Circulation of ideas: Firms versus Markets

Thomas Hellmann (Stanford GSB) and Enrico Perotti (University of Amsterdam)

Abstract

We describe ideas as incomplete concepts requiring feedback from agents with complementary expertise, but, once shared, may be stolen. We study the comparative advantage of different environments for invention and implementation. Open exchange systems are good for circulation and thus matching, but may fail to reward idea generation. Controlled idea exchange systems, such as firms, reward idea generation but restrict their circulation. This identifies a basic trade-off between protecting the rights of invention and the best implementation of ideas. When firms and markets compete, the possibility of leakages will change optimal firm size and affects its internal organization.

1 Introduction[1]

The role of innovation in economic growth is well recognized (Paul Romer, 1990), yet the process of generation of innovative ideas is still a novel field of inquiry. The literature has focused on the role of intellectual property rights as incentive for invention (Aghion and Tirole, 1994), recognizing that new ideas may be easily appropriated once shared (Arrow, 1962; Anton and Yao, 1994). Yet patent protection can prevent idea appropriation only when the idea is deemed to be a fully developed novel concept. We focus here on an earlier stage in the innovation process, where ideas are still half-baked and need to be further elaborated before they are patentable. We ask what environment promotes the emergence of novel concepts which need to be shared to be screened and developed, when communicating them exposes the inventor to the risk that they will be stolen. In the process we explore the notion that

[1] We are grateful for comments by Amar Bhide. The usual disclaimer applies.

a free circulation of ideas may be at least as critical for generating innovation as their protection.

We start with a simple definition of a novel concept. Schumpeter (1929) argued that new ideas are original combinations of existing factors. Most random combinations are useless, while valuable ideas are combinations which "fit" together, aggregating the component resources in a novel, functional way (Biais and Perotti, 2003; Weitzman, 1998). Accordingly, we define an early stage idea as an incompletely specified combination, which requires elaboration along some dimension. If the concept has some merit, it will become implementable once elaborated by some agent whose expertise is complementary to the inventor's.

Because the concept is novel, it is not obvious who may be the agent with the right skills to complete it. Thus elaborating an idea requires trying to match it with various agents, seeking the right missing element. A broad circulation of ideas is thus critical for the process of innovation, as it allows maximum scope for any concept to find the right complementary match.[2]

Yet there is a fundamental problem with an open circulation of ideas, namely that information is not an excludable good (Arrow, 1962).[3] We assume that an incomplete idea is too vague for an independent patent office to grant exclusive property rights. We also assume that agents will not contract to cooperate on an idea without knowing its content in advance. We then study how alternative governance structures offer different trade-offs between the need to circulate ideas and to reward their generation. In particular, we study the relative merit of hierarchies (firms) versus markets.

In our model, an inventor who has developed an incomplete idea screens other agents on competence, seeking the right partner to assess and complete her concept (we call this person the "complementor"). Once the idea is elaborated, both the inventor's and the complementor's skills are required for implementation. If the inventor shares his idea with a complementor, this agent receives an informative signal, which he can either report it truthfully (i.e. give feedback), or not. If the idea is not feasible, there is nothing gained by misreporting. If instead the idea has potential, the complementor can give explicit feedback, which clearly explains how to complete the idea, and propose a partnership with the speaker. Since both parties have valuable implementation skills, an im-

[2] In her work on Silicon Valley, Saxenian (1994) suggests the term "cross-pollination" to highlight the innovative potential of open random matches.

[3] Arrow states that none would agree to pay for information before its content is revealed. But once one has been told the information the buyer cannot any longer be induced to pay for it.

mediate agreement is always efficient. The complementor has a stronger bargaining position, since he can easily find a substitute partner.[4]

However, given that the idea is both new and incomplete, it is difficult to identify the agent with the relevant complementary skills. Most often the inventor will present the idea not to a complementor, but to an agent whose skills, conditional on knowing the idea, essentially replicate his or her own. We call such an agent a "substitute".

If the inventor presents the idea to a substitute, this agent is unable to elaborate the idea, but once he has heard it can match the skills of the inventor. Substitutes are thus useless to the inventor and will seek to steal the idea. In addition, he may discourage the inventor by giving a false negative feedback. As a result, an idea may circulate through a sequence of substitutes agents, until matched to a complementor, who will either dismiss or implement the idea.

From a social perspective, an open exchange environment, i.e. a pure market environment where anybody can talk to anybody else, ensures a full screening and maximum implementation of good ideas. This, however, may not ensure a sufficient reward to the inventor, since unless he is lucky to share the idea immediately with a complementor, the idea is taken and (s)he does not receive any benefit from it. As a result, an open environment is efficient at elaborating and implementing ideas, but may fail to restrain idea stealing and thus to reward idea generation.

In response to this risk of idea appropriation, agents may choose to join a closed, "structured" exchange system, in which all participants are bound by an ex-ante commitment not to compete with any internally generated idea. A prime example of such a closed system is a firm, which we define as a collection of agents who relinquish to the firm owner, as part of their employment contract, any claim on the ideas generated by any employee, in exchange for a performance-based reward.

Our main results contrast the effectiveness of firms as a governance mechanism for the development of ideas against an open exchange system. Firms' main advantage is that they can control idea stealing and guarantee a reward for invention. Their main drawback, in addition to any monitoring cost, is that they restrict the set of agents that the inventor can talk to. While firms are better than markets a providing incentives for the generation of ideas, they are worse at ensuring that these ideas are elaborated and implemented.

We view joining a firm as equivalent to signing a contract, agreeing in advance not to compete with a set of yet unknown ideas. Why would an agent agree to such a contract, if he would refuse to sign a "blind"

[4]The complementor may also be able to withhold part of his complementary information until after an agreement is reached.

contract with another individual ? Here we rely on two elements: the infinitely repeated structure of our game and the multi-project nature of firms.[5] A reputation for fairness is valuable to firms as they may capture the residual value of many employees' ideas, and thus enables firms to credibly commit not to expropriate them without the agreed reward. In principle, any individual can set up a firm, by investing some resources in some sunk capital, which we term reputation (see Kreps, 1994). The value of this investment is more vulnerable to public scrutiny than an individual reputation for the infinitely many agents in the open system.[6]

Although the owner does not control directly his employees' ideas, he can prevent them from implementing any ideas outside the boundary of the firm. Employees can thus profit from their ideas only by reporting them to the owner in exchange for a promised reward. The owner can monitor idea generation and their internal circulation by establishing an internal recording system (read: a "bureaucracy") to create a verifiable paper trail of registered ideas and projects. This system creates some mutual guarantees for the owners and the employees: once an idea is recorded, the owner cannot expropriate it, nor can an employee take it outside the firm. Naturally, there must be an incentive for employees to report their ideas to the owner, and for the owner to record them rather than expropriate them.

Firm ownership thus represent a claim on the use of registered ideas (rather than tangible assets) vis-a-vis employees, the firm boundary constraints their escape, and its internal structure is aimed at ensuring both a controlled circulation of ideas and a reliable reward system. Indeed, employment law protects firms' claims explicitly on internally generated ideas in a number of ways, such as with trade-secrets and non-compete clauses. Naturally, in order to enforce their claim, firms need to record and control employee initiatives in order to create a verifiable paper trail to both reward invention and restrain idea theft. Our theory thus provides a justification for the frequently bemoaned bureaucracy associated with large hierarchical firms, in that it creates a protection for ideas that are too premature to be patented against theft by other employees.[7]

We next consider idea circulation when firms and markets interact. When an open idea exchange can reward idea generation, markets are strictly more efficient than firms. We thus focus attention to the case

[5] We refer here to large, multi-project firms, rather than start-ups created to implement a single idea.

[6] A firm reputation includes both a commitment to reward inventors as well as to fight idea stealing by employees.

[7] Of course, the internal recording does not protect the idea against a third party's claim. We assume here that an employee who would illegally share an internal idea with an outsider would not be able to profit secretly from it.

when they do not (so that markets are not viable by themselves). As a result, inventors do not wish to take an idea outside the firm if they can complete it inside. (They could do so by not registering the idea with the monitor, and leave.)

Yet many internally generated ideas indeed are developed after an employee leaves the firm.[8] We consider here the simple case of "efficient mobility", which firms may condone. It is in fact efficient for firms to allow inventors to develop independently those ideas not completed internally, in order to increase their incentive to generate ideas. This intuition is developed in much greater detail in Lewis and Yao (2003), who also consider the issue of idea-specific "fit" of firms' and employees' competencies.

We show that mobility produces a general equilibrium effect of increasing the reservation utility of potential employees, and thus their compensation. The reason is that now there is a positive reward for agents to operate in an open market, as they may be matched with an agent who has taken an idea from a firm. This forces an increase in average compensation for all employed agents. In conclusion, we can talk about a symbiosis of markets and firms. For an individual firm it is always efficient to let employees depart after exhausting internal matching opportunities. As this generates a flow of ideas to the open system, this promotes independent activity there, even though market exchange by itself does not support ideas generation.

Finally, we show how markets may specialize in ideas with lower cost of development, for which they can offer a sufficient reward, while costlier, more ambitious initiatives will be generated inside firms, which can commit to reward a larger investment.

What alternatives are there to markets and firms ? In an open system containing an infinite number of agents, a conventional reputation process may be ineffective. However, within a closed system whose membership is valuable, a reputation mechanism may indeed work. In an extension we discuss partnerships and networks as special governance structures for idea exchange.

