uses n in the production phase). In case (ii) he adopts o and $T_{t+1} = o$ $(\forall t < K)$. If $T_t = n$, then manager t adopts n (without exerting effort) and uses n during the production phase.

We assume that n is a public good. This means that if, e.g. manager one invents n, then all his successors can freely (i.e. without paying manager one a "royalty fee") adopt the new technology without exerting effort. Similarly, if e.g. manager K invents the new technology, all his predecessors can freely switch their old technologies for the new one (without paying manager K a royalty fee and without exerting effort). This public good assumption is modeled by the following two assumptions:
(a) if manager t invents n, then $T_{t'} = n$ $(t' = t + 1, ..., K)$
(b) if manager t invents n, manager t's predecessors instantaneously switch their old technologies for the new one and use n during the production phase.

The profits of the firm of manager t, Π_t, are realised during the production phase. $\Pi_t \in \{\overline{\pi}(ns), \overline{\pi}(s), \underline{\pi}\}$. $\overline{\pi}(ns)$ $(\overline{\pi}(s))$ means that manager t is using n in the production phase and adopted n from scratch (switched from o to n). $\underline{\pi}$ means that manager t is using o in the production phase. For simplicity, we assume that $\overline{\pi}(ns) = 1 > \overline{\pi}(s) = \gamma > \underline{\pi} = 0$.

We model the managers' utilities as a v.N.M. utility function. Formally, manager t's payoff in the adoption phase, U_t, equals $\Pi_t - E_t$. Of course prior to making his effort decision, manager t doesn't necessarily know Π_t's realisation. To illustrate how expected utilities are written, consider manager one's effort decision. If he decides to exert effort his expected payoff equals $E(U_1|E_1 = e) = p - e$. If manager one exerts effort, with probability p there exists a new technology which he will find (for sure) and in the production phase this will give him a payoff equal to $1 - e$. However our diligent manager may be unlucky because with probability $(1 - p)$ there does not exist a new technology. In that case manager one must adopt o at time one. During the production phase his payoff then equals: $-e$.

In order to highlight the interesting features of our model, we suppose that:

A1: $p > e$.

Under assumption one all our (risk neutral) managers have - based on their priors - an individual incentive to exert effort at their time of entry. We also assume that:

A2: $p - e < p\gamma$

To understand A2, assume that manager one knows that manager two will exert effort for sure. Manager 1's ex ante payoff of free-riding (with "free-riding" we mean that manager 1 exerts no effort at time 1 and adopts o in the hope to switch from o to n at time 2) then equals $p\gamma$. This is easy to see: if manager 1 free-rides, with a prior probability p manager 2 will invent n. In that case manager 1 will switch from o to n and will get a profit of γ during the production phase. Hence A2 means that - due to relatively low switching costs - our managers rather imitate instead of innovate.

2.2 Two examples.

To fix ideas, we first detail an example of how payoffs may look like. Consider therefore the following table:

Table One: Illustration of payoffs when manager two invents n

manager	1	2	3	4	5	6	7
T_t	o	o	n	n	n	n	n
E_t	0	e	0	0	0	0	0
initial techn.	o	n	n	n	n	n	n
final techn.	n	n	n	n	n	n	n
payoffs	γ	$1-e$	1	1	1	1	1

In the example above seven managers enter sequentially in the emerging market.

Manager one doesn't exert effort and adopts the old technology at time one. At time two, manager two observes that $T_2 = o$ (or, equivalently, manager two observes manager one adopting the old technology), exerts effort and invents n (the example assumes that n exists). Manager two then adopts n as his initial technology and manager one switches his old technology for the new one. At time t' ($t' = 3, 4, ..., 7$), $T_{t'} = n$ and manager t' adopts n as his initial technology (without exerting effort). All managers thus use the new technology during the production phase (this is represented on the fifth line). However manager one did not adopt the new technology from scratch. Therefore, manager one incurs a switching cost and gets γ instead of one. Manager t' ($t' = 3, 4, ..., 7$) adopted n from scratch and did not exert effort. Therefore manager t' gets one in the production phase. Manager two also adopted n from scratch, but he also exerted effort. Hence, in the production phase manager two gets $1 - e$ instead of one.

Finally, we detail an example which should facilitate the interpretation of our general framework. Consider therefore the following quote:

> In many emerging industries the pressure ... is so great that ... problems are dealt with expediently ... Confronted with the need to set a pricing schedule, for example, one firm adopts a two-tiered price that the marketing manager used in his previous firm, and the other firms in the industry imitate for lack of a ready alternative. (Porter, 1980,p.219)

One can interpret Porter's example along the lines of the general framework we set out above. Call NLP (=a more elaborate non-linear pricing schedule) the best alternative to TTP (=two-tiered price). In our model TTP can be thought of as the "old" technology, while NLP represents the "new" one. As NLP is a fairly unknown pricing policy its adoption is hindered by two different types of uncertainty: *profit* uncertainty and *development* uncertainty. With profit uncertainty we mean that no manager knows whether NLP is more profitable than TTP. With development uncertainty we mean that no manager knows how NLP really looks like. As NLP represents a fairly unknown pricing policy managers must first think about how to design the best possible alternative to PPT (this typically will require developing market surveys, negotiating with input suppliers, etc...). Both kinds of uncertainties are captured by my payoffs. If manager one doesn't want to adopt TTP, he must first incur a development cost equal to e. We assume that immediately after having incurred the development cost, a manager learns about its profitability characteristics[11]. There are then two possibilities: (i) NLP is more profitable than PPT (i.e. yields a revenue equal to one) or (ii) NLP is less profitable than PPT (i.e. yields a negative revenue). We assume that (i) prevails with probability p. In case (i) the innovating manager adops NLP as his pricing policy. In case (ii) the manager who incurred the development cost adopts PPT as his pricing policy. Hence, a manager who incurs the development cost gets an expected payoff equal to $p-e$. As pricing policies cannot be patented, it is obvious that any superior pricing schedule will be imitated by other firms (as Porter's example suggests). Therefore, if manager one incurs the development cost, manager two gets, in expected terms, p. Similarly, if manager one free-rides on manager two's development costs, he gets, ex ante, $p\gamma$ (where the switching costs can be thought of as menu costs, psychological costs due to severing existing relationships with the firm's customers and/or its input suppliers, etc...).

At this point there are two possibilities: (i) $p\gamma > e$ or (ii) $p\gamma \leq e$. Inequality (ii) implies that managers only want to incur the development cost at their time of entry. In our general framework we assumed that managers can only exert effort at their time of entry. Therefore our model is appropriate to study "technology" adoption in emerging markets when inequality (ii) holds. However, we don't believe our model to be inappropriate when inequality (i) holds. If $p\gamma > e$, managers have an incentive to exert effort not only at their time of entry but also in future periods. Analysing this kind of game is obviously hard. If we understand what happens in a simpler context where managers only exert effort at their time of entry, this should already help us in analysing the equilibrium outcome of this game with an enriched strategy space.

3 Efficiency analysis

In this section we introduce a social planner in our model who has the authority to oblige any manager to exert effort at his time of entry. She uses her authority to maximise the sum of the (ex ante) utilities of the K managers (=social welfare).

We define $q_t \in [0, 1]$ as the instrument of the social planner: it is the probability

[11] Alternatively, we could assume that a manager only learns about the profitability of NLP after having "adopted" it during a certain period. However, a manager's payoff should then depend on the amount of time he may have adopted a suboptimal "technology". This payoff structure would only complicate matters and would not affect our main result.

with which the social planner obliges manager t to exert effort if $T_t = o$ (if $T_t = n$, by assumption manager t exerts no effort). $\tilde{q}_t$ denotes the first best probability: it is the probability with which, if $T_t = o$, manager t must exert effort to maximise ex ante social welfare. A strategy profile (Q) is defined as $(q_1, q_2, ..., q_K)$. $Q_{-t} = (q_1, q_2, ..., q_{t-1}, q_{t+1}, ..., q_K)$. $\tilde{Q} = (\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_K)$. $E(U_t|Q)$ denotes the ex ante utility of manager t given q_t and Q_{-t}. The social planner's maximisation problem can be written as:

$$\max_Q W(Q) = \sum_t E(U_t|Q) \qquad s.t. \qquad q_t \in [0, 1]$$

To avoid any possible confusion we first describe the sequencing of events:

0) At time zero: nature decides whether n exists or not.
1) At time $t = 1, ..., K$:
(i) Manager t observes $(T_1, T_2, ..., T_t)$[12]
(ii) If $T_t = o$, manager t exerts effort with probability q_t (determined by the social planner), otherwise he exerts no effort and adopts n.
(iii) If manager t successfully invented n in (ii), $\forall t < K$ $T_{t+1} = n$, manager t adopts n and all his predecessors switch to n. If manager t did not invent n, $T_{t+1} = o$ and manager t adopts o.
2) At time $t = K + 1, ..., \overline{K}$: managers receive their payoffs and the game ends.

