Scapegoats and Optimal Allocation of Responsibility[1]

Eyal Winter[2]

Abstract

We consider a model of hierarchical organizations in which agents have the option of reducing the probability of failure by investing towards their decisions. A mechanism specifies a distribution of sanctions in case of failure across the levels of the hierarchy. It is said to be investment-inducing if it induces all agents to invest in equilibrium. It is said to be optimal if it does so at minimal total punishment. We characterize optimal investment-inducing mechanisms in several versions of our benchmark model. In particular we refer to the problem of allocating individuals with diverse qualifications to different levels of the hierarchy as well as allocating tasks of different importance across different hierarchy levels. We also address the issue of incentive-optimal hierarchy architectures.

1. Introduction

When an organization experiences a failure there is usually an urge to appoint blame at some level in the organizational hierarchy. Indeed, if the failure is clearly attributed to one or several individuals, then calling the responsible persons to account is a relatively easy task. But sometimes, indeed very often, associating wrong actions with the outcome of failure is ambiguous. Still, even in these cases the issue of responsibility is raised and sanctions are eventually imposed.

The problem of allocating responsibility in hierarchical organizations touches upon a variety of issues in economics, management, law and politics. It goes right to the definition of leadership and has an impact upon the incentives within organizations. Following repeated training accidents in the Israeli military during the mid- Nineties, several investigative committees were formed to look into the events that led up to some of these incidents. Part of the committees' mandate in this respect was to recommend punitive actions against individuals depending on the responsibility that they were

[1] I wish to thank seminar participants at Berkeley, Stanford, CORE, DELTA, EUI Florence, INSEAD Mannheim, Jerusalem and Tel Aviv for various comments and suggestions.

[2] The Economics Department and the Center for Rationality, The Hebrew University of Jerusalem. E-mail: mseyal@pluto.mscc.huji.ac.il.

carrying. On some occasions the committee's decision was followed by a public outcry that it failed to point the finger at a high enough level of the hierarchy. Some journalists even coined the phrase "Guard Syndrome" suggesting that the committee pushed the responsibility down to lower and lower levels - indeed, as low as the private who guarded the base's gate on the night of the event.

In the corporate world the allocation of responsibility seems to have a completely different pattern. Here, the tendency is much greater to go after the top of the pyramid, i.e. the CEO, in case of a failure. James and Soref (1981) have estimated that among America's 300 largest industrial firms in 1965 approximately 11 percent of the chief executives were dismissed in a given year. A 1999 article in the *Economist*[3] also claims that boards are getting soft fingers on the trigger when it comes to dismissing CEOs. It asserts that a third of the CEOs who ended their job in a Fortune 500 company between 1980 and 1996 were dismissed, often at substantial cost to the company in the form of lavish compensation packages and full option entitlements.

Whether it is an investigative committee or a board decision concerning a CEO, it is fair to assume that the allocation of sanctions following a failure is partly affected by factors that are not directly related to the optimal functioning of the organization. In the former case political pressures may result in looking for scapegoats and punishing only excessively low levels in the hierarchy, while in the latter concerns about shareholders' attitudes may lead to aiming only at the top.

In contrast this paper concerns itself with the issue of *optimal* allocation of responsibility (and sanctions). We will consider hierarchical structures in which decisions are taken by all levels of the hierarchy. The allocation of responsibility is reflected by the level of punishment that each level incurs following a failure. When discussing the optimality of the allocation we will consider two contravening factors: first, the threat of sanctions creates the incentive to exert effort or invest towards making the right decision, and this at each level of the hierarchy. Thus stiffer punishments will result in a greater incentive to invest and a lower probability of failure. On the other hand, sanctions and even the threat of sanctions are costly for the organization. The ex-post cost, when

[3] October 30 1999.