We define a partnership as a firm which requires idea circulation and project implementation within its boundary but does not rely on monitoring, saving on monitoring costs. The drawback is an inability to reward directly the generation of ideas. We show that partnerships may

[8] According to Bhide', over 70 % of the founders of firms in the Inc 500 list of fast growing young firms replicated or modified ideas encountered in their previous employment. Gompers, Lerner and Scharfstein (2003) show how a very high number of new entrepreneurs in Silicon Valley and Massachussets left larger firms and started their firm thanks to contacts in the venture capital network.

be viable in cases where markets or firms fail, but their ability to reward invention dissipates fast as the number of partners grows, so their size is limited by a free-riding constraint.

We also present a stylized idea of a network of free agents as an intermediate solution between a close and an open system. Our approach assume a loose network discipline which condones idea stealing but demands a high frequency of idea contribution. We argue that a looser discipline allows easier and more mutual monitoring relative to a reputational model of cooperation. It may thus sustain a larger number of participants and thus of potential matches relative to stricter punishment schemes. We plan to pursue this theme in future research.

This paper contributes to a rich literature. The theme of the optimal allocation of control over innovative ideas is analyzed in Aghion and Tirole (1994). Prior work has recognized the problem faced by an inventor in talking to potential partners in an unstructured market environment. Anton and Yao (1994) show that the inventor can avoid idea stealing and secure some rents by threatening to transmit the idea to competitors. Cestone and White (1998) and Baccara and Razin (2002) also study how the threat of competition can deter information leakages. Anton and Yao (2003) considers partial disclosure of ideas. Rajan and Zingales (2001) suggest that hierarchy may address the problem of idea-stealing. Granting selectively access to the technology to employees with relation-specific investment may avoid defection. Biais and Perotti (2003) show that contractual cooperation over new ideas is feasible only when the experts' incentive or ease to steal the idea is not too large, and that this depends critically on the complementarity of their information. We pursue this notion of complementarity further, by allowing the agent with complementary expertise to elaborate the idea. Related work include Hellmann (2001), who studies the sequence of resource commitments in an open bargaining setting when potential partners can add information. Ueda (2002) examines the trade-off of talking to uninformed investors (banks) that cannot appropriate an idea, versus venture capitalists that can. A related theme is carefully developed in Chemmanur and Chen (2003).

In section 2 we explain the base model. In section 3 we contrast open systems (markets) and closed systems (firms). In section 4 we extend the model to allow for the coexistence of firms and markets, and we study the mobility of ideas across firm boundaries. Section 5 discusses the implications of the model for firms' internal organization. Section 6 considers as intermediate governance forms the case of non hierarchical closed systems sustained by mutual monitoring, namely partnerships and networks. We close with a brief conclusion.

2 The Base Model

2.1 Ideas and Agents

In our model, ideas are the sole productive asset. They are generated by individuals at some cost, and they are initially private information (as it is private information whether an agent has generated any idea). Ideas are incomplete, so they need to be elaborated by someone with a set of skills complementary to those of the inventor before they are patentable. They may be good, with prior chance p, or bad. Inventors seek to elaborate ideas by sharing them with others, seeking to find someone with a good match of skills. However, since the idea is novel and incomplete, it is not obvious what the required complementary skills are, or who might have them.

There are some search costs to screen out incompetent agents (incompetent relative to the specific idea). An incompetent agent has little comprehension of the idea, and can neither complement it nor steal it. Since the narrator discounts future payoffs, he is willing to incur some search costs to screen ideas faster by identifying knowledgeable agents. [9] We assume that search costs are low enough that searching for an idea with a prior chance p of being good is valuable.

Competent agents have skills of two types, which cannot be distinguished ex ante. The first type are "complementors", who understand whether the idea is good, and if so are able to provide some complementary feedback to complete it.[10] The second type are "substitutes," who do not know whether the idea is good, or how it can be completed; however, once they hear the idea, have the same implementation skills as the inventor. Because they have nothing to contribute, they have no incentive to report truthfully their ignorance, and are likely to steal the idea.

2.2 Firms and Markets

An open exchange system offers an inventor unrestricted matching with an infinite set of agents. Inventors are not subject to any governance structure that restricts their quest to implement their ideas. However, they are also exposed to the danger of having their idea stolen, as inventors cannot contractually protect their idea. We discuss this assumption in the next subsection.

In contrast, a firm defines the set of individuals on whose output the owner can lay an ownership claim, and whose transactions are restricted

[9] The screening may also avoid that incompetent agents circulate the idea to potential competitors.

[10] The frequency of this critical match may be rare for very novel ideas.

to take place inside the boundary. By becoming firm employees, agents willingly accept a blanket non-compete clause over all ideas generated or heard during their employment contract. We assume initially that this claim is perfectly enforceable, only to relax it later.

The owner of the firm can capture the value of employees' ideas by policing the boundary, which forces internal completion and transform ideas in tangible projects on which the firm can claim ownership. For this, the firm needs to keep a record of employee activities. This produces verifiable evidence (a "paper trail") on what ideas are generated within the firm, and who generated them. These records also help the firm to commit to an internal compensation system. The monitor cannot appropriate all the of a registered idea, since this observable would undermine the firm reputation in front of all employees.

Since the hierarchical firm is the classic Coasian counterpart to a market exchange system, in the model we mostly focus on the contrast between such firms and free markets (open exchange systems). Later we will discuss partnerships and networks.

2.3 The impossibility of ex ante contracting on ideas

We take the view that ideas are contractible: in principle, if an idea can be described verbally, it can also be contracted upon. Yet incomplete ideas are not yet patentable. We assume that individuals will not agree to sign a contract binding them to cooperate on a single idea whose content they have not seen, as this may unduly restrict future opportunities or expose them to extortion risks. In legal terms, such contracts are called "non-disclosure agreements" or NDAs. For instance, a NDA may contain ideas and knowledge that the agent already possesses, potentially expropriating him. [11] Anton and Yao (1994, 2003) point out that in practice, NDAs are rarely used, are very difficult to enforce, and many institutions involved in funding new ideas, such as venture capitalists, explicitly refuse to sign them.

An agent joining a firm signs away the right to use any of the ideas circulating within the firm.[12] There are several reasons why would agents be willing to join firms, but unwilling to sign NDAs with other individuals. The employment contract may be signed at an early stage, before the agent has developed any idea. Firms build reputations by some sunk

[11] In addition, even if a listener were to sign a NDA, it may be costly for the talker to enforce it. A firm which has invested in a reputation and has a constant flow of projects has a stronger In addition, even if a listener were to sign a NDA, it may be costly for the talker to enforce it. A firm which has invested in a reputation and has a constant flow of projects has a stronger incentive to enforce the agreement.

[12] Trade secret and non-compete clauses are standard features of employment contracts.

investment, and thus has an interest in maintaining a long-term reputation for honoring its commitments. Finally, the firm keeps bureaucratic records on the flow of internal initiatives as a tool to enforce its intellectual property and its commitment to reward inventors. A large firm has also a larger incentive to fight to enforce its intellectual property, since beyond the value of the individual case, it has its long-term reputation at stake.

Once an idea is recorded its origin is verifiable, so the monitor agent is bound by the same rules as the productive employees. Obviously, the monitor might also steal an idea by not recording it, but this exposes him to more serious reputational consequences than an individual because he would miss on a steady flow of multiple projects.[13]

(for evidence and an extensive discussion, see Anton and Yao, 2003). Under these circumstances it would seem that no specific contractual commitment is possible without sharing the idea. An incomplete idea must thus be necessarily shared with potential partners before they commit to cooperate, and thus be by nature exposed to appropriation.

2.4 Stage game

The model describes an infinite sequence of dates in which agents interact. Assume there is a unit mass of agents. At the beginning of each period, an agent may decide to generate an idea at a private cost ψ. An idea comes with probability γ. It may be good, with probability p, or bad. Ideas are incomplete and need to be completed before they are productive. Ex ante, it is not known what the complementary piece is, so ideas must circulate among many agents to find their match. The choice of matching depends on the governance arrangement. To seek a knowledgeable partner, the carrier of the idea incurs a search cost c.[14] After an agent with an idea has presented it, the listener receives a private signal on the quality of the idea and gives a report. Finally, the idea is either implemented, or dropped, or taken to another listener next period by one or both agents. Agents' discount factor is δ.

Competent listeners understand the idea, but most of them cannot tell whether it can be completed. Such agents (substitutes) occur with probability $1 - \phi$. While he has no additional insight, a substitute has an incentive to steal the idea and try to complete it with the help of someone else next period. Thus, a substitute may become a talker.

With probability ϕ the listener has the right complementary expertise

[13] The firm may want to hire as monitors agents who are not knowledgeable, an interesting reason for having dumb bureaucracies.

[14] For instance, (s)he needs to expend some effort to avoid being matched with agents with no chance of having any related skill.

to assess the idea. The signal received by this complementor is always correct. If the idea is good, the complementor knows how to complete the idea. He can choose to deliver this information, which establishes his type and thus reveals the idea to be good. Since two agents are always needed to implement an idea, the complementor can engage the talker as a collaborator. The type of agreement they may strike depends on the governance arrangement. The complementor may also steal the idea.

If a complementor recognizes the idea as bad, he does not want to pursue idea alone, and gives negative feedback. However, a negative report by a complementor cannot be distinguished from that of a substitute listener. Thus a negative feedback leads to a downward revision in the narrator's beliefs on the quality of the idea, even though it may hide idea stealing. Under a condition derived below (called the discouragement condition), the talker stops pursuing the idea after a negative report. We will impose this condition to simplify the analysis.

When two teams implement the idea simultaneously, we assume that they engage in Bertrand competition, so both teams get zero returns. Once the idea has been implemented, everybody knows about it, and no one else tries it again.