Proposition 1 $\tilde{Q} = (1, 0, 0, ..., 0)$[13]

Proof: See Appendix one

Proposition (1) shows that two conditions must be satisfied in order to obtain ex ante efficiency: (i) only one manager must exert effort and (ii) manager one should exert effort. Both conditions are intuitive. By assumption n is invented if it exists and if a manager exerts effort e. The inexistence of n is thus proven as soon as a manager exerts effort and adopts o. Hence, to avoid costly duplication of effort a social planner will never oblige more than one manager to exert effort. By assumption adopting n from scratch is more efficient than first adopting o and

[12]This assumption is equivalent to the one present in the standard herding models where it is assumed that each player observes her predecessors' actions. For example, if manager seven observes $(T_1 = o, T_2 = o, T_3 = o, T_4 = o, T_5 = n, T_6 = n, T_7 = n)$, then manager seven knows that manager one up to manager four adopted the old technology and that managers five and six adopted the new one.

[13]$\tilde{Q}$ does not depend on our informational assumptions. If the social planner could also make q_t contingent on $(E_1, E_2, ..., E_{t-1})$ this wouldn't affect $\tilde{Q}$. See Appendix one for a formal proof.

then switching to n. Hence to avoid the occurence of switching costs manager one should exert effort.

The following corollary shows that we can efficiency-rank different strategy profiles.

Corollary 1 $W((0,0,...,0,1)) < W((0,0,...,0,1,0)) < ... < W((0,1,0,...,0)) < W((1,0,...,0))$

Proof: $W(\text{only manager } t \text{ exerts effort for sure}) = p[(t-1)\gamma + (K-t+1)] - e$. As $\gamma < 1$, it is immediate that the rhs of our last equation is decreasing in t.Q.E.D.

Corollary (1) is intuitive: the later a manager exerts effort, the later the innovation may occur in the entry sequence, the more managers who will then switch from o to n and who will then incur the inefficient switching cost $(1-\gamma)$.

4 Observable effort levels.

In this section we compute equilibrium strategies in our game in the absence of a social planner and if all managers observe their predecessors' effort levels on top of the available technology. For reasons that will become clear in our next section, we call this environment a non-herding one.

Call $h^{E_t} = (E_1, E_2, ..., E_{t-1})$ the history of past effort levels prior to manager t's effort decision. H^{E_t} denotes the set of all possible past effort levels (prior to manager t's effort level). $H^{E_1} = \emptyset$. A behavioural strategy for manager t is defined as a function $q_t : H^{E_t} \to [0,1]$ with the interpretation that $q_t(h^{E_t})$ denotes the probability with which manager t is going to exert effort if $T_t = o$ and given h^{E_t}. A subgame perfect equilibrium (SPE) is a $Q^s = (q_1^s(\cdot), q_2^s(\cdot), ..., q_K^s(\cdot))$ s.t. manager t cannot gain by choosing a $q_t(\cdot)$ different from $q_t^s(\cdot)$, given $(q_1^s(\cdot), q_2^s(\cdot), ..., q_{t-1}^s(\cdot), q_{t+1}^s(\cdot), ..., q_K^s(\cdot))$, $\forall t$, $\forall h^{E_t} \in H^{E_t}$.

As already explained above, manager t is sure that n doesn't exist if one of his predecessors exerted effort and if $T_t = o$. Hence, $\forall h^{E_t} \neq (0,0,...,0), q_t^s(h^{E_t}) = 0$. Therefore to fully describe our SPE we only need to compute the K different $q_t^s((0,...,0))$'s. Therefore in this section and with a slight abuse of notation we simply rewrite $q_t((0,...,0))$ as q_t. Manager t is said to be active, if $q_t^s > 0$. Manager t is said to follow the free-riding strategy if $t < K$ and if $q_t^s = 0$. Stated differently, manager t follows the free-riding strategy if he doesn't exert effort (not even with a very small probability) in the hope to free-ride on the effort level of one of his successors.

In this case we obtain the following result:

PROPOSITION 2 *If manager t observes* $(T_1, T_2, ..., T_t)$ *as well as the effort levels of all his predecessors,* $Q^s = (0, 0, ..., 0, 1)$ *constitutes the unique SPE.*

Proof: see Appendix one.

To understand the intuition behind proposition (2) let's focus on the two-manager case. Consider the following graph.

[Insert here Graph One]

Obviously, Graph one does not represent a complete extensive-form representation of the two-manager case. For example, in a complete extensive-form representation one should allow nature to move first. In Graph one nature only moves if manager one exerts effort. As usual, nature is represented as a non-strategic "player": if player one exerts effort, then with probability p, nature decides "to go right" (and to reach node (c)) and with probability $1 - p$, nature decides "to go left" (and to reach node (b)). Moreover in Graph one payoffs are sometimes written in conditional expected terms. For instance, $(p\gamma, p - e)$ represents the expected payoff manager one and two get, conditional on $E_1 = 0$ and $E_2 = e$. Graph one actually represents a simplified extensive-form representation which allows us to show in a more transparent way how equilibrium behaviour changes when comparing an environment with observable effort levels to another one with unobservable effort levels.

Manager one must decide whether to exert effort or not. If he exerts effort, ex ante he gets $p - e$. Manager one doesn't exert effort if $p - e < E(U_1|E_1 = 0)$. Suppose manager one doesn't exert effort and adopts o as his initial technology. At time two, manager two receives "bad news" in the sense that he observes that $T_2 = o$. Manager two knows that this observation can be explained in two different ways: (i) manager one did not exert effort and adopted the old technology in the hope to free-ride on the effort level of the second manager and (ii) manager one exerted effort but the new technology doesn't exist. In Graph one explanation (i) is represented by node (a). Explanation (ii) is represented by node (b). Node (c) represents the case in which $T_2 = n$. In this section it is assumed that manager two also observes manager one's effort level. Therefore, if manager one doesn't exert effort, manager two knows that the game has reached node (a). Therefore manager two knows that if he exerts effort, he gets $p - e$. If manager two doesn't exert effort, he gets zero (manager two knows that no other manager is entering the market after him, so he knows that he's the only manager who can try to invent the new technology). Therefore, if manager one doesn't exert effort, it is optimal for manager two to do so. Manager one knows this. Therefore, $E(U_1|E_1 = 0) = p\gamma$. By A2, $p - e < p\gamma$ and in equilibrium manager one doesn't exert effort and adopts the old technology in the hope to switch to the new one at time two.

Crucial in our two-manager example is that with observable effort levels, manager two can never wrongly attribute $T_2 = o$ to explanation (i) or (ii). Manager one acknowledges this and as he rather imitates (due to A2) he doesn't exert effort in the hope to free-ride on manager two's effort level. The reader can easily understand that this intuition remains valid in a world with K managers (manager K can never wrongly attribute $T_K = o$ to the explanation "no one exerted effort" or to the explanation "at least one manager exerted effort but n doesn't exist"). Note that $Q^s = (0, 1)$ is not the most efficient strategy profile because in case manager two invents the new technology, manager one incurs an inefficient switching cost of $(1 - \gamma)$. As we have shown in Corollary (1), with only one manager who exerts effort for sure, no strategy profile is doing worse in terms of ex ante efficiency than the unique SPE we obtain with observable effort levels.