sanctions are imposed, is quite clear. In addition to the financial burden that replacement of top executives or high officials imposes on the organization, it involves the loss of experience and specific skills that have been acquired during their tenure as managers. However, the threat of sanctions is costly even when they are not realized, i.e., when no failure occurs. The threat of excessive punishment following a failure may result in a witch-hunt atmosphere within the organization, thereby reducing the incentives to collaborate or to take calculated risks. We will present a game theoretic model to discuss the issue of optimal allocation of responsibility in hierarchical organizations by taking into account the tradeoff described above. Specifically, we will attempt to design mechanisms that induce agents to invest towards their decision at a minimum threat of sanctions. What are the exogenous differences between levels of the hierarchy (i.e., between a boss and his subordinates) that we incorporate into our model? Admittedly, in reality bosses differ from their subordinates in terms of their available actions as well as in a wide range of other characteristics. Our model will not try to maximize on the number of such characteristics that it incorporates. On the contrary: we opt for parsimony by focusing only on one aspect to distinguish between levels of the hierarchy: namely, the flow of information concerning effort. Specifically, hierarchies in our model are characterized by the assumption that a boss can observe whether his subordinates have invested towards making their decision or performing their task, while subordinates have no access to similar information concerning their bosses. Indeed, this asymmetry affects incentives and thus also the optimal allocation of responsibility.

While we prefer to cast our results in the context of hierarchies, all our results can be interpreted outside this scope as well. In fact the only feature that the organization should possess in order for these results to be meaningful is that the information concerning exertion of effort is not symmetrically distributed across the agents in the organization. Consider for example a production process that involves an assembly line in which tasks are performed sequentially. The output of department j's task is shifted to department $j+1$ for further development. Here, the asymmetric information about exertion of effort is inherited in the order of moves in the sense that the decisions of early movers are revealed to those who move later. Interpreted in this fashion, our results will have implications about the allocation of responsibility between early and late movers in a sequential production process, and will also address the issue of comparing production processes from the point of view of optimal incentive.[4] We expand on this interpretation and its relation to hierarchies later in the discussion.

We will now illustrate the questions that we are interested in by presenting an elementary version of our model. Consider a cube consisting of 27 boxes arranged in three layers with three rows and three columns per layer. An object is hidden in one of the boxes with equal probability for each box. Three individuals are assigned by the principal to (jointly) locate the object. Player 1 has to determine the layer at which the object resides, player 2 the row, and player 3 the column. Each player can purchase the information concerning his correct decision at cost c. If he doesn't purchase the information he chooses one of the three options randomly. To locate the object the players now move in sequence. First, player 1 decides whether to purchase the information, then player 2 (observing player 1's decision) and then player 3 (observing the purchasing decisions of 1 and 2). Finally, players submit their location decisions, and the corresponding box is opened. If the object is found in the box each player receives a payoff of zero. If the object is missing (failure) sanctions v_1,v_2 and v_3 are imposed on the three players. Note that our informational definition of hierarchy implies that player 3 is at the top of the hierarchy while player 1 is at the bottom. Now, the principal seeks to design the mechanism of sanctions so that in equilibrium all players are induced to purchase the information, and in addition among all such mechanisms he searches for the one that yields minimal total sanctions. As we shall see in the sequel the only allocation of sanctions satisfying these conditions imposes a punishment of 1.038c on player 1, 1.125c on 2 and 1.5c on player 3.

Two remarks about the above example are in order: First, we comment that our general model will allow the purchased information to be noisy so that a failure (and thus also sanctions) may occur on the equilibrium path.

Second, note that the information structure that we assume to characterize the difference between bosses and subordinates imposes an exogenous order of moves in

[4] See also Sobel (1992) and Kremer (1993) who address similar issues in a non-strategic context.

which bosses act after their subordinates. As we will argue later this is the natural order of moves if all the underlying tasks are devoted to reaching a decision and no other activities.

Our general model will be developed in stages. We will start in Section 2 with the benchmark model in which a project has to be managed by n individuals at different hierarchy levels, each of whom has to perform a different task. Each individual has an option to invest towards his decision at a cost. Investment will increase the probability that his task is performed successfully from $\alpha \in [0,1]$ to 1. The project fails if at least one task fails, in which case sanctions are imposed on all agents.