We assume that agents can only carry one idea from one period to the next. They can thus share at most one idea, and listen at most to one idea from their matched partner. This structure allows us to treat talking and listening as independent activities.

Throughout the paper we denote the per period utilities of agents with lower cases u and v, and the lifetime utilities (i.e., the net present value of the current and all future per-period utilities) with upper cases U and V. u and U pertain to agents in the open system, while v and V pertain to agents inside firms.

3 Analysis of the model

3.1 Analysis of the open exchange system

In an open exchange system, there are no restrictions about who can talk to whom or what they can do thereafter. Consider first the bargaining between a complementor and the agent that told him the idea, and who is now a potential collaborator. If the idea is implemented, it generates an expected value z. On his own, the complementor could gain z_0 from the continuation game, where he finds another partner to implement the idea with. The original talker cannot produce anything without the complementor. Using Nash bargaining, the talker gets $az = \frac{z - z_0}{2}$ and complementor gets $\overline{a}z = (1 - \frac{z - z_0}{2z})z = \frac{z + z_0}{2}$. To compute z_0, note

that the original talker does not want to take alone the idea further. He knows that the complementor will pursue the idea, and he does not want to end up in competition. The complementor thus takes the idea to somebody else. With probability ϕ the complementor finds another complementor. Their skills set are actually not matched, since they both have the complementary skill, but lack the skills of the original inventor, or its substitutes. The optimal bargain at this point is that one of two complementor pursues the idea. Since they are symmetric, they can flip an even coin. Thus, with probability ϕ the complementor gets $\frac{z_0}{2}$. With probability $\overline{\phi}$, the complementor finds a partner that is a substitute. In this case, he can again bargain for $\overline{a}z$. Thus $z_0 = \delta(\phi\frac{z_0}{2} + \overline{\phi}\overline{a}z - c)$. After simple transformations we obtain $z_0 = \frac{\delta(\overline{\phi}\overline{a}z - c)}{(1 - \frac{\delta\phi}{2})}$ and

$$a = \frac{1 - \delta + \frac{\delta\phi}{2} + \frac{\delta c}{z}}{2 - \delta}$$

The bargaining share of the talker (a) is larger when the complementor has higher relative search costs (higher $\frac{c}{z}$), when the complementor is more likely to encounter another complementor (higher ϕ), and when the complementor is more impatient (lower δ).

Straightforward calculations reveal that $\frac{\phi}{2}\frac{c}{z} < a < \frac{1}{2}$, so that the share of talker is also always less than half. This shows how the open system favors the complementor rather than the inventor of an idea.

With this, we can compute the return to generating and stealing an idea. Conditional on having an idea, the utility of the talker is

$$u_T = \phi paz - c$$

The expected period utility of an idea inventor is given by

$$u_G = \gamma u_T - \psi$$

Part of the expected benefit of being a listener comes from the chance of becoming a complementor, as captured by

$$u_L = \phi p\overline{a}z$$

Below we derive the total benefit of being a listener, which also includes the benefit of stealing the idea.

We now introduce an intuitive assumption that allows us to maintain stationarity of the process of idea circulation. Namely, we identify

some condition under which a talker receiving negative feedback would abandon the idea. We call this the discouragement condition.

To derive it, consider the continuation game after negative feedback. With probability $\frac{\phi\overline{p}}{\phi\overline{p}+\overline{\phi}}$ the reported negative signal was honest, so the utility of pursuing the idea for one period will be $-c$. With probability $\frac{\overline{\phi}}{\phi\overline{p}+\overline{\phi}}$ the negative signal was dishonest. If the talker pursues anyway, he always incurs the search cost c. With probability $\overline{\phi}$ the original talker finds no match, and gets no benefits. With probability ϕ^2 both the talker and the substitute listener find their respective match, each getting the Bertrand payoff, which is zero. And with probability $\phi\overline{\phi}$ the talker finds a match, but the substitute listener does not. In this case the talker gets paz. Thus the expected return for the talker from pursuing an idea one more period after negative feedback is $\frac{\overline{\phi}}{\phi\overline{p}+\overline{\phi}}\phi\overline{\phi}paz - c$. The discouragement condition ensures that an agent prefers to generate a new idea, rather than pursuing an old idea on which he has received negative feedback. Formally, the condition is given by

$$c > \underline{c} \equiv \frac{\phi\overline{\phi}^2 paz}{\phi\overline{p}+\overline{\phi}} - (\gamma\phi paz - \psi).$$

Throughout the analysis we assume that this condition holds. Note that for low values ψ, $\underline{c}$ may be negative anyway.

To derive the steady state properties of the open system, denote by s the fraction of agents who just stole an idea and by $\overline{s}$ those without any (stolen) ideas. The number of people who have an idea are all those who stole an idea (s), and those who did not but generated one ($\overline{s}\gamma$). Thus the number of people talking is given by $t = s + \overline{s}\gamma$. Naturally, this is also the number of people listening.

We can thus establish the following. In any period, $tp\phi$ projects get implemented, and $t\overline{p}\phi$ projects get stopped because a complementor realizes they are bad and discourage the talker without taking them further. Finally, in any period, $t\overline{\phi}$ projects get stolen. Thus $s = t\overline{\phi}$. Using $t = s + \overline{s}\gamma$ we get $t = \frac{\gamma}{\phi+\gamma-\gamma\phi}$ and $s = \frac{\gamma-\gamma\phi}{\phi+\gamma-\gamma\phi}$. It is easy to see that the rate of talking (t) and stealing (s) both increase with the probability of generating an idea (γ) and both fall with the probability of finding a complementor (ϕ).

Consider now the ex-ante utility of an agent in such a system, assuming that everyone invest in generating new ideas. With probability s the agent steals an idea, and gets u_T.With probability $\overline{s}\gamma$, the agent

has a new idea and also gets u_T.With probability $\overline{s\gamma}$, the agent has no idea and gets nothing. In addition, with probability t, an agent becomes a listener and gets u_L. We get[15]

$$U = \frac{\overline{s}u_G + su_T + tu_L}{1-\delta}$$

Note that U is increasing in p, z, δ, ϕ and γ, which is all very intuitive.

The open exchange system generates efficient outcomes. In particular, ideas circulate until they find their complementary match. They are then either implemented if good, or dropped if bad. The above expression for U also shows that the payoff for an agent in the open system is exactly the net present value of generating ideas. In the open exchange system each agent on average receives a payoff equal to his per capita quota of ideas. But the agent does not receive the payoff from his own ideas. Instead one large fraction of his returns stem from being the complementor to other people ideas. And another fraction stems from circulating ideas that were stolen from other agents.

The main issue is whether agents have an incentive to generate ideas. So far the calculations assume that agents want to generate ideas. We now examine for what values of ψ this is true. Let the critical cost threshold is given by

$$\psi_M \equiv \gamma u_T = \gamma(\phi paz - c) = \gamma\left[\phi pz\frac{1-\delta+\frac{\delta\phi}{2}+\frac{\delta c}{z}}{2-\delta} - c\right]$$

Proposition 1 *Consider an open exchange system.*

- *For $\psi < \psi_M$, the system achieves a first-best outcome, where ideas circulate until they are either implemented if good, or dropped if bad.*
- *For $\psi > \psi_M$, the expected reward of generating an idea os too low because of idea stealing. No ideas are therefore generated.*
- *Markets are more likely to fail*
    1. *if finding an idea is rare (low γ),*

[15] To see the derivation, denote the lifetime utility with a stolen idea by U_T and without a stolen idea by U_G, then $U_T = u_T + tu_L + \overline{s}\delta U_G + s\delta U_T$ and $U_G = u_G + tu_L + \overline{s}\delta U_G + s\delta U_T$. Using $U = \overline{s}U_G + sU_T$, we rewrite this as $U_T = u_T + tu_L + \delta U$ and $U_G = u_G + tu_L + \delta U$, so that $U = \overline{s}(u_G + tu_L + \delta U) + s(u_T + tu_L + \delta U) = \overline{s}u_G + su_T + tu_L + \delta U$.

2. *if the idea is unlikely to be good (low p),*
3. *if ideas have little value (low z),*
4. *if finding a right match is rare (low ϕ),*
5. *if it is costly to search for a match (high c),*
6. *and if the complementor is sufficiently patient (high δ).*

This results echoes Grossman and Stiglitz (1980), who derive the impossibility of perfectly informative financial prices in a frictionless environment, also because there are no incentives to generate information. Here, the problem with generating ideas is that a large fraction of the returns are appropriated by the complementor, not the inventor of an idea.

3.2 Analysis of hierarchical firms

We now consider the firm as a possible solution to the weakness of the pure open system. We define the firm as a closed exchange system in which employees sign a NDA relative to all internally registered ideas. The monitor establishes a record for all such ideas, which allows the firm to prevent their use outside the firm and thus to capture their value. A verifiable record also serves as a visible commitment to compensate idea generation and internal completion.

Specifically, we assume that each firm has a monitor (or boss). She chooses the size of the firm by choosing the number of employees F. The total (lifetime) cost of monitoring is given by $M(F)$, where $M > 0$, $M' > 0$ and $M'' > 0$. We assume that the monitor can perfectly control the external boundaries of the firm (we relax this later). She can also register the generation of ideas inside the firm, and assign rewards to their generation and completion. Let b be the compensation for the inventor of the idea. It is easy to see that in this simple model the firm has no need to compensate the complementor, so that the utility of listening is zero.