Proposition (2) is not a very interesting one: it merely states that in the presence of a public-good problem, players (managers) follow inefficient strategies in the hope to free-ride on someone else's effort level. Clearly this is an unsurprising result. Already David Hume (1739) was aware of the fact that in the presence of a public good the economic outcome need not be efficient. Similarly, Holmström (1982) showed that in a moral-hazard-in-teams set-up (which can also be seen as a public-good problem), agents underprovide effort in equilibrium.

5 Unobservable effort levels.

In this section we consider the same game as the one we analysed in our previous section except that now our managers only observe the evolution of the available technology and not one another's effort levels. We refer to this environment as a herding one.

q_t denotes the strategy of manager t. It represents the probability with which manager t is going to exert effort given that $T_t = o$. As effort levels are unobservable this is a dynamic game of imperfect information. Each manager - upon observing his predecessors adopting the old technology - must have a belief concerning which node in the game tree was reached. If $T_t = o$, then manager t knows that $X_t = \sum_{i=0}^{t-1} C_{t-1}^i$ nodes are present in his information set (where C_{t-1}^i denotes the number of different combinations of i managers out of $t-1$ managers[14]). β_t denotes a $(1 \times X_t)$ vector where element j represents the probability that node j was reached when $T_t = o$. A perfect Bayesian equilibrium (PBE) is a (Q^*, β^*) (where $\beta^* = (\beta_1^*, \beta_2^*, ..., \beta_K^*)$) such that:
(i) manager t cannot gain by choosing a q_t different from q_t^*, given β_t^* and $(q_1^*, q_2^*, ..., q_{t-1}^*, q_{t+1}^*, ..., q_K^*)$ $\forall t$,
(ii) β_t^* is computed via Bayes' law, given $(q_1^*, q_2^*, ..., q_{t-1}^*)$, $\forall t$.

A manager is said to be active if $q_t^* > 0$. Before stating and proving our most important result (which is summarised in proposition four), we first focus on a simple PBE which illustrates how herding may increase efficiency in our model.

5.1 An equilibrium with one active manager.

Our main result with one active manager is summarised below:

PROPOSITION 3 *With unobservable effort levels:*
1) For every manager, there exists a unique PBE in which he exerts effort for sure and in which the other managers are inactive.
2) With only one active manager, a herding environment can never perform worse (in terms of efficiency) than a non-herding one.

Proof: In the next paragraphs we prove the proposition for the case where $K = 2$. The proof when $K > 2$ appears in the appendix.

To understand the intuition behind proposition (3) it is useful to go back to the two-manager case. Consider the graph below:

[14] For example, manager three knows that $T_3 = o$ can be explained in four different ways: (i) no one exerted effort, (ii) only manager one exerted effort (but the new technology doesn't exist), (iii) only manager two exerted effort (but n doesn't exist) and (iv) both manager one and two exerted effort (but n doesn't exist). And $4 = C_2^0 + C_2^1 + C_2^2$.

[Insert here Graph Two]

Graph two is identical to Graph one, except that now (a) and (b) lie in the same information set (i.e. manager two doesn't know if $T_2 = o$ because manager one wants to free-ride on his effort level or because the new technology doesn't exist). As usual, in every PBE manager two must have a belief concerning which node in the game tree was reached, and this belief must be consistent with equilibrium behaviour and vice versa. x denotes the probability that the game has reached node (a) conditional on $T_2 = o$. Suppose manager two puts $x = 0$ whenever $T_2 = o$. Given this belief, it's optimal for manager two not to exert effort and adopt the old technology because $E(U_2|E_2 = e, T_2 = o, x = 0) = -e < E(U_2|E_2 = 0, T_2 = o, x = 0) = 0$.

Manager one knows that manager two puts $x = 0$ whenever $T_2 = o$. If manager one deviates and doesn't exert effort, then manager two would wrongly attribute $T_2 = o$ to the explanation "manager one exerted effort but the new technology doesn't exist". In that case manager two doesn't exert effort (and manager one knows this). Therefore $E(U_1|E_1 = 0) = 0 < p-e$, and it's optimal for manager one to exert effort with probability one. Hence, there exists a PBE in which manager one (and only manager one) exerts effort for sure (and this strategy profile is the most efficient one).

In our introduction we defined herding as a behaviour where one person, after observing the action(s) of her predecessor(s) (and after updating her prior beliefs), has more incentives to imitate her predecessor(s). Note that in the equilibrium detailed in proposition (3), manager two's behaviour complies to this definition. Upon observing $T_2 = o$, manager two computes:

$$p_2 = \Pr(n \text{ exists}|Q^*, T_2 = o) = \frac{(1-q_1^*)}{(1-q_1^*) + q_1^*(1-p)}p$$

in which he replaces q_1^* by one. Hence, $p_2 = 0 < e$. In the beginning of the game manager two knew that the new technology existed with probability p. However, after observing manager one adopting the old technology (and after updating his prior beliefs) he puts zero probability on the event that the new technology exists, and adopts the same technology as the one adopted by manager one.[15] In this

[15]This behaviour is identical to the one followed by all players inside an "informational cascade".

sense, manager two herds whenever manager one adopts the old technology. In the equilibrium highlighted in proposition (3), and in contrast with the existing herding literature, this herding is actually good for efficiency reasons because it obliges manager one to follow a more efficient strategy.

Of course this is not the unique PBE with one active manager. For example, there also exists a PBE in which $x = 1$ and in which $Q^* = (0, 1)$. Point 2) is easy to prove. From Corollary (1)we know that:

$$W(Q^* = (1, 0)) > W(Q^* = (0, 1)) = W(Q^s = (0, 1))$$

Q.E.D.

There is little reason to assume that an equilibrium with only one active manager constitutes a natural "focal point" of our game with unobservable effort levels. We started by focusing on this kind of PBE's for pedagogical reasons only. However, in our next subsection we will see that the insights we get out of proposition (3) remain to a large extent valid if we focus on a PBE with more than one active manager.

5.2 Equilibria with N ($1 \leq N \leq K$) active managers.

In this section we provide the reader with a complete analysis of all the PBE's which exist in our game with unobservable effort levels. In our previous subsection we proved the existence of K equilibria in which only one manager exerts effort. Therefore, we now work under the assumption that $1 < N \leq K$.

Without loss of generality, we assume it are the first N managers who are active. With N managers, all managers compute their q_t^*'s out of the following set of $N + (N-1)$ nonlinear simultaneous equations. The first N equations merely state that, as long $T_t = o$, manager t ($t \leq N$) must be indifferent between exerting effort or not. In the equations below p_j denotes $\Pr(n \text{ exists}|Q^*, T_j = o)$:

$$p - e = q_2^* p\gamma + (1 - q_2^*)q_3^* p\gamma + ... + \prod_{j=2}^{N-1}(1 - q_k^*)q_N^* p\gamma$$

$$p_2 - e = q_3^* p_2\gamma + (1 - q_3^*)q_4^* p_2\gamma + ... + \prod_{j=3}^{N-1}(1 - q_k^*)q_N^* p_2\gamma$$

$$\vdots$$

$$p_{N-1} - e = q_N^* p_{N-1} \gamma$$

$$p_N = e$$

For instance, the second equation states that manager two must be indifferent between exerting effort or not when $T_2 = o$. The lhs of the second equation represents manager two's gain of exerting effort conditional on $T_2 = o$. The rhs of the second equation represents manager two's gain of not exerting effort conditional on $T_2 = o$ (if manager two doesn't exert effort with probability $q_3^* p_2$ manager three will invent n (in which case manager two gets γ), with probability $(1 - q_3^*) q_4^* p_2$ manager four will invent n (in which case manager two gets γ), etc...). The remaining $(N - 1)$ equations determine the posteriors of manager two to manager N (since $p_1 = p$, the "posterior" of the first manager is exogenously given and need not be endogenously computed). These equations can be summarised as:

$$p_j = \frac{\prod_{k=1}^{j-1}(1 - q_k^*)}{\prod_{k=1}^{j-1}(1 - q_k^*)p + (1 - p)} p \qquad \forall j = 2, ..., N$$

We can now state our next (and most important) proposition:

Proposition 4 *With unobservable effort levels:*
1) For every set of managers there exists a unique PBE in which they are all active.
2) Depending on the values of the parameters and on the selected equilibrium, a herding environment may be more efficient than a non-herding one.