Our first result will characterize the optimal investment inducing mechanisms, i.e., the allocation of sanctions that will induce all players to invest at a minimal total punishment. An interesting feature of the mechanism is the fact that people higher up in the hierarchy carry a heavier load of responsibility, i.e., they receive a stiffer punishment following a failure. In Section 3 we extend the model by allowing the cost of investment to vary across individuals. We interpret low cost as higher qualification, i.e., more qualified individuals need to exert less effort in order to increase the probability of their task's success. In this framework we require the mechanism to specify both the allocation of sanctions as well as the allocations of individuals across levels of the hierarchy. We show that the optimality of the mechanism implies that individuals with low investment cost should be assigned to higher levels of the hierarchy. A dual treatment is offered with respect to differential probability of task success α_i (i.e., absence of investment). We interpret tasks with low α_i as more important than ones with high α_i, since the investment decisions on such (low α_i) tasks are critical to the success of the whole project. Investment on tasks whose α_i is high has only a marginal effect on the success of the whole project. In this framework of differential success probability a mechanism has to specify the allocation of sanctions across levels in case of failure, as well as the allocation of tasks across the levels of the hierarchy. Concerning the latter, we show that in an optimal mechanism more critical tasks should be assigned to higher levels of the hierarchy.

We continue in Section 4 by adopting a slightly different approach, which allows the principal to bear part of the risk of the project's failure for the sake of reducing the total cost of deterrence. Here, the principal sets a target of success probability p and seeks the mechanism that minimizes the total punishment under the constraint that agents' (equilibrium) behavior leads to the project's success with probability of at least p. We show that if α (i.e., the probability that each task succeeds without investment) is low, then the principal cannot reduce the total sanctions by bearing part of the risk of the project's failure. Indeed, allowing for a positive probability of failure will require an exemption of some of the agents from responsibility. This will reduce the incentives of the others to invest, thereby requiring much harsher sanctions on the individuals who share responsibility. In contrast, for high α the principal can reduce the total sanctions if he allows for a higher probability of failure. The optimal mechanisms in this case are characterized in Section 4. We show that if the cost of sanctions increases with the hierarchy level, then the heaviest load of responsibility rests upon *intermediate* levels of the hierarchy, a feature that we find quite akin to the outcome of many investigative committees following a failure. Such a scapegoat effect balances between the fact that on the one hand incentive considerations require sanctions to increase with the level of the hierarchy and on the other sanction costs increase with the hierarchy level as well.

In Section 5 we investigate optimal architectures of hierarchy in our framework. In contrast to the benchmark model which requires a complete hierarchy we allow for general hierarchy structures represented by trees. The consequence of this is that the information structure cannot be described by a perfect information game as some tasks are performed simultaneously. We will show that optimal architecture imposes a complete hierarchy. Any other architecture will require an increase in the total sanctions. We argue, however, that other objectives that are not considered in our model (such as reducing the time it takes for information to flow across the hierarchy levels) may yield other architectures as optimal structures (see for example Keren and Levhari (1983), Radner (1993) and Hart and Moore (1999)).

In Section 6 we formulate a generalized version of our model by referring to general success probability functions, i.e., functions that map a profile of investment decisions to a probability of the project's success. This framework does not require that each individual be responsible for a different task, nor does it assume that the failure of a particular task results in the failure of the whole project. We argue in Section 6 that almost all our results (appropriately modified) hold in this more general framework. Finally, we conclude with a couple of remarks in Section 7. Proofs are relegated to the Appendix.

Models of hierarchical organizations have been introduced by various authors in the past. However, to the best of our knowledge none of these papers addresses the issue of responsibility allocation. One strand of this literature examines the engineering of hierarchical networks to optimize the information processing within organizations. In Sah and Stiglitz (1986) the role of individuals in the organization hierarchy is to evaluate and screen projects and the hierarchy design should take into account type-I and type-II errors. Radner (1992, 1993) as well as Keren and Levhari (1983) are some of the other classics within this literature, which mainly focuses on the time it takes for the information to flow across the different levels of the hierarchy. A more recent literature that views hierarchies as mechanisms for information aggregation is due to Van Zandt (1998, 1999). Some other papers address the related issue of authority in organizations. Aghion and Tirole (1997) discuss the allocation of authority in a model where the main tradeoff is between promoting initiatives by subordinates and losing control of their actions. Hart and Moore (1999) discuss optimal hierarchies within a model of ideas formation. Hierarchical organizations are also discussed by Rosen (1986), which views career paths as tournaments in which individuals compete for higher job levels within the organization. Rosen notes that benefits should increase with rank as a convex function (i.e., marginal benefits are increasing as well) in order to induce individuals to compete for higher ranks. The intuition behind this result is rather appealing: in addition to the benefits of the new job, winning the competition for a certain rank awards the winner also with the right to compete for higher ranks. As one moves higher up in the hierarchy the value of this right diminishes as there are less levels up the road to compete for. In order to maintain the incentive it is thus necessary to compensate with extra benefits. Interestingly, though for a completely different reason, our model predicts that sanctions should increase with rank as a convex function as well. Roughly speaking, in our framework sanctions should increase sufficiently fast with the level of the hierarchy to offset the effect that higher in the hierarchy agents' actions are monitored by less individuals thus increasing the incentive to shirk.