Concretely, suppose that an inventor "registers" the idea with the owner, who, although never able to complete it, assigns to the inventor the task to try to find an internal match (i.e., act as an internal "champion" for the idea). The fact that employees are unable to escape the boundary of the firm allows the champion to obtain reliable feedback from all listeners within the firm. We assume that the firm will not hire another agent to complete an unmatched idea. We also assume that the inventor must abandon the idea if no match is found, but will relax it later. For comparability with the open system, we include an equivalent

search cost also inside the firm.[16]

We denote the per period utility by v and the lifetime utility by V. The expected utility of being a talker in any period is given by

$$v_T = \phi pbz - c$$

To calculate the lifetime utility, we need to take into account how long an idea has already circulated. Let V_G denote the agent's utility when he does not have an idea, and thus needs to generate a new one. Let f be the fth round of talking about an idea, so V_T^f is the lifetime utility of an employee who is about to talk to the fth listener. For any f with $1 \leq f < F$, we have

$$V_T^f = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^{f+1}$$

This says that in the fth round an agent has an expect per-period return of v_T. With probability ϕ the idea is resolved (the idea is either implemented or dropped) so that the agent can start afresh, generate new ideas and get V_G. And with probability $\overline{\phi}$ the idea is still unresolved, the agent continues to circulate his idea, and gets V_T^{f+1}. At $f = F$, the agent has no one left to talk to, so

$$V_T^F = V_G \text{ where } V_G = -\psi + \gamma V_T^1 + \overline{\gamma}\delta V_G$$

Clearly, these utilities form a recursive system. The following lemma characterizes their solution.

Lemma 1:
(i) $V_G = \tau v_T - \rho\psi$ where

$$\tau = \frac{\gamma \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i}{1 - [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]}$$

$$\rho = \frac{1}{1 - [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]}$$

(ii) V_G is increasing in F.
(iii) $V_T^1 > V_T^2 > ... > V_T^{F-1} > V_T^F = V_G$.

The proof is in the appendix.

[16] In reality, search costs may be lower inside firms.

Here, τ is the discounted number of times that an employee has an idea to talk about, and this also the discounted number of times an agent gets to listen to an idea. And ρ is the discounted number of times that an employee generates an idea. Thus V_G is the discounted value of talking about ideas, minus the cost of generating them.

The intuition why V_G is increasing in F is that a larger firm provides more opportunities to circulate the idea, implying a greater value of generating an idea. The intuition why $V_T^f > V_T^{f+1}$ is that having more people left to talk to (i.e., having a fresher idea) is better for the inventor, since there is a greater probability of completion.

The monitor's net present value of profits is given by

$$\Pi = F(1-b)\tau\phi pz - M(F)$$

The monitor's objective is to maximize Π, subject to the agent's incentive constraint. For this, she chooses the optimal size F, as well as a compensation package b.

The incentive constraint naturally concerns the willingness to generate ideas in the first place, and is given by $V_G \geq 0$. V_G is increasing in b (through v_T). Since the monitor wants to minimize b, we get

$$V_G = 0 \Leftrightarrow b^* = \frac{c + \psi\frac{\rho}{\tau}}{\phi pz}$$

Substituting b^* into Π yields

$$\Pi(F) = F\tau(\phi pz - c) - F\rho\psi - M(F)$$

where $F\tau(\phi pz - c)$ is the number of employees times how many ideas they have, times the expected value of an idea net of search costs, $F\rho\psi$ is the number of employees, times how many times they have to seek new ideas, times to cost of generating ideas.[17]

Proposition 2 *The monitor's optimal choice of F^* is*

- *decreasing in the marginal cost of monitoring (M').*
- *decreasing in the cost of generating ideas (ψ)*

[17] In this setting, the monitor's profit function equals the social return, since he can hold agents down to their reservation utility. This does not imply that firms are efficient, since from a social perspective it would be better to let ideas circulate freely.

- *decreasing in the search cost (c)*
- *and increasing in the profitability of new ideas (p and z).*

The optimal compensation b^* is

- *increasing in the cost of generating ideas (ψ)*
- *increasing in the search cost (c)*
- *and decreasing in the profitability of new ideas (p and z).*

The proof is in the appendix.

This proposition generates some intuitive results about the optimal size of the firm. The boundaries of the firm are determined by the cost of monitoring the activities of employees. Naturally, the greater the marginal cost of monitoring an additional employee, the smaller the firm. The boundary of the firm is also affected by the fundamental properties of ideas. The greater the cost of generating ideas (high ψ), the smaller the firm, because the marginal benefit of having another employee is lower. The same applies to search costs (c). Viceversa, the more profitable the agent's ideas (higher p or z), the larger are firms.

The comparative statics of b^* are also intuitive. If the cost of generating an idea is greater (higher ψ, lower γ), inventors need to be compensated more. Interestingly, in addition to this direct effect, there is also an indirect effect. Higher generation costs reduce the optimal firm size F^*, which in turn reduces the return to generating an idea. The change in the optimal reward b^* must take this into account. A same logic applies to all other comparative statics.

4 The interaction of firms and markets

4.1 Comparing firm and market governance

We are now in a position to compare firms and markets. We first state how the two systems differ in terms of their ability to resolve and implement new ideas.

Proposition 3 *In an open system, all ideas are resolved. By contrast, in a closed system, restrictions on the circulation of ideas imply that not all ideas are resolved. The probability that an idea gets resolved is given by* $\sum_{i=0}^{i=F-2} \phi(\overline{\phi})^i < 1$. *This is an increasing function of F.*

Thus our basic result is that firms restrict the circulation of ideas to ensure a sufficient reward for idea generation, but this limits the set of possible matches and thus the realization of the full potential of new ideas. In particular, some good ideas will not be elaborated and implemented.

To make the open and closed system directly comparable, we have assumed so far that search costs are just as high in an open as in a close exchange system. It may well be that firms reduce search costs, by providing a collection of like-minded people. On the other hand, for any given idea, it is less probable to find the perfect match within the small group of people inside the firm, relative to the vast universe of people outside the firm. This suggests another interesting structural trade-off: in the open system people face higher search costs (high c), but are more likely to get better matches (high ϕ). The relative performance of the open system depends obviously on how much better the matches become, relative to how much more expensive it is to search. An interesting implication of this kind of comparison, however, concerns the expected speed of idea resolution. In the market, ideas are likely to be resolved faster. This is for two reasons: first, as noted in Proposition 3, ideas within the firm may never be resolved. But even conditional on firms resolving the idea, markets are faster when they have a higher value of ϕ.

We next turn to the question under what circumstances firms emerge in this model. For this we note that as long as markets are viable ($\psi \leq \psi_M$), there is no role for firms. In this model markets are efficient when they work. And firms are inefficient, in the sense that they have to restrict the circulation of ideas, and that they incur additional monitoring costs. The model predicts that firms may emerge when markets fail. Naturally, whether or not firms emerge depends on their own viability. We note that Π is decreasing in ψ. We define ψ_F so that $\Pi(F^*, \psi_F) = 0$, where F^* is the optimal choice of F. ψ_F is the critical value, above which firms are not viable. For $\psi \leq \psi_M$, open exchange system is feasible. In general, we cannot say whether ψ_F is greater or smaller than ψ_M, since this depends on the cost of monitoring (which may also include a fixed cost of setting up a firm). If monitoring costs are high, then $\psi_F < \psi_M$, and firms are never efficient. But if monitoring costs are not too high, then $\psi_F > \psi_M$.

4.2 Specialization in idea generation

We sketch here the question of what kind of ideas are better suited for markets or firms. In particular, we consider a trade-off between ambitious but costly ideas versus small and cheap ideas. To provide a fair comparison, we consider the set of ideas that all have the same social return. We define this to be the expected total return of an idea, net of the cost of generating and circulating it. This is given by

$$S = -\psi + \gamma[(\phi pz - c) + \overline{\phi}\delta(\phi pz - c) + (\overline{\phi}\delta)^2(\phi pz - c) + ...] = -\psi + \gamma\frac{\phi pz - c}{1 - \overline{\phi}\delta}$$

The following proposition examines the trade-off in generating small vs big ideas in firms vs markets.

Proposition 4 *For a given set of ideas, that all have the same social return, but that differ in terms their size (z) and generation costs (ψ), markets dominate for small ideas (low z and ψ), but firms dominate for big ideas (high z and ψ).*

This says that bigger ideas are more difficult for the market. Put differently, markets are good for lower cost (low ψ) but also lower returns (low z) projects. Overall, this suggests that markets are good a small, high risk projects. firms, by contrast, have a relative advantage at larger projects, bigger ideas, and also more safe projects.

Another very interesting comparison concerns incentives and compensation. We are arguing that firms provide a better environment for protecting ideas, and that this means better incentives to generate ideas. This may seem counter-intuitive at first, since it is commonly believed that firms provide fewer incentives and less generous compensation for innovations than markets. Put bluntly, intrapreneurs make less money than entrepreneurs.

Our model is actually consistent with this, and provides some additional insights. First, we have to be careful with who we call an entrepreneur. In markets, we may observe large returns, but for the outside observer it is difficult to ascertain whether these returns are made by the generator, by some substitute who heard the idea, or by a complementor. Our model predicts that we would see the largest returns for the complementor. In contrast, inside the firm a complementor does not get any (extra) reward. Providing feedback is considered part of being

a good citizen inside a firm, but it does not give rise to large financial returns.

What about the financial returns of the generator inside a firm. First, the total reward for innovation is not overly generous, since in fact $V_G = 0$, i.e., the bonus compensation for generating ideas is just enough to compensate the generator for his costs of generating the idea. Second, suppose we had two comparable projects and we wanted to compare how the market and the firm compensate the generator. If the firm matches the compensation that the agent would get in the market, we would find that the financial returns in the market look larger. But the reason for this is simply that in the market, the agent has one chance to implement his idea. The firm, by contrast, can offer the generator several opportunities to try out his ideas. The observed compensation inside the firm (b) would therefore be lower than the observed compensation inside the market (a). However, what we should also observe is that the likelihood of getting compensated (i.e., the likelihood of success) is much lower in the market (ϕ), and considerably higher inside the firm ($\phi \sum_{i=0}^{i=F-2} \overline{\phi}^i$). Again, out model predicts that innovation in the market looks more risky (even though in this case we are comparing two identical projects!).