Proof: The proof of point 1) is divided in two parts. If the set of active managers is a singleton, then the proof appears in our previous subsection. If the set of active managers contains more than one manager, see Appendix two. The proof of point 2) appears below.

What is the intuition behind point 1) of proposition (4)? Which "forces" influence the equilibrium effort level of each active manager?

Before answering these questions we first introduce some new notations. z_t denotes the probability with which manager t does not exert effort (i.e. $z_t = 1 - q_t$). Z_{j+1} denotes the probability that no manager moving after manager j ($j = 2, ..., N$) will search for the new technology (i.e. $\forall j = 2, ..., N-1$, $Z_{j+1} = z_{j+1}.z_{j+2}...z_N$ and if $j = N$, $Z_{N+1} = 1$).

Suppose the first N managers are active. Let's focus on two managers: manager $j-1$ and manager j (where $1 < j < N$). We know that if $T_{j-1} = o$:

$$p_{j-1} - e = q_j^* p_{j-1}\gamma + z_j^* q_{j+1}^* p_{j-1}\gamma + ... + \prod_{k=j}^{N-1} z_k^* q_N^* p_{j-1}\gamma \tag{1}$$

This last equality can be written as: $p_{j-1} - e = (1 - Z_j^*)p_{j-1}\gamma$. By definition, $Z_j = z_j Z_{j+1}$. Therefore we can rewrite equation (1) as:

$$p_{j-1} - e = (1 - z_j^* Z_{j+1}^*)p_{j-1}\gamma \tag{2}$$

This last equation shows that manager j's equilibrium behaviour can be summarised by the following rule: "manager j takes p_{j-1} and Z_{j+1}^* as given and chooses q_j^* such that manager $j-1$ is indifferent between exerting effort or not".[16] This simple rule also provides a partial intuition why, once we fix the set of active managers, there exists a unique vector of equilibrium probabilities. Z_{j+1}^* is fixed such that manager j is indifferent. Since $q_{j-1}^* > 0$ (otherwise manager $j-1$ is not active), $p_{j-1} > p_j$. This implies that:

$$p_{j-1} - e > p_j - e = (1 - Z_{j+1}^*)p_j\gamma$$

Hence, if $z_j = 1$ ($q_j = 0$), manager $j-1$ strictly prefers to exert effort. If $z_j = 0$ ($q_j = 1$), manager $j-1$ strictly prefers to wait because $p_{j-1} - e < p_{j-1}\gamma$. By monotonicity, there exists a unique q_j^* such that manager $j-1$ is indifferent (and this reasoning is valid $\forall j \in \{2, ..., N\}$).

We can rewrite equation (2) as:

$$z_j^* = \frac{1}{Z_{j+1}^*}[C + \frac{e}{p_{j-1}\gamma}] \tag{3}$$

(where $C = (1 - \frac{1}{\gamma})$). Equation (3) shows that q_j^* is influenced by two opposing effects (or "forces"). The first effect, which we call the *information externality effect*,

[16]The intuition behind manager one's behaviour is more complicated and is not explained in this paper. Somewhat surprisingly, it turns out that manager one chooses q_1 such that manager N is indifferent. The intuition why manager one is able to make the last manager indifferent involves a lengthy and, in our opinion, not very interesting discussion. For a mathematical explanation see proposition (6) in appendix two.

is captured by the term p_{j-1}. q_j^* is increasing in p_{j-1}: the more one advances in the queue, the higher the sum of the expected efforts spent by all previous managers, the lower the posterior of manager $j-1$, the lower the probability with which manager j must exert effort to make manager $j-1$ indifferent (ceteris paribus). The information externality also gives birth to a countervailing force which we call the *herd effect*. This herd effect is captured by the term Z_{j+1}^*. q_j^* is increasing in Z_{j+1}^*. This is also logical: the more one advances in the queue, the lower the probability that manager $j-1$ can hope to free-ride on the effort of a subsequent manager (i.e. the higher Z_{j+1}^*), the higher the probability with which manager j should exert a high effort to make manager $j-1$ indifferent (ceteris paribus).

Shortly said, the information externality effect states that if manager j observes more managers who adopted the old technology, then he exerts effort with a lower probability (i.e. the higher manager j's incentives to adopt the same technology as the one adopted by his predecessors). Obviously this is not an original effect. This effect has already been extensively studied by Banerjee (1992), BHW (1992), Chamley and Gale (1994), Chamley (1997), Zhang (1997),.... The herd effect basically states that a player's incentives to exert effort increases due to future herding. This effect is also not original. As argued in the introduction, Chamley (1997) and Zhang (1997) already showed how the occurence of an informational cascade adversely affects a player's gain of waiting and thereby induces her to invest[17].

Both effects highlight how q_j^* varies with j. The reader should bear in mind that this paper is intended to show how herding, by reducing free-rider problems, may increase efficiency. Therefore, the originality of this paper is based on our finding that in the herding environment q_j^* may be strictly positive when $j < K$ (and not on how q_j^* varies with j). Therefore, the fact that both effects are present in other papers does not bother us.[18]

Next, to prove point 2), we detail two numerical examples which compare the ex ante efficiency of a herding versus a non-herding environment.

[17] As mentioned in the introduction, the herd effect is not present in Hendrick and Kovenock's (1989) model. This is because they only work with a two-period model. In their model both firms don't care about each other's second-period posteriors because (by construction) they cannot free-ride on each other's second-period drilling decision.

[18] However, we want to mention that this model incorporates both effects in a more elegant way (none of the two papers mentioned above were able to synthesize the interplay between both effects in one "summary equation").

Example 1: $K = 3; p = 0.9; e = 0.65$ and $\gamma = 0.7$
In this example $q_1^* \simeq 0.79$ and $q_2^* = 0.4$.

$$W((0.79, 0.4, 0)) = q_1^* p[3 - e] + q_1^* q_2^* (1-p)[-2e] + q_1^*(1-q_2^*)(1-p)[-e] +$$

$$(1 - q_1^*) q_2^* p[2 + \gamma - e] + (1 - q_1^*) q_2^* (1-p)[-e] + (1-q_1^*)(1-q_2^*)[0] \tag{4}$$

The terms between square brackets represent conditional aggregate payoffs. For example $[3 - e]$ represents the sum of the payoffs of the three players if manager one invents n (which happens with probability $q_1^* p$).

$$W((0, 0, 1)) = p[2\gamma + 1] - e \tag{5}$$

If we substitute all the (exogenous and endogenous) variables in (4) and (5) by their numerical counterparts, we see that $W((0.79, 0.4, 0)) \simeq 1,75 > W((0, 0, 1)) = 1,51$. Hence in this example a herding environment increases ex ante social welfare.

The intuition goes as follows: introducing herding (i.e. assuming non-observable effort levels) in our public-good problem has efficiency-increasing as well as efficiency-decreasing properties. On the one hand herding induces the first two players to exert effort which is good for efficiency. On the other hand a herding environment may also be undesirable from an efficiency point of view for the following two reasons: (i) in a herding environment both managers may inefficiently exert effort (this happens with an ex ante probability of $q_1^* q_2^*(1-p)$) and (ii) all players may end up using a suboptimal technology (this happens with an ex ante probability $(1-q_1^*)(1-q_2^*)p$). Point (ii) is not a surprising one: this efficiency-decreasing consequence of herding has already been extensively studied in the herding literature (see, e.g., Banerjee (1992), BHW (1992), Chamley and Gale (1994), Hendricks and Kovenock (1989), Scharfstein and Stein (1990), Zwiebel (1995), ...).

In our first example q_1^* is relatively high. This increases the efficiency-enhancing effect of herding and decreases $(1-q_1^*)(1-q_2^*)p$. Therefore a herding environment is a more efficient environment than a non-herding one. Note also that in our first example the duplication-of-effort problem is relatively small. If we could overcome that problem $W((0.79, 0.4, 0))$ would only increase by $2q_1^* q_2^*(1-p)e$, i.e. by 0.041. The next example shows that a herding environment may also be an ex ante more inefficient environment than a non-herding one.