2. The Model

The organizational project involves n activities performed by n individuals (henceforth players) who are ordered in an increasing order according to their hierarchy position in the organization. That is, player i+1 supervises players i, i-1,…,1. The consequence of supervision is purely informational. i supervises j means that player i can observe the behavior of player j and in particular the effort that has been exerted by player j towards the performance of his/her activity but subordinates cannot similarly observe the behavior of their bosses. This relation dictates the order of moves in our sequential (extensive form) game. Players act sequentially in the order 1, 2, ... , n. Each player in his turn decides whether to invest (exerts effort) towards the performance of his activity. This investment can be interpreted as an acquisition of costly information relevant to that player's decision making. We denote by d_i the investment decision of player i.

$d_i \in \{0,1\}$, where 1 stands for a decision to invest and 0 stands for non-investment decision. The cost of investment in our benchmark model is c and is constant over all players. Each player before making his investment decision observes the decision of all his predecessors (i.e., his subordinates).

Each player's activity results in either success or failure. If player i invests, i.e., $d_i = 1$, then his activity is successful with probability 1. However, if $d_i = 0$, his success probability is $0 \leq \alpha \leq 1$, which again, for the time being, is constant over all players. The events that determine the success of the tasks are independent across the players and they occur after all players have made their decisions.

The project terminates successfully if and only if all activities have been performed successfully. If the project is successful all players receive a payoff of zero. If the project fails, then player i has to endure the punishment v_i, i.e., he receives a payment of $-v_i$. Thus players' punishments depend only on the project's realization and not on their investment decision, which we assume is unobservable by the principal. Players are all assumed to be risk-neutral. To summarize the payoff function of the game more formally, let $d = (d_1,\ldots,d_n) \in \{0,1\}^n$, denote the action combination taken by all players. We note that the payoff for player i depends on the combination of actions taken by all players. The payoff for player i when d is played is:

$f_i(d) = -v_i(1-\alpha^{s(d)})- c$ if $d_i = 1$ and

$f_i(d) = -v_i(1-\alpha^{s(d)})$ if $d_i = 0$,

where $s(d) = |\{j|\ d_j = 0\}|$ is the number of individuals who choose to shirk.

We will denote by G(v) the extensive form game induced by the vector of punishments $v = (v_1,v_2,\ldots,v_n)$. In the sequel we will analyze this game by means of its subgame perfect equilibria (SPE).

The responsibility allocator (henceforth the principal) wishes to design a mechanism that will induce all players to invest (in equilibrium). A mechanism is an allocation of punishments in case of a failure, i.e., a vector v. We say that the mechanism v is investment-inducing (INI) if all the SPEs of the game G(v) entail investment by all players, i.e., d = (1,…,1). In addition to inducing players to invest the principal attempts to achieve this goal with minimal punishment. We will say that an INI mechanism v is optimal[5] if $\Sigma_{i\in N}v'_i \geq \Sigma_{i\in N}v_i$ for every other INI mechanism v'.

Two comments about the model are in order:

We have assumed in this benchmark model that the investment by all agents guarantees the project's success with probability 1. As we shall see later this assumption is inessential. Our general model in Section 6 allows the project to fail also when all agents invest, implying that sanctions may be inflicted also along the equilibrium path, and explains why the principal wishes to reduce it. We also note that instead of imposing

[5] In the sequel we will use the term "optimal mechanism" also when the vector of minimal punishments induces investment in some but not all SPEs as long as this vector is arbitrarily close to a mechanism in which investment is the *unique* SPE. We need this technical caveat because punishments take continuous values.

sanctions following failure we could have required the mechanism to allocate rewards when the project succeeds (and zero payoffs in case of failure) and then search for the mechanism that minimizes total spending (rewards) on part of the principal, binding to the fact that it induces all agents to invest. This dual model leads to the same results as the one based on sanctions.