4.3 Coexistence of firms and markets

So far we considered firms and markets in isolation. We now discuss how firms and markets may interact. We are particularly interested in the flow of ideas across firms and markets.

To what extent does the presence of markets affect firms? We will show it is efficient for firms to circulate further any idea which could not find an internal completion. In particular, the firm can improve incentives by releasing any employee who was unable to complete his idea inside the firm. Since employees value the opportunity to seek to extract some value from their ideas, firms can reduce their incentive costs by allowing mobility of initiatives after exhausting all internal matching opportunities. This generates an interesting coexistence of firms and markets, where the open system, even if unable to reward independent idea generation, may thrive by feeding off ideas generated, but not completed inside firms.

To derive such a coexistence equilibrium, we need to model the flow of ideas and people across firms and markets. Assume that there is a continuum of agents normalized to one. Suppose there is a fixed number of firms , and let θ be the steady state fraction of agents that work for some firm. The remaining fraction $(1-\theta)$ resides in the open system. A

firm that releases an employee always rehires another agent to maintain its optimal size, so that if $\lambda\theta$ is the fraction of employees that are leaving firms to explore their ideas in the open system, then firms rehire in each period the same fraction $\lambda\theta$. We continue to assume that $\psi > \psi_M$, so none would wish to generate ideas in the open system, which is thus by itself not viable.

We calculate λ as follows. For any firm, let Q_0 be the expected number of employees that have no idea at beginning of period. With probability $\gamma\overline{\phi}$ they generate a new idea which is not resolved within the period. This means that in steady state every period $Q_0\gamma\overline{\phi}$ employees start with an idea that is exactly one period old. Similarly $Q_0\gamma\overline{\phi}^f$ (for $f < F-1$) have an idea that is exactly f periods old. Adding up the total number of employees, we obtain $Q_0(1+\gamma\overline{\phi}+\gamma\overline{\phi}^2...+\gamma\overline{\phi}^{F-2}) = F$ $\Leftrightarrow Q_0 = \dfrac{F}{\overline{\gamma}+\gamma\sum_{i=0}^{F-2}\overline{\phi}^i}$. The number of employees that leave because they cannot find a match inside the firm is $Q_0\gamma\overline{\phi}^{F-1}$. The fraction λ is therefore given by

$$\lambda = \frac{Q_0\gamma\overline{\phi}^{F-1}}{F} = \frac{\gamma\overline{\phi}^{F-1}}{\overline{\gamma}+\gamma\sum_{i=0}^{F-2}\overline{\phi}^i}$$

It is easy to verify that this is increasing in γ and decreasing in ϕ.[18]

We now need to modify the utility of employees. While the expression for V_T^f and V_G remain the same, V_T^F now includes the utility that the employee gets from taking his idea outside. Once the idea leaves the firm, it becomes exposed to idea expropriation. A departing employee receives the same expected returns as any other agent in the open system, given by $u_T = \phi paz - c$.

Let U_0 be the lifetime utility of an agent that currently resides in the open system and does not have any idea to pursue. We derive this below. The utility of the departing employee is then simply given by $V_T^F = u_T + U_0$. For simplicity we ignore the possibility that employee turnover may increase the possibility of a match inside the firm, assuming the firm hires an "identical" worker type.

The ability to move to the open system now endogenizes the employee's reservation wage. In order to hire employees, firms have to match their utility of being in the open system, which requires $V_G \geq U_0$. Firms will optimally set their compensation b so that this condition is satisfied with equality, i.e., $V_G = U_0$. Straightforward calculations reveal

[18] Thus the flow of ideas to the market increases with the rate of idea generation and falls with the rate of completion.

that V_G is now given by

$$V_G = \tau v_T - \rho\psi + \widetilde{\rho} u_T$$

where the first term measures the expected reward to complete ideas inside the firm, the second the expected cost of generating them, and $\widetilde{\rho} = \rho\gamma(\overline{\phi}\delta)^{F-1}$ measures the (appropriately discounted) probability that the employee will not find a match inside the firm and will seek a match outside.

Letting the employee leave increases his utility of generating ideas, and thus allows the firm to reduce its incentive compensation. Without the option to leave, b would have to satisfy $v_T(b) = \dfrac{U_0 - \rho\psi}{\tau}$, while now it only has to satisfy $v_T(b) = \dfrac{U_0 - \rho\psi - \widetilde{\rho} u_T}{\tau}$. Note that this result is closely reminiscent of the results in Lewis and Yao (2003).

From the perspective of an individual firm, allowing an employee to pursue ideas that could not be completed internally is beneficial. However, from a general equilibrium perspective, the open system raises employees' reservation utilities.

We continue to assume that $\psi > \psi_M$, so that no agent wants to generate ideas in the open system. However, agents currently in the open system may hear ideas of released employees, and either steal them or become their complementor. This "escape of ideas from firms" makes it attractive to stay in the open system.

To formally see this, we derive the equilibrium value of U_0. For this, we need to consider the circulation of these ideas in the open system. If there are $\lambda\theta$ new ideas every period, a fraction ϕ gets resolved (either implemented or stopped), but a fraction $\overline{\phi}$ continues to circulate. The number of ideas that have circulated for exactly one period is thus $\lambda\theta\overline{\phi}$. Similarly, the number of ideas that have circulated for i periods is $\lambda\theta\overline{\phi}^i$. In steady state, the total number of ideas circulating is thus $\lambda\theta\sum_{i=0}^{i=\infty}\overline{\phi}^i = \dfrac{\lambda\theta}{\phi}$.

Consider now the probability that an agent in the open system is approached by another agent with an idea. Since there are $\dfrac{\lambda\theta}{\phi}$ ideas, and $\overline{\theta}$ people to talk to, the probability of becoming a listener is $\dfrac{\lambda}{\phi}\dfrac{\theta}{\overline{\theta}}$. We assume that there are enough agents in the open system, such that $\dfrac{\lambda}{\phi}\dfrac{\theta}{\overline{\theta}} < 1 \Leftrightarrow \theta < \dfrac{\phi}{\lambda + \phi}$. The utility of an agent without an own idea in the open system is thus given by

$$U_0 = \frac{\lambda}{\phi}\frac{\theta}{\overline{\theta}}\phi(p\overline{\alpha}z + \delta U_0) + \frac{\lambda}{\phi}\frac{\theta}{\overline{\theta}}\overline{\phi}(\delta p\alpha z + \delta U_0) + (1 - \frac{\lambda}{\phi}\frac{\theta}{\overline{\theta}})\delta U_O$$

The per period utility has two terms: the first is the gain from listening to an idea and being able to complement it; the second term from listening to an idea and stealing it. We can term their sum the period utility of being a listener in the open system. (With some probability the agent does not hear any idea.) Firms must now pay at least this amount of reservation utility to attract workers.

An agent may in each period remain in the open system, or alternatively join a firm. Its compensation inside a firm must match the utility of remaining in the open system waiting to hear an idea, $V_G = U_0$. We can now solve the above equation for U_0, and we obtain the reservation utility of all agents in the economy, given by

$$U_0 = \frac{1}{1-\delta}\frac{\lambda}{\phi}\frac{\theta}{\overline{\theta}}(\phi p\overline{\alpha}z + \delta\overline{\phi}p\alpha z)$$

This reservation utility is increasing in θ. The more firms there are, the more employees depart, and the more attractive it becomes for an agent to be in the open system. This equilibrium is more efficient, as now all good ideas are realized.

The implication of the movement of ideas between firms and markets on equilibrium compensation is clear: agents now earn more inside firms because they must receive a positive opportunity cost. Additionally, a market which in itself does not support ideas generation may nevertheless thrive in implementation of new ideas thanks to the presence of many firms.

The implications for firm size are not unambiguous. Firms may pay a lower compensation because employees receive some additional gain from each idea. Thus firms indirectly gain part of the increase in value due to the implemented ideas which had to be discarded before. Agents in the open system also gain a fraction of such ideas, which means that the reservation utility of employees certainly increases.

It is easy to see that the final effect on the cost of hiring employees, and thus on the optimal size of the firm, will depend on the fraction of the value of the additionally completed ideas gained by the departing employees relative to their matches in teh open system. Perhaps less obviously, it depends also on how often free agents gets to hear an idea which left a firm, and thus on the density of firms.

The number of firms is therefore endogenous. Were firm profitability and thus firm size or number increase as a result of mobility, there would be a beneficial effect on the value of being in the open system, and the cost of employing individuals will thus increases, limiting the ability of firms to capture the additional value and thus their size.

In conclusion, we can talk about a symbiosis of markets and firms.

For an individual firm it is always efficient to let employees depart after exhausting internal matching opportunities. As this generates a flow of ideas to the open system, this promotes independent activity in the open system, and increases employee's reservation utility.

4.4 Discussion on Firms versus Markets

A main distinctive feature of firms in our model (as opposed to markets, or start-ups aimed at implementing one specific project) is that firms are long term, multi-project institutions, which cannot be replicated by arm's length transacting individuals. (Of course, an individual may stop acting as an independent agent and start a firm by incurring some fixed costs, which may include the cost of building some reputation.) These features ensure a high value of reputation for protecting ideas. Specifically, firms seek a reputation which underlies a commitment not to expropriate inventors, and to prosecute idea stealing by employees, which is essential to enforce the firm boundary. This combination ensures that talented individuals will join firms as a way to commit ex ante to cooperate with other people's ideas without stealing them, in exchange for a reliable mechanism to share ideas for implementation.