Example 2: $K = 3; p = 0.7; e = 0.65$ and $\gamma = 0.7$
Here $q_1^* \simeq 0.2$ and $q_2^* \simeq 0.1$ and $W((0.2, 0.1, 0)) \simeq 0.39 < W((0, 0, 1)) = 1.03$.

The intuition mirrors the one we explained in our previous example: in this case q_1^* and q_2^* are very low. This decreases the efficiency-enhancing consequence of herding and increases $(1 - q_1^*)(1 - q_2^*)p$.

6 A discussion of our model and its results.

6.1 On the existence of multiple equilibria.

The aim of this paper is to show how herding, by reducing free-rider problems, may increase efficiency. We don't need to focus on a particular PBE to get this insight! In our model a herding environment may increase efficiency as soon as $t < N$ and $q_t^* > 0$. All our PBE's with more than one active manager share thus in common that some manager(s), due to subsequent herding, will internalise his (their) information externality(ies) by exerting effort (and this favours efficiency). Our main insight does thus not depend on the selected equilibrium and in that respect the fact that our game possesses a potentially large number of equilibria is not a big problem.

6.2 How robust is our main insight?

The main contribution of this paper can be divided in two different parts: (i) information externalities induce early entrants to exert effort and (ii) if, for efficiency reasons, one wants to avoid delay, a herding environment may be more efficient than a non-herding one. We now discuss the robustness of this insight. We first argue how our main insight may be recovered in an endogenous queue context. Next we tackle the issue of incomplete information. Finally, we discuss to what extent our results rely on the absence of a discount factor.

6.2.1 The endogenous queue case.

Following Banerjee (1992) and BHW (1992) we worked with an "exogenous queue" in the sense that managers only exert effort at their time of entry. One may believe that our main result crucially depends on our exogenous queue assumption (and on the assumption that $\gamma < 1$). To understand this point, let's analyse the following "endogenous queue" case: assume K firms produce their goods in a market using an old technology. At time one, due to exogenous technological reasons, all managers simultaneously realise that, with probability p, a new technology exists. Assume $\gamma = 1$ and δ (the discount rate) < 1. In the non-herding environment all players play a war of attrition to determine who is going to exert effort in order to invent the public good. As all players observe one another's effort levels and technologies, this is a war of attrition without information externalities. As usual, in the symmetric equilibrium, all players at time one must be indifferent between exerting effort or not. Therefore, ex ante everyone gets $p - e$. In the herding environment all players play a war of attrition with information externalities. As before, in the symmetric equilibrium, at time one everyone must be indifferent between exerting effort or not. Therefore, ex ante everyone also gets $p - e$. Hence, one may believe our main result not to be robust in the endogenous queue case. In our emerging market set-up we argued that - due to switching costs - efficiency requires early adoption of the new technology. However, even in an endogenous queue context one can easily find a reason why, for efficiency reasons, one wants to avoid delay.

For example, take the same set-up as the one we detailed in the previous paragraph. Assume our K firms use a technology to manufacture a consumption good. If our K firms use the old technology then consumer surplus equals zero. If our K firms use the new technology to manufacture the consumption good, then - due for instance to a reduction in costs which allow for lower prices - consumer surplus equals $CS > 0$. Consumers are impatient and discount the future at the rate δ. In this set-up firms are indifferent between a non-herding environment and a herding one, because, as argued above, they get the same payoff ($= p - e$) in both environments. However, due to the herd effect, the new technology may be invented at an earlier date (in expected terms) in the herding environment than in the non-herding one. Therefore discounted expected consumer surplus may be higher in the herding environment than in the non-herding one. In that case the herding environment is, ex ante, more efficient than the non-herding one. We believe the endogenous queue case to be an interesting topic for future research.

6.2.2 The incomplete information case.

In our model no manager possesses some private information concerning the existence of the new technology. However, the example below suggests that our result can also hold in an incomplete information set-up.

Assume that $K = 2$ and that $\Pr(n \text{ exists}) = \frac{1}{2}$. Both managers receive a conditionally independent signal, s_t $(t = 1, 2)$, concerning the existence of n. The signal can either be "high" or "low". Formally, $\Pr(s_t = \text{high}|n \text{ exists}) = \Pr(s_t = \text{low}|n$ does not exist$) = p > \frac{1}{2}$. Assume that:

A1': $\frac{(1-p)^2}{p^2+(1-p)^2} < e$

Assumption A1' states that even if both managers revceive a low signal, still each manager faces an individual incentive to exert effort because of a low effort cost. Assume assumption A2 holds (i.e. managers prefer to imitate instead of innovate). The non-herding environment corresponds to the case in which effort levels are observable and all managers share their private information. In the herding environment effort levels are unobservable. It should be obvious to the reader that our non-herding environment is characterised by a unique SPE in which only the last manager exerts effort for sure. For the same reason as the one we explained above it should be obvious that our herding environment is characterised by a PBE in which only manager one exerts effort.

6.2.3 On the absence of a discount factor.

Clearly, the introduction of a discount factor would affect proposition (2). If managers discount the future, early movers may not find it profitable to wait and (hope to) free-ride on the last manager's effort level. However, with observable effort levels, our unique[19] SPE would, most probably, still feature some (inefficient) strategic waiting. For example, in the presence of a "moderate" discount factor we may find an equilibrium in which the first two managers don't exert effort, the third manager exerts effort with probability one, and managers four, five and six don't exert effort. But a discount factor also discourages managers to follow the free-riding strategy in the herding environment! Moreover proposition (3) does not depend on the absence of a discount factor; even with very low interest rates there still exists a PBE in which only the first manager exerts effort for sure. Therefore, in our opinion, our main result does not hinge on the absence of an interest rate in our model.

[19]With observable effort levels our game belongs to the class of games with complete and perfect information. The equilibrium in this class of games is computed by backwards induction and is always unique (independently of the presence or absence of a discount factor).

7 Conclusions.

In this paper we introduced a public good in a variant of the standard herding models à la Banerjee (1992) and BHW (1992). In our model players care about their successor's actions because if someone invents the new technology, previous adopters of the old technology can freely switch to the new one and enjoy a higher payoff. We showed how a herding environment, by reducing free-rider problems, may be more efficient than a non-herding one. We therefore developed a complete information model in which managers move sequentially. We argued that our model is appropriate to analyse herding in emerging markets. However we also believe that our main insight can be applied to other public good problems. For instance the hold-up problem (or more generally problems due to moral hazard in teams) may become less severe in a herding environment because individuals realise that if they don't work enough this may reveal some bad information and this will discourage others from working hard too. In that case our theory can also be applied to the field of (complete and incomplete) contract theory. Similarly, organisations may structure themselves such as to introduce information externalities in their organisations which may induce some people to work. We believe all this to be interesting topics for future research.

References:

Avery, Christopher and Zemsky, Peter "Multi-Dimensional Uncertainty and Herd Behavior in Financial Markets" *American Economic Review*, September 1998, 88(4), 724-748

Banerjee, Abhijit "A Simple Model of Herd Behavior" *The Quarterly Journal of Economics*, August 1992, 57(3), 797-817

Bikhchandani, Sushil, Hirshleifer, David and Welch, Ivo "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades" *Journal of Political Economy*, October 1992, 100(5), 992-1026

Caminal, Ramon and Vives, Xavier "Why market shares matter: an information-based theory" *Rand Journal of Economics*, Summer 1996, 27(2), 221-239

Chamley, Christophe "Social Learning, Delays, and Multiple Equilibria" *Unpublished Working Paper, EHESS*, December 1997

Chamley, Christophe and Gale, Douglas "Information Revelation and Strategic delay in a Model of Investment" *Econometrica*, September 1994, 62(5), 1065-1085

Cripps, Martin, Keller, Godfrey and Rady, Sven "Strategic Experimentation with Public or Private Information" *Mimeo*, March 2000

Hendricks, Kenneth and Kovenock, Dan "Asymmetric Information, Information Externalities, and Efficiency: the Case of Oil Exploration" *Rand Journal of Economics*, Summer 1989, 20(2), 164-182

Holmström, B. "Moral Hazard in Teams" *Bell Journal of Economics*, 1982, 13, 423-40

Hume, David "A Treatise of Human Nature" 1739, London: J.Noon

Scharfstein, David and Stein, Jeremy "Herd Behavior and Investment" *American Economic Review*, June 1990, 80(3), 465-479

Porter, Michael E. "Competitive Strategy, Techniques for Analyzing Industries and Competitors" 1980, New York: The Free Press

Vives, Xavier "How fast do rational agents learn?" *Review of Economic Studies*, April 1993, 60(2), 329-347

Zhang, Jianbo "Strategic Delay and the Onset of Investment Cascades" *Rand Journal of Economics*, 28(1), Spring 1997, 188-205

Zwiebel, Jeffrey "Corporate Conservatism and Relative Compensation" *Journal of Political Economy*, 1995, 103(1), 1-25

Appendix one
Proof of proposition (1):

We must compute $\tilde{Q}$ when the social planner can make q_t only contingent on $(T_1, ..., T_t)$ and when q_t can also be made contingent on past effort levels. We first state and prove the following two lemmas.