A specific feature of hierarchical firms concerns their active regulation of the internal generation and circulation of ideas. Their formal structure for recording and managing internal initiatives create a formal "paper trail" which facilitates the verifiability by third party of idea stealing, either by the monitor or by employees. Thus the "bureaucratic" nature of firms, while adding costs and possibly slowing down initiatives relative to more informal arrangements, have an essential function in ensuring that ideas are better protected in firms than in an open exchange.

The notion that markets produce poor incentives relative to firms for idea generation may sound at first counter-intuitive. Typically, the reward to develop a new venture on one's own are much greater than for an employee. Yet our model is consistent with this intuition. Firms can encourage employee initiatives by offering low but reliable rewards because the open exchange offers high potential returns, but also a high risk of idea expropriation.

The model offers some suggestions also about the relative profile of employed and free agents. Firms constrain idea circulation in order to capture their value, and offer reliable but minimal rewards for inventors and complementors up to their reservation utility. Thus firms (and universities) may be uniquely important as a protected environment for the generation of initial concepts, even though they may capture most of the value created by the inventor. In contrast, markets tend to re-

ward those who elaborate and implement ideas rather than those who originate them. The open exchange system may attract those agents who are best at elaborating and implement new ideas, rather than pure inventors.

We have assumed that an inventor always choose to disclose the idea to the firm, in order to protect his claim internally. Yet in reality many employees leave to start an own business, without disclosing their idea to their employer. If some ideas fit better with the firm's competence than others, employees will rather leave than disclosing the ideas to avoid wasting time with the "bureaucratic" process of seeking an internal match, when the chance of an internal match is low and the search cost and discount rate are high.[19]

Even an inventor with a registered idea may become increasingly impatient as the chance of a match decrease over time. Firms may adopt a rigid approach about letting employees leave the firm, lest they lose too many projects, forcing inventors to seek all possible internal matches. The drawback is that it discourages inventors from reporting some ideas. Firms may in contrast choose for flexible rules, allowing employees discretion on how long to search internally before they are allowed to take the idea outside the firm. In general, the optimal internal organization will depend on the relative frequency of general and firm specific ideas, and on firms' specific cluster of competence.

An alternative notion of mobility arise when an employee may escape with someone else's registered idea. We consider it in a related paper, as the treatment is quite complex, not least because it may lead to parallel development of the same idea. A basic intuition is that opportunistic mobility forces firms to adjust compensation for those agents with the greatest incentive to leave. As the ability to escape increases, the higher cost of providing incentives to retain ideas reduces the profitability of employing agents, and induces a reduction in firm size.

Whatever the causes of the flow of ideas from firms to markets, it suggests that there is a natural symbiosis between the ability of firms to sustain exploration in ambitious ideas and the comparative advantage of open market partners, such as venture capitalists, in screening and implementing ideas which leave firms.

Because many independent entrepreneurs reap large rewards, it often look as if markets by themselves promote more innovation than firms. Our approach helps to qualify this point, suggesting that successful entrepreneurs may not necessarily be the original inventors of new ideas, but rather those with a creative capacity for idea elaboration, a contri-

[19] Note that this provides a simple intuitive reason for why firms choose to pursue "narrow" strategies (see Rotemberg and Saloner, 1994).

bution sometimes as significant as the initial invention. The main role of markets may in fact be to be able to implement ideas that originate in firms but cannot be elaborated there, which would lead to conclude that large firms are a crucial feature of highly innovative environment. After all, any account of Silicon Valley's success starts with a list of the remarkable number of novel ventures started by individuals who left employment with a larger firm. Our model confirms an incubating role for enterprises suggested by many authors (e.g. Saxenian, 1994), and suggest a symbiotic relationship between firms, other institutions such as universities, and free agents operating in an open exchange market. In contrast, a secretive corporate culture which resists interaction with markets, such as those which dominated on the East Coast or in Europe, will not sustain an innovative environment because firms may suffocate the elaboration of internally generated ideas and thus their potential. The hierarchical approach to R&D in Japan and Europe, as well as in the large high tech companies on Route 128 in Massachusetts, is often contrasted with the loosely organized open environment of Silicon Valley in California (Saxenian, 1994; Aoki, 2002). The success of the Silicon Valley model is attributed to a free movement of ideas and individuals among small ventures assembled via informal arrangements. Yet the intense exchange of ideas in Silicon Valley is puzzling, since California actually has a fairly weak tradition of protecting intellectual property (Gilson (?) and Hyde (?)), so it is not clear how idea generation may be rewarded. In practice, many new ventures in Silicon Valley develop ideas which originated in large firms. According to Bhide, over 70 % of the founders of firms in the Inc. 500 list of fast growing young firms replicated or modified ideas encountered in their previous employment. The open environment in Silicon Valley may thus thrive also thanks to the historical presence in the area of large firms, which have acted as incubators of new ideas, particularly those which were costly to develop.

In any coexistence of markets and firms, employees will leave to implement their projects outside firms only if there are some reliable channels for open exchange of ideas. We need thus to understand better the informal arrangements supporting an open exchange of ideas. The role of informal "networks of practice" has been highlighted by all researchers of innovative clusters. Evidence of a major role of venture capitalists in large corporations in entrepreneurial spawning is offered by Gompers, Lerner and Scharfstein (2003), who also offer evidence on the role played by venture capital and other informal contacts. Hellmann (2003) provides a related theory of when employees leave, and how intellectual property rights affect their incentives.

5 Mutual monitoring mechanisms

So far we have assumed that agents outside a hierarchy cannot commit to reward idea generation. In practice, there are decentralized market mechanisms which limit idea stealing, such as mutual monitoring. We now consider informal arrangements where subsets of free agents create self-enforcing mechanisms to share ideas among themselves.

Any such mutual arrangement may function only within a limited set of agents (thus a closed system), since there are limits to the scope of mutual monitoring. A larger arrangement is more efficient at producing matches, and will adopt the maximum punishment available in repeated interaction among free agents: if a member is found to be violating the rules, (s)he is excluded from any further interaction with any other agent in the set. We assume that an ostracized agent cannot find another mutual arrangement willing to adopt him/her. Finally, we assume that agents perceive that a failure to enforce such a threat leads to a collapse of the arrangement.

We still assume that the open system is not viable by itself, so that the continuation payoff after stealing an idea and being banished from the network is zero.

We now turn to describe two special cases, which we term partnerships and networks.

5.1 Partnership

We define a simple non-hierarchical closed exchange system, such as a partnership where agents share all output. We assume that the informal rule of a partnership (as opposed to a legal rule) requires internal completion of all ideas, and that partners can see whether one of them implements a project with a non partner. Thus just as a firm, a partnership has a clear border, but has no monitoring mechanism to record new ideas or control idea stealing directly, nor make rewards contingent on idea generation or completion. The function of the border is here to define a violation of the partnership rule that all projects must be implemented internally, so that profits from all transactions can be shared among all partners. [20]

Since ideas can only circulate internally, they may not find their efficient match. Thus the optimal size for a partnership is the maximum number of partners which still provides incentives for idea generation and discourages idea stealing.

[20] Assuming that employees are identical, the appropriate reward system grants everybody the same fraction of firm returns.

As a partnership expands, the direct reward for inventors falls rapidly, as the returns are shared with more and more partners. Let P be the number of partners. The incentive constraint that ensure idea generation is given by

$$V_G(P) = \frac{-\psi + \gamma v_T^P \sum_{i=0}^{i=P-2} (\overline{\phi}\delta)^i}{1 - [\overline{\gamma}\delta + \gamma\phi\delta \sum_{i=0}^{i=P-2} (\overline{\phi}\delta)^i + \gamma(\overline{\phi}\delta)^{P-1}]} \geq 0$$

where $v_T^P = \phi p \frac{z}{P} - c$. [21] The optimal size of the partnership is thus given by P^* such that $V_G(P^*) = 0$. This ensures the largest size for a maximum chance for internal completion.

The incentive constraint can be rewrittrn as

$$V_G(P) = \tau v_T^P - \rho\psi \geq 0 \text{ where } v_T^P = \phi p \frac{z}{P} - c$$

where τ and ρ are evaluated at $F = P$.

A second condition is necessary to ensure that neither an inventor nor a complementor would steal the idea (and thus be expelled). If they did, they would get some expected return from the stolen idea, and 0 ever after. The continuation utility of staying in the partnership, since each partner receives the average product, is given by

$$V(P) = \tau v^P - \rho\psi \text{ where } v^P = \phi p z - c$$

The exact continuation value depends on the number of ideas inside firm. Note that $V(P)$ is computed under the assumption that no one in the firm has any new idea. We impose this extreme case, as it represents the most severe existence constraint faced by the partnership. If this condition is satisfied, none will ever steal an idea from another partner. Then the overall incentive constraint for the generator is $V(P) \geq p\phi a z$, while for the complementor it is $V(P) \geq (1-a)z$. Clearly, the more binding constraint is due to the generator.

In summary, to be sustainable under all circumstances, a partnership needs to satify both $V(P) \geq (1-a)z$ and $V_G(P) = \tau v_T^P - \rho\psi \geq 0$

We can next show that partnerships may be viable when markets fail. Define ψ_P the lowest value of ψ, so that $V_G(P, \psi) \leq 0$ for all P. Partnerships are feasible for $\psi \leq \psi_P$, but not for $\psi > \psi_P$.

Proposition 5 *Partnerships may be feasible at intermediate idea generation costs. In particular, when markets cease to be feasible, partnerships are still viable.*

[21] It can be shown that $V_G(P)$ is concave in P, and as $P \to \infty$ we have $V_G(P) < 0$.

Proof. We need to show that $\psi_P > \psi_M$. Note that at $P = 2$ we have $V_G(P) = -\psi + \gamma(\phi p \frac{z}{2} - c)$. At $\psi = \psi_M = \gamma(\phi paz - c)$ we thus get $V_G(P, \psi_M) = \gamma(\phi p \frac{z}{2} - c) - \gamma(\phi paz - c) > 0$ since $a < \frac{1}{2}$. Thus $\psi_M < \psi_P$. ■

Thus the size of the partnership is limited by the fact that rewards need to be shared equally among all partners, irrespective of their contribution to the innovation. For high costs of idea generation $\psi > \psi_P$ it is impossible to satisfy the participation constraint, and partnerships become infeasible.

Beyond firms and partnerships, one can think of further closed systems. Some extreme cases are families or clans, which have a predetermined (often genetic) definition of membership, so that joining them is largely impossible, and leaving them extremely costly. More generally, the threat of exclusion may be most effective to discipline idea stealing if there is no option to join another form of close exchange and open exchange is unreliable.

5.2 Networks

We sketch here a stylized model of a "diffuse network", defined as a completely decentralized arrangement, without any internal tracking system nor any coordinating monitor. We take the view that a network is a more impersonal arrangement than reputation mechanisms, where the disciplining of idea stealing requires a direct and precise observation of individually actions.

The more specific features of our definition of a network are that unlike other reputational arrangements, such as those in a partnership, a network does not constrain individuals to trade only among members, whose identity is common knowledge. The reason is that monitoring by network members cannot establish whether a talker has originated or stolen an idea, nor can it observe all members' productive activities.

Participation in the network requires that agents regularly contribute new ideas to other members. Agents who just listen to ideas or simply repeat ideas by others are expelled. Since ideas circulate freely in the network, agents can recognize at least some of those taken by others.

Thus we view a network as too diffuse to be able to monitor idea stealing or to provide any form of profit sharing on the set of ideas implemented by its members. Agents are not punished for taking around ideas generated by others, or for implementing some of them elsewhere, as long as they share enough ideas with network members. This is perhaps a cynical but realistic view of a diffuse network, which condones opportunism, as long as it achieves a long term fair share to its members.

We assume that the size of the network N is such that this form of mutual monitoring can occur at no cost. Note that the size of the network is optimally set as large as possible, provided that its members have an incentive to keep presenting ideas to other members.

We assume, as in the general open system, that the discouragement condition holds, so that if the inventor of an idea receives a negative feedback (s)he stops pursuing it and seeks another one.

Since a substitute agent will earn nothing from reporting her true signal, (s)he will always give negative feedback in order to take the idea further. Thus an idea inventor earns a return only when immediately matched with a complementor. The direct return to idea generation is exactly as in the open system, namely

$$u_G = \gamma u_T - \psi$$

where $u_T = \phi p \alpha z - c$. In such a network the main benefit of generating ideas is to be admitted to listen to other network agents' idea. The expected benefit of being a listener comes from the chance of being a complementor, or to find one next period:

$$u_L = \phi p \alpha z + (1 - \phi)\delta[\phi p \alpha z - c]$$

Being a listener is strictly better than the return to idea generation. Since there are no penalties, substitute agents will always steal the idea. Thus the main incentive problem is again to reward idea generation.

Let U_0 be the continuation value in network. For a complementor, the utility from stealing is then as in the open system

$$\delta[(\phi(1 - \alpha)z + (1 - \phi)\frac{z}{2} - c - c]$$

which is less than agreeing to complete the idea immediately with the talker, which yields $\phi(1 - \alpha)z$ and thus avoids discounting and search costs. Therefore meeting a complementor in a network always leads to a cooperative implementation. This confirms that a negative feedback will be informative, since a complementor with a good signal would report it, and is consistent with the discouragement condition assumed earlier.

To complete the model, we assume that the effort to generate a new idea is observable, so that mutual monitoring can police the requirement that each network member contributes his fair share of ideas. In conclusion, inventors sharing ideas within a network often have their idea taken but are able to maintain membership and thus to gain access to others' ideas.

Provided agents indeed have an incentive to generate ideas, it is easy to see that

Proposition 6 *The expected payoff to participate in a diffuse network is the same as the payoff to join a partnership.*

The proof is very intuitive: over time, network agents are like partners who contribute and complete ideas with a common frequency. In both contexts each agent receives the average per capita value of all ideas generated within the group. The difference is that in the partnership each partner receives exactly the same share from all projects, while in a network each agent receives half of the value of all ideas which he participate in implementing, whether as a inventor, substitute idea-taker or complementor. Over time, all agents contribute the same average number of ideas, and have access to the same number of ideas by other network members.

What if other agents could not observe this effort ? An alternative mechanism to ensure an incentive for idea generation is that the network members records over time whether each agent contributes his/her fair share of own ideas.

Specifically, the rule has to ensure that any agent who did not take an idea from the last period has an incentive to attempt to generate an idea. With probability γ this leads to an idea, with probability $(1-\gamma)$ it does not.

Network participation would then require that each agent over time presents an idea with a minimum probability θ, where θ equals the chance to have taken an idea last period plus the chance of not having taken an idea but having attempted to generate an idea. The chance to have taken an idea last period equals the chance to hear an idea on which the agent is substitute. Thus $\theta = (1-\theta)\gamma + \theta\overline{\phi}$ which implies $\theta = \frac{\gamma}{\gamma+\phi}$.

Thus the rule states that an agent must presents a new idea to other network members with frequency θ in order to be admitted to participate further. For this rule to be self-enforceable, it must be optimal to comply, which requires that the capitalized value of remaining in the network as an active idea inventor be positive, namely

$$U_0 = \theta u_T - (1-\theta)\gamma\psi + (1-\theta)u_L =$$

$$\theta[\phi p\alpha z + (1-\phi)\delta[\phi p\alpha z - c - c] - (1-\theta)\gamma\psi + \phi p\alpha z \geq 0$$

and that it would exceed the gain from a deviation from the network norm. Here the best deviation would save idea generation costs by free riding on the ideas circulating in the network, which implies that the agent would have an idea to talk about only with probability $\theta\overline{\phi}$. If we

define t as the period over which such a deviation would be detected, then the network is viable if the potential loss of network participation is greater than the gain from saving on idea generation costs:[22]

$$\delta^t U_0 > \sum_{i=o}^{t} \delta^i (1-\theta)\gamma\psi$$

A special version of this threat would be to assume that an agent who presents ideas more often is sought more often as a listener by other agents in the network, which is attractive since it reduces the number of periods without a new idea. This produces an equilibrium in which all agents would choose the same degree of idea generation effort.

The type of network we have outlined is only one of many possible mutual monitoring mechanisms. It is a particularly weak form of governance, as it does not attempt to control idea stealing nor to capture output to reallocate it. It simply predicates that all bilateral deals must take place among members of the network and that all members must present their fair share of own ideas to other members.

There are other possible mechanisms for mutual monitoring. Another classic example is the classic reputation game, which assumes that individual behavior is observed by others "within sight" and that any opportunistic action be punished by ostracism from this group. This mechanism could in principle control idea stealing, and involve some monitoring costs; it would produce an outcome similar to a close system. However, it may be extremely hard in practice for other agents to identify who, among two people talking to each other, was the agent with the idea and who the idea thief. At a minimum, this requires a high degree of effort, and since all agents in the group must have a precise signal to be able to attribute fault, it would entail high, duplicate monitoring costs. More likely, such precise monitoring may not be feasible except in small, geographically localized circles. Our definition of network is probably a description of a more diffuse, if also less precise, form of peer monitoring.

6 Conclusions

We have proposed a novel trade-off between the necessity to protect idea generation and the need to share ideas in order to screen and elaborate them. A free circulation of ideas is thus most efficient at the elaboration

[22] An additional gain from deviating may be the capture of a greater share of profits in case of trading with non network members. This would not usually be the case for a clueless agent, if network members split their profits from cooperation, but may be an issue for complementors.

of incomplete concepts, but fail to reward any personal search for novel concepts. Individuals may then voluntarily join close exchange systems such as firms to ensure that their initiatives may receive support and feedback without being appropriated.

Much further work is needed to elaborate the important theme of the circulation of ideas among individuals and institutions. We have sketched some questions and outlined some directions of research we intend to pursue. Yet much is left to be done. Curiously, this paper says little about some institutions, such as universities, which do encourage a broad circulation of ideas across their boundaries. Interestingly, academic researchers often do not capture much of the value created by their discoveries, often gained by agents more practical at implementation, although the open publication system ensures some form of reward. This suggests that an open process of generating and exchanging ideas is a very valuable public good.

7 References

- Aghion, P. and J. Tirole, 1994, "On the management of innovation", Quarterly Journal of Economics, 1185–207.

- Anton, J., and D. Yao, 1994, "Expropriation and Inventions", American Economic Review, 190–209.

- Anton, J., and D. Yao, 2002, "The sale of ideas: Strategic disclosure, property rights and contracting", Review of Economic Studies.

- Anton, J. and D. Yao, 2003, "Attracting skeptical buyers", Working paper, Duke University.

- Arrow, K., 1962, Economic Welfare and the Allocation of Resources for Inventions, in R. Nelson (ed), The rate and direction of inventive activity: Economic and social factors, Princeton University Press, Princeton.

- Baccara, and Razin, 2002, "From thought to practice: Appropriation and endogenous market structure with imperfect intellectual property rights," Working paper, Princeton University.