LEMMA 1 *Assume the social planner can make q_t only contingent on $(T_1, ..., T_t)$. Assume there exists at least one j $(j = 2, ..., K)$ such that $q_j = 1$. Then, $W((q_1, ..., q_{j-2}, q_{j-1}, 1, q_{j+1}, ..., q_K)) \leq W((q_1, ..., q_{j-2}, 1, 0, q_{j+1}, ..., q_K))$, where the inequality becomes strict whenever $j = 2$.*

Proof: Consider the two following strategy profiles:

$$Q_1 = (q_1, ..., q_{j-2}, q_{j-1}, 1, q_{j+1}, ..., q_K),$$

$$Q_2 = (q_1, ..., q_{j-2}, 1, 0, q_{j+1}, ..., q_K).$$

Q_2 is identical to Q_1 except for q_{j-1} and q_j. Define z_t as $z_t = 1 - q_t$. The expected utility of manager l $(l = 1, ..., j-2)$ under strategy profile Q_m $(m = 1, 2)$ equals:

$$E(U_l|Q_m) = \Pr(T_l = n|Q_m) + \Pr(T_l = o|Q_m)[q_l(p_l - e) + (1 - q_l)p_l\gamma].$$

Observe that $\Pr(T_l = n|Q_1) = p(1 - z_1...z_{l-1}) = \Pr(T_l = n|Q_2)$ and that $p_l = \Pr(n \text{ exists}|T_l = o, Q_1) = \frac{z_1...z_{l-1}}{z_1...z_{l-1}p+(1-p)}p = \Pr(n \text{ exists}|T_l = o, Q_2)$. Hence manager l is indifferent between Q_1 and Q_2.

The expected utility of manager l $(l = j+1, ..., K)$ under strategy profile Q_m equals:

$$E(U_l|Q_1) = p + (1-p)q_l(-e) = E(U_l|Q_2)$$

Hence all manager j's successors are also indifferent between Q_1 and Q_2.

The expected utility of manager $j-1$ in Q_1 equals:

$$E(U_{j-1}|Q_1) = \Pr(T_{j-1} = n|Q_1) + \Pr(T_{j-1} = o|Q_1)[q_{j-1}E(U_{j-1}|E_{j-1} = e, T_{j-1} = o, Q_1)$$

$$+(1 - q_{j-1})E(U_{j-1}|E_{j-1} = 0, T_{j-1} = o, Q_1)]. \quad (6)$$

After some computations,one can check that (6) can be rewritten as:

$$E(U_{j-1}|Q_1) = p(1-z_1...z_{j-2}) + (1-z_{j-1})[pz_1...z_{j-2} - pz_1...z_{j-2}e - (1-p)e] + z_{j-1}pz_1...z_{j-2}\gamma.$$

Similarly, one can write the other relevant expectations as:

$$E(U_{j-1}|Q_2) = p(1 - z_1...z_{j-2}) + pz_1...z_{j-2} - pz_1...z_{j-2}e - (1-p)e,$$

$$E(U_j|Q_1) = p(1 - z_1...z_{j-1}) + pz_1...z_{j-1} - pz_1...z_{j-1}e - (1-p)e,$$

$$E(U_j|Q_2) = p.$$

Combining all our previous insights, we can say that $W(Q_2) \geq W(Q_1)$ if and only if:

$$(1-p)e \geq z_{j-1}(1-p)e - pz_1...z_{j-1}(1-\gamma),$$

where z_{j-1} refers to the probability with which manager $j-1$ will not exert effort when $T_{j-1} = o$ under Q_1. It is easy to see that the inequality above is trivially satisfied. Note also that the inequality becomes strict whenever $j = 2$. Q.E.D.

LEMMA 2 *Assume the social planner can make q_t contingent on $(T_1, ..., T_t)$ and $(E_1, ..., E_{t-1})$. Assume there exists at least one j ($j = 2, ..., K$) such that $q_j = 1$. Then, $W((q_1, ..., q_{j-2}, q_{j-1}, 1, q_{j+1}, ..., q_K)) \leq W((q_1, ..., q_{j-2}, 1, 0, q_{j+1}, ..., q_K))$, where the inequality becomes strict whenever $j = 2$.*

Proof: Observe that if $t < K$ and $E_t = e$, then $\tilde{q}_j = 0$ ($\forall j = t+1, ..., K$). If $E_t = e$ there are two possibilities: manager t invents n or n doesn't exist. In both cases any future effort level is costly and unnecessary. Therefore if the social planner can make q_t contingent on $(E_1, ..., E_{t-1})$, q_t should be interpreted as the probability with which the social planner obliges manager t to exert effort if $(E_1, E_2, ..., E_{t-1}) = (0, 0, ..., 0)$.

Consider manager l ($l = 1, ..., j-2$). $E(U_l|E_l, (E_1, ..., E_{l-1}), Q_m)$ denotes manager l's expected utility given his effort level, given his predecessors' effort levels and given Q_m. Observe that $E(U_l|e, (0, ..., 0), Q_1) = E(U_l|e, (0, ..., 0), Q_2) = p - e$ and $E(U_l|0, (0, ..., 0), Q_1) = E(U_l|0, (0, ..., 0), Q_2) = p\gamma$. Hence, as before, manager l is indifferent between Q_1 and Q_2.

For reasons identical to the ones we detailed in our previous lemma, manager j's successors are also indifferent between Q_1 and Q_2. The relevant expectations in this case can be written as:

$$E(U_{j-1}|Q_1) = (1 - z_1...z_{j-2})p + (z_1...z_{j-2})[(1 - z_{j-1})(p - e) + z_{j-1}p\gamma],$$

$$E(U_{j-1}|Q_2) = p(1 - z_1...z_{j-2}) + z_1...z_{j-2}(p - e),$$
$$E(U_j|Q_1) = p(1 - z_1...z_{j-1}) + z_1...z_{j-1}(p - e),$$
$$E(U_j|Q_2) = p.$$

After some computations we get that $W(Q_2) \geq W(Q_1)$ if and only if:

$$(p - e) - z_{j-1}p(1 - \gamma) \leq (p - e)$$

(where z_{j-1} denotes the probability that manager $j - 1$ will not exert effort when $(E_1, ..., E_{j-2}) = (0, ..., 0)$ under strategy profile Q_1). As before, the above inequality is trivially satisfied and becomes strict whenever $j = 2$. Q.E.D.

We can now prove proposition (1). $I_t \in \{(T_1, ..., T_t), (T_1, ..., T_t)$ and $(E_1, ..., E_{t-1})\}$ denotes the information at the disposal of the social planner when choosing q_t.