- Biais, Bruno and Enrico Perotti, 2003, "Entrepreneurs and New Ideas", CEPR Discussion Paper no 3864, London

- Bidhé, Amar V., 2000, "The Origin and Evolution of New Businesses", Oxford University Press

- Cheung, "Property Rights in Trade Secrets", Economic Inquiry, 1992.
- Chemmanur, T. and Z. Chen, "Angels, Venture Capitalists, and Entrepreneurs: A Dynamic Model of Private Equity Financing", Boston College mimeo, 2003
- Garmaise, M., 2001, "Informed Investors and the Process of Financing Entrepreneurial Projects", University of Chicago mimeo
- Hellmann, Thomas, 2001, "Entrepreneurship and the Process of Obtaining Resource Commitments", mimeo, Stanford University
- Hellmann, Thomas, 2003, "When do emplyees become entrepreneurs", mimeo, Stanford University
- Lazear, E., 2002, Entrepreneurship, Working paper, Graduate School of Business, Stanford University.
- Lewis, Tracy and Dennis Yao, 2003, "Innovation, Knowledge Flow, and Worker Mobility", mimeo, University of Pennsylvania, Philadelphia
- Rajan, R. and L. Zingales, "The Firm as a Dedicated Hierarchy", Quarterly Journal of Economics, 2001
- Romer, P., Endogenous Technological Change, Journal of Political Economy, Vol. 98, 1990
- Saxenian, Anna Lee, 1994, "Regional Advantage: Culture and Competition in Silicon Valley and Route 128", Harvard Business School Press, 1994
- Schumpeter, J., 1926, Theorie der wirtschaftlichen Entwicklung, Duncker and Humblot, Berlin.
- Schumpeter, J., 1942, Capitalism, Socialism and Democracy, George Allen and Unwin, London.
- Weitzman, M., 1998, Recombinant growth, Quarterly Journal of Economics, 331–360.

8 Appendix

8.1 Proof of Lemma 1

For part (i) consider first the case of $F = 3$ then $V_T^2 = v_T + \delta V_G$, $V_T^1 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^2$ and $V_G = -\psi + \gamma V_T^1 + \overline{\gamma}\delta V_G$. This is a system of three variables and three unknown. Solving it, we find that $V_G = -\psi + \overline{\gamma}\delta V_G + \gamma V_T^1 = -\psi + \overline{\gamma}\delta V_G + \gamma v_T + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta V_T^2 = -\psi + \overline{\gamma}\delta V_G + \gamma v_T + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta v_T + \gamma\overline{\phi}\delta\delta V_G = -\psi + \gamma v_T + \gamma\overline{\phi}\delta v_T + \overline{\gamma}\delta V_G + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta\delta V_G$ so that

$$V_G = \frac{-\psi + \gamma v_T + \gamma\overline{\phi}\delta v_T}{1 - (\overline{\gamma}\delta + \gamma\phi\delta + \gamma\overline{\phi}\delta\delta)}$$

Consider next the case of $F = 4$, where $V_T^3 = v_T + \delta V_G$, $V_T^2 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^3$, $V_T^1 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^2$ and $V_G = -\psi + \gamma V_T^1 + \overline{\gamma}\delta V_G$. From $V_G = -\psi + \overline{\gamma}\delta V_G + \gamma V_T^1 = -\psi + \overline{\gamma}\delta V_G + \gamma v_T + \gamma\phi\delta V_G + \gamma\phi\delta V_T^2 = -\psi + \overline{\gamma}\delta V_G + \gamma v_T + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta v_T + \gamma\overline{\phi}\delta\phi\delta V_G + \gamma\phi\delta\overline{\phi}\delta V_T^3 = -\psi + \overline{\gamma}\delta V_G + \gamma v_T + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta v_T + \gamma\overline{\phi}\delta\phi\delta V_G + \gamma\phi\delta\overline{\phi}\delta v_T + \gamma\phi\delta\overline{\phi}\delta\delta V_G = -\psi + \gamma v_T + \gamma\overline{\phi}\delta v_T + \gamma\overline{\phi}\delta\overline{\phi}\delta v_T + \overline{\gamma}\delta V_G + \gamma\phi\delta V_G + \gamma\overline{\phi}\delta\phi\delta V_G + \gamma\overline{\phi}\delta\overline{\phi}\delta\delta V_G$ we obtain

$$V_G = \frac{-\psi + \gamma v_T + \gamma\overline{\phi}\delta v_T + \gamma\overline{\phi}\delta\overline{\phi}\delta v_T}{1 - (\overline{\gamma}\delta + \gamma\phi\delta + \gamma\overline{\phi}\delta\phi\delta + \gamma\overline{\phi}\delta\overline{\phi}\delta\delta)}$$

For the general case, it is now easy to see that

$$V_G = \frac{-\psi + \gamma v_T \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i}{1 - [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]}$$

where for the denominator we use $[\overline{\gamma}\delta + \gamma\delta(\overline{\phi}\delta)^{F-2} + \gamma\phi\delta \sum_{i=0}^{i=F-3} (\overline{\phi}\delta)^i] = [\overline{\gamma}\delta + \overline{\phi}\gamma\delta(\overline{\phi}\delta)^{F-2} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i] = [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]$. Naturally, this maps directly into $V_G = \tau v_T - \rho\psi$ where

$$\tau = \frac{\gamma \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i}{1 - [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]}$$

$$\rho = \frac{1}{1 - [\overline{\gamma}\delta + \gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i]}$$

For part (ii) we simply note that the numerator of V_G is increasing in F (since $\sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i$ increases with F) and the numerator is decreasing in F (since $\gamma(\overline{\phi}\delta)^{F-1} + \gamma\phi\delta \sum_{i=0}^{i=F-2} (\overline{\phi}\delta)^i$ increases with F). Note also that for the same reasons, both τ and ρ are increasing in F.

For part (iii) we consider again the case of $F = 4$. To see that $V_T^3 > V_G$, assume to the contrary that $V_T^3 < V_G$. Then we have $V_T^2 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^3 < v_T + \delta V_G = V_T^3$, so that $V_T^1 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^2 < v_T + \phi\delta V_G + \overline{\phi}\delta V_T^3 < v_T + \delta V_G = V_T^3$ and thus $V_G = -\psi + \gamma V_T^1 + \overline{\gamma}\delta V_G < -\psi + \gamma V_T^3 + \overline{\gamma}\delta V_G = -\psi + \gamma v_T + \delta V_G \Leftrightarrow V_G < \dfrac{\gamma v_T - \psi}{1-\delta}$. But this is not possible, since $V_G(F) > V_G(F=2) = \dfrac{\gamma v_T - \psi}{1-\delta}$. Thus $V_T^3 > V_G$. To see that $V_T^2 > V_T^3$, note that $V_T^2 - V_T^3 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^3 - v_T - \delta V_G = \overline{\phi}\delta(V_T^3 - V_G) > 0$. For $V_T^1 > V_T^2$ note that $V_T^1 - V_T^2 = v_T + \phi\delta V_G + \overline{\phi}\delta V_T^2 - v_T - \phi\delta V_G - \overline{\phi}\delta V_T^3 = \overline{\phi}\delta(V_T^2 - V_T^3) > 0$. Thus $V_T^1 > V_T^2 > V_T^3 > V_G$. The proof for $F > 4$ is analogous.

8.2 Proof of Proposition 2

The first-order condition is given by $\Delta \equiv (\tau + F\tau')(\phi pz - c) - (\rho + F\rho')\psi - M' = 0$. Note that $\tau' > 0$ and $\rho' > 0$. The second order condition $\Delta' < 0$ is always satisfied for M sufficiently concave (M'' sufficiently large). We thus have $\dfrac{dF}{d\psi} = \dfrac{1}{-\Delta'}\dfrac{d\Delta}{d\psi} = \dfrac{-\psi - F\rho'}{-\Delta'} < 0$, $\dfrac{dF}{dp} = \dfrac{\phi z(\tau + F\tau')}{-\Delta'} > 0$, $\dfrac{dF}{dp} = \dfrac{p\phi(\tau + F\tau')}{-\Delta'} > 0$ and $\dfrac{dF}{dc}\dfrac{-\tau - F\tau'}{-\Delta'} < 0$.

For the comparative statics of b^*, we note that there is a direct and an indirect effect. The direct effect is simply given by $\dfrac{\partial b^*}{\partial \psi} = \dfrac{\rho}{\tau}\dfrac{1}{\phi pz} > 0$. The indirect effect is given by $\dfrac{\partial b^*}{\partial F^*}\dfrac{\partial F^*}{\partial \psi} > 0$, where $\dfrac{\partial b^*}{\partial F^*} < 0$ since $\dfrac{\rho}{\tau}$ is decreasing in F, and $\dfrac{\partial F^*}{\partial \psi} < 0$ from the above proposition. The comparative statics for c, p and z follow the logic.

8.3 Proof of Proposition 4

For a given social value S, we ask whether markets or firms specialize in big project. We saw that markets are better than firms as long as $u_G = \gamma(\phi paz - c) - \psi \geq 0$. To keep the social value constant, we consider an increase in z and ψ that leaves S constant. Totally differentiating S, we obtain $\dfrac{d\psi}{dz} = \dfrac{\gamma\phi p}{1 - \overline{\phi}\delta}$. We use this, as well as $\dfrac{da}{dz} = \dfrac{-\delta c}{(2-\delta)z^2}$ to obtain after transformation $\dfrac{du_G}{dz} = \gamma\phi pz\dfrac{da}{dz} - \dfrac{\gamma\phi p(1-a)}{1 - \overline{\phi}\delta} < 0$.