Consider manager l ($l = 2, ..., K$). Assume,without loss of generality[20], that $\tilde{q}_{l+1} = ... = \tilde{q}_K = 0$. In that case there are two possibilities: (a) $p_l[(K-l+1)+(l-1)\gamma] \geq e$ or (b) $p_l[(K - l + 1) + (l - 1)\gamma] < e$, where $p_l = \Pr(n \text{ exists}|I_l)$. Inequality (b) states that if manager l exerts effort, the social cost exceed the social benefits. In case (a)

$$W((q_1, ..., q_{l-1}, q_l, 0, ..., 0)) \leq W((q_1, ..., q_{l-1}, 1, 0, ..., 0)).$$

But then from lemmas (1) and (2) we also know that:

$$W((q_1, ..., q_{l-1}, 1, 0, ..., 0)) \leq W((q_1, ..., q_{l-2}, 1, 0, 0, ..., 0) \leq ... < W((1, 0, ..., 0))$$

In case (b) $W((q_1, ..., q_{l-1}, q_l, 0, ..., 0) < W((q_1, ..., q_{l-1}, 0, 0, ..., 0)\ \forall q_l \neq 0$. Hence, if $q_l \neq 0$ ($l = 2, ..., K$), then the social planner can always find another Q in which $q_l = 0$ and in which welfare is strictly greater. Hence $\tilde{q}_l = 0$.

Consider now manager 1. By assumption $pK > e$ and therefore $\tilde{q}_1 = 1$. Q.E.D.

[20] This assumption is without loss of generality because one can apply the reasonning in this paragraph first to manager K

Proof of proposition (2):

This is a game of complete and perfect information. Hence we solve this game by backwards induction. Assume the first $K-1$ managers exert no effort and adopt o. Manager K observes that $T_K = o$, but he also observes that none of his predecessors exerted effort. Given that none of his predecessors exerted effort, manager K knows that n still exists with a probability p. Hence, $E(U_K|E_K = e, (E_1, E_2, ..., E_{K-1}) = (0, 0, ..., 0)) = p - e > E(U_K|E_K = 0, (E_1, E_2, ..., E_{K-1}) = (0, 0, ..., 0)) = 0$. Hence $q_K^* = 1$. Consider now manager one. If manager one exerts effort he gets $p - e$. However manager one knows that at least one of his successors will exert effort (ultimately manager K will exert effort if no manager did so before him). Hence manager $1'$s payoff of following the free-rider strategy equals $p\gamma$. As manager one rather imitates instead of innovates (assumption A2), $q_1^* = 0$. Consider now manager two. Manager two observes that $E_1 = 0$ and that $T_2 = o$. But now manager two is in the same position as the one in which the first manager was one period before. Hence $q_2^* = 0$. Continuing this reasonning until manager $K - 1$, it follows that $q_1^* = q_2^* = q_3^* = ... = q_{K-1}^* = 0$. Q.E.D.

Proof of proposition (3):

Assume all managers anicipate that only manager i exerts effort and that $q_i^* = 1$. Given this anticipation we show that:
(i) manager i's predecessors don't want to exert effort.
(ii) manager i's successors don't want to exert effort.
(iii) manager i wants to exert effort with probability one.

(i) Assume there exists a manager $l < i$. Manager l knows that $q_i^* = 1$ and he also know that no other manager exerted effort (or will exert effort). Manager l's gain of exerting effort, given that none of his predecessors exerted effort, equals $p - e$. Manager l's gain of not exerting effort, given that he anticipates that $q_i^* = 1$, equals $p\gamma$. Under A2 it's optimal for him not to exert effort and $q_l^* = 0$, $\forall l < i$.

(ii) Assume there exists a manager $l > i$. Assume $T_l = o$. Manager l computes:

$$p_l = \Pr(n \text{ exists}|T_l = o, Q^* = (0, ..., 0, 1, 0, ..., 0)) = 0$$

Therefore, manager l's gain of exerting effort equals $-e < 0$. Therefore, $q_l^* = 0$, $\forall l > i$.

(iii) Consider manager i. Manager i knows that no one else in the economy exerted effort (or will exert effort). Hence, $E(U_i|E_i = e) = p - e$, $E(U_i|E_i = 0) = 0$. Therefore, it's optimal for manager i to exert effort with probability one. Therefore, there exist a PBE in which only manager i is active and in which $q_i^* = 1$. Q.E.D.

Appendix two
Proof of point 1) of proposition (4):

by N.Melissas and M.Pauly

The proof is subdivided in two different steps. First we show that $\forall N \geq 2$ there exists a unique $Q^*, p_2, p_3, ..., p_N$ such that our $N + (N-1)$ equalities are respected. In this first step we don't look at the incentives of the non-active managers, we just take their (inactive) behaviour as given. Next, we show that an inactive manager cannot gain by deviating, given the behaviour of the other managers.

Step one: we start by reducing our system of $N + (N-1)$ simultaneous equations into a more tractable set of N equations in N unknowns. To illustrate our way of working, consider the first equation: $p - e = q_2 p\gamma + z_2 q_3 p\gamma + z_2 z_3 q_4 p\gamma + ... + z_2 ... z_{N-1} q_N p\gamma$ (where $z_t = 1 - q_t$). This equation can be rewritten as: $p - e = (1 - z_2 z_3 ... z_N) p\gamma$. This last equation can be rewritten as: $\kappa_2 + \gamma z_2 z_3 ... z_N = \kappa_1$ where $\kappa_2 = 1 - \gamma - e$ and $\kappa_1 = e\frac{1-p}{p}$.

Similarly we can rewrite equation t as: $(1-\gamma) + \gamma z_{t+1} ... z_N = \frac{e}{p_t}$ (if $t = N$ then $z_{N+1} = 1$). Replacing $\frac{1}{p_t}$ by $1 + \frac{c}{z_1 ... z_{t-1}}$ (where $c = \frac{1-p}{p}$) in this last equation we get $\kappa_2 z_1 ... z_{t-1} + \gamma z_1 ... z_{t-1} z_{t+1} ... z_N = \kappa_1$. We thus obtain the following set of N simultaneous equations:

(7) $$\kappa_2 + \gamma z_2 z_3 ... z_N = \kappa_1$$

(8) $$\kappa_2 z_1 + \gamma z_1 z_3 ... z_N = \kappa_1$$

$$\kappa_2 z_1 z_2 + \gamma z_1 z_2 z_4 ... z_N = \kappa_1$$

.

.

.

$$\kappa_2 z_1 ... z_{N-2} + \gamma z_1 ... z_{N-2} z_N = \kappa_1$$

(9) $$\kappa_2 z_1 ... z_{N-1} + \gamma z_1 ... z_{N-1} = \kappa_1$$

definition: We say that $(z_1, ..., z_N)$ is an equilibrium if z_t, computed out of our system of N simultaneous equations, $\in$ (0,1) $\forall t$

We first state and prove our Uniqueness Theorem.

Proposition 5 *(Uniqueness Theorem) There cannot be more than one equilibrium.*

Proof:
Suppose there are two equilibria $(z_1, ..., z_N)$ and $(z'_1, ..., z'_N)$. Assume that $z'_1 > z_1$. Out of the first equation we know that $z'_2...z'_N = z_2...z_N = \frac{\kappa_1 - \kappa_2}{\gamma}$. Out of the second equation we know that $\kappa_2 z_1 + \gamma \frac{z_1 z_2 ... z_N}{z_2} = \kappa_1 = \kappa_2 z'_1 + \gamma \frac{z'_1 z'_2 ... z'_N}{z'_2}$. It follows that:

$$z_2 = \frac{\gamma z_1}{\kappa_1 - \kappa_2 z_1}(z_2 ... z_N)$$

Hence, $z'_2 > z_2$. Similarly, we learn from equation t that $z'_t > z_t$. But this contradicts that $z'_2...z'_N = z_2...z_N = \frac{\kappa_1 - \kappa_2}{\gamma}$. If $z'_1 = z_1$, it's easy to show that $\forall t \; z'_t = z_t$ and that both equilibria are equal.Q.E.D.

Definition: $(z_1, ..., z_N)$ is a candidate equilibrium if $z_t \in (0, 1]$ $\forall t$ and,

$$\kappa_2 z_1 + \gamma a \frac{z_1}{z_2} = \kappa_1 \tag{10}$$

$$\kappa_2 z_1 z_2 + \gamma a \frac{z_1}{z_3} = \kappa_1$$

$$\vdots$$

$$\kappa_2 z_1 ... z_{N-1} + \gamma a \frac{z_1}{z_N} = \kappa_1 \tag{11}$$

where $a = \frac{\kappa_1 - \kappa_2}{\gamma}$.

Note that the system of equations starting from (10) to (and including) (11) merely represent a rewriting of the system of equations starting from (8) to (and including) (9). Note also that equation (7) does not intervene in our definiton of our candidate equilibrium. Therefore, every equilibrium is a candidate-equilibrium (to see this, replace (7) by $z_2 z_3 ... z_N = a = \frac{\kappa_1 - \kappa_2}{\gamma}$ and insert it in all the subsequent equations). However, every candidate equilibrium constitutes an equilibrium only if:
(a) all the $z_t \neq 1$,

(b) $z_2 z_3 ... z_N = \frac{\kappa_1 - \kappa_2}{\gamma}$

Our next proposition shows that once we know z_1, we are able to compute all the remaining $z_j's$ $(j = 2, ..., N)$.

Proposition 6 $\forall\ \alpha \in (0, 1]$, *there exists one and only one candidate equilibrium such that* $z_1 = \alpha$.

Proof:
If $z_1 = \alpha$, then equation (10) gives:

$$z_2 = \frac{(\kappa_1 - \kappa_2)\alpha}{\kappa_1 - \kappa_2 \alpha}$$

Note that $\kappa_1 - \kappa_2 > 0$, because of A2. Therefore $z_2 \in (0, 1]$. We can see from equations (10) to (and including) equation (11) that $z_j \in (0, 1]\ \forall j$. Q.E.D.

We know enough now to state our last proposition.

Proposition 7 *(Existence Theorem) Under A1, A2 and A3,* $\forall N > 1$ *there exists an equilibrium vector* $(z_1, ..., z_N)$.

Proof:
Assume that $z_1 = 1$. Then out of our system of equations starting from (10) to (and including) (11) we know that $z_2 = z_3 = ... = z_N = 1$ (because $\kappa_2 + \gamma a = \kappa_1$). But then $z_2 ... z_N = 1 > \frac{\kappa_1 - \kappa_2}{\gamma}$ (the reader can easily verify that under A1: $\kappa_1 - \kappa_2 < \gamma$).

Assume now that $z_1 < \epsilon$ (where ϵ represents an arbitrarily small strictly positive number). From (10) we know that:

$$\lim_{z_1 \to 0} z_2 = \lim_{z_1 \to 0} \frac{(\kappa_1 - \kappa_2) z_1}{\kappa_1 - \kappa_2 z_1} = 0$$

Hence $\exists\ \epsilon_2 > 0$ such that $\forall z_1 \leq \epsilon_2, z_2 < a^{\frac{1}{N-1}}$
More generally,

$$\lim_{z_1 \to 0} z_j = \lim_{z_1 \to 0} \frac{(\kappa_1 - \kappa_2) z_1}{\kappa_1 - \kappa_2 z_1 ... z_{j-1}} = 0$$

Hence $\forall j$, $\exists\ \epsilon_j > 0$ such that $\forall z_1 \leq \epsilon_j$, $z_j < a^{\frac{1}{N-1}}$.
Let's define:

$$\epsilon = \min_j \{\epsilon_j\} > 0$$

Then $\forall z_1 \leq \epsilon$, $z_j < a^{\frac{1}{N-1}}$ $\forall j$.
In that case , $z_2 ... z_N < a$.
We know that the function $f : z_1 \in (0,1) \rightarrow z_2 ... z_N \in \Re$ is a continuous one because $\forall j$, z_j is a continuous function of z_1 and we know that the product of continuous functions also yields a continuous function. We also know that $f(1) > a$ and $f(\epsilon) < a$. Hence there exists at least one $z_1 \in (0,1)$ such that $f(z_1) = a$. Q.E.D.

Step two: In this step we prove that an inactive manager cannot gain by deviating, given the behaviour of the other managers. Assume all managers anticipate that $q_l^* = 0$. We show that:
(i) if manager l moves before the first active manager, manager l cannot gain by deviating.
(ii) if manager l moves after the first active but before the last active manager, manager l cannot gain by deviating.
(iii) if manager l moves after the last active manager, manager l cannot gain by deviating.

(i) Call manager t the first active manager. Assume there exists a manager $l < t$. Manager l knows that none of his (potential) predecessors exerted effort. Therefore, if manager l exerts effort, he gets $p - e$. However, manager l knows that in the future the first active manager (= manager t) will be indifferent between exerting effort or not. This implies that:

$$E(U_l|Q^*, E_l = e) = p - e < q_t^* p\gamma + (1 - q_t^*)[p - e] = E(U_l|Q^*, E_l = 0)$$

Where the term between square brackets denotes manager t's payoff of not exerting effort. Therefore, manager l strictly prefers to exert effort with probability zero, $\forall l < t$.

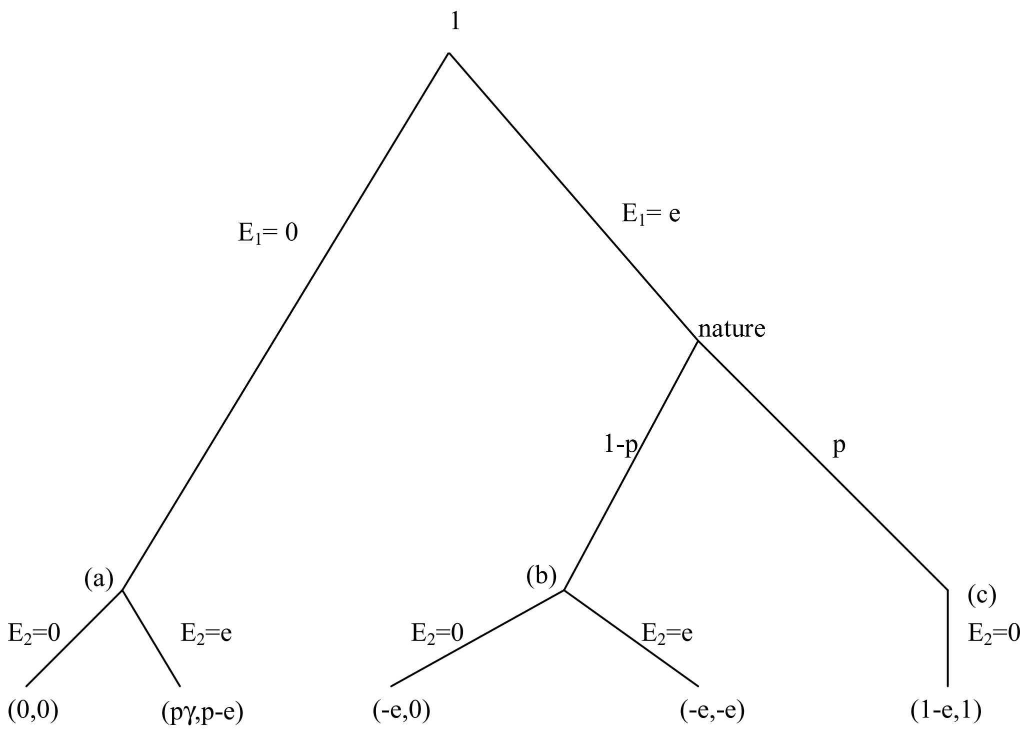
(ii) Call t the first active manager. Call t' the last active manager. Assume manager l is inactive and $t < l < t'$. Assume $T_l = o$. Assume, without loss of generality, that manager $l+1$ is active. If manager l exerts effort, he gets $p_l - e$. Manager l knows that manager $l+1$ will be indifferent between exerting effort or not. This implies that:

$$E(U_l|Q^*, E_l = e) = p_l - e < q_{l+1}^* p_l\gamma + (1 - q_{l+1}^*)[p_l - e] = E(U_l|Q^*, E_l = 0)$$

Where the term between square brackets represents manager $l+1$'s gain of not exerting effort. Therefore, manager l strictly prefers to exert effort with probability zero, $\forall l \in (t, t')$.

(iii) Call t the last active manager. Assume there exists a manager $l > t$. As we are focusing on PBE's with more than one active managers, we know that $p_t = e$. As $q_t^* > 0$, $p_l < p_t = e$. Therefore, $q_l^* = 0$, $\forall l > t$. Q.E.D.

Graph One: A simplified extensive-form representation in a world with two managers and observable effort levels.



Graph Two: A simplified extensive-form representation in a world with two managers and unobservable effort levels.