An important assumption in our model is that sanctions can be contingent only on the event of failure, not on the agents' investment decision. Furthermore, while each investment decision of an agent is observable by his superiors, such an action is not verifiable (or contractable). This meets our intuition that it is often hard to produce evidence that a certain agent did not exert effort, while it is much easier to prove that the project has failed. Furthermore, in many environments sanctions are imposed by an authority which is external to the organization (e.g., a board or an investigative committee). Information about agents' devotion can only be supplied by the agents themselves, but coming from an interested party, this information is doomed to be unreliable. This distinction between observable and contractable actions is standard in the contract theory literature (see for example Holmstrom (1982), Hart and Moore (1990), Shleifer and Vishny (1997)).

We start with a simple characterization of optimal INI mechanisms in our framework:

Proposition 1: A mechanism v is an optimal investment-inducing mechanism if and only if $v_j = c/(1-\alpha^{n-j+1})$.

Note the following comparative statics observations that are implied by Proposition 1. First and most obvious is the fact that punishments increase with c. This is simply because the incentive to shirk increases with c. Second, punishments are increasing with α. Again as α increases the risk involving no investment reduces and shirking becomes more attractive. This is being offset by a higher punishment. But perhaps the most interesting observation is the fact that v_j is increasing in j and that as a function of the hierarchy level the sanctions increase as a convex function (i.e., marginal punishment increases with j). Namely, players higher up in the hierarchy are punished more severely than lower-ranking players. This follows from the particular information structure by which information concerning investment decision flows in one direction only. Players higher up in the hierarchy should be provided this extra incentive in terms of greater punishment because the information structure provides them with more opportunity to free ride on the investment decision of players lower in the hierarchy. More specifically, as we will see later, the equilibrium strategies in the optimal mechanism prescribe each agent to invest if and only if he observed all his subordinates investing. This gives agents low in the hierarchy (who appear early in the order of moves) less incentive to shirk than agents higher up in the hierarchy, which implies that the optimal mechanism can ease sanctions on some of these agents. Note that this result relies solely on the information structure without assuming any other asymmetries between the players.

Note that the assembly line interpretation of this result suggests that agents who have to act at a later stage of the production process should be provided with stronger incentives. Marketing is a good example for such a task, and indeed one which is typically very well rewarded.

Our benchmark model has assumed that the principal wishes to guarantee that the project terminates successfully with probability 1. This has led to an optimal mechanism in which punishment increases with rank. In the sequel we will show that we obtain a different result if the principal is allowed to bear part of the risk of failure for the sake of further reducing the total punishment, which will give rise to the scapegoat effect. But before getting to this twist in the model, we will discuss two simple extensions, from which we can draw two interesting results concerning the allocation of activities and individuals across the different levels of the hierarchy. These results will mainly serve as a test for the relevance and adequacy of our proposed model.

3. Competence, Importance and Hierarchy

Following the tradition of the signaling literature since Spence (1974), we define a player's *competence* as his or her cost of investment. This definition becomes particularly relevant when we think of the performed activity in terms of acquisition of information. If one agent is more socially connected or more computer-literate than another agent he will reasonably be able to extract the same information with less effort. This of course will affect his incentive to invest. In contrast to the benchmark model we will now assume that players have differential investment cost, i.e., the game will be characterized by a vector $c = (c_1,\dots,c_n)$. For simplicity we assume that c_i are distinct and $c_1 > c_2,\dots,> c_n$ and the payoff function for player i is changing by replacing c with c_i in the payoff function (here the index i does not necessarily correspond to the hierarchy level).

The principal's task is now to determine the allocation of individuals to different levels of the hierarchy and a vector of punishments. Again, the principal wishes to induce all players to invest at a minimal total punishment. Formally, a mechanism is now a pair m = (w,v) where w is a permutation of N = {1,2,…,n} (representing the assignment of players to levels of the hierarchy) and v is a vector of punishments. For a mechanism m and a vector of costs c we denote by $G_{m,c}$ the game described in Section 2 with respect to c and m. We say that m is an optimal INI mechanism if every SPE of $G_{m,c}$ leads to a probability 1 success and moreover there exists no other mechanism m' = (w',v') such that all SPEs of $G_{m',c}$ lead to a probability 1 success and $\Sigma_{i\in N}v_i > \Sigma_{i\in N}v'_i$.

We can now state the following:

Proposition 2: In the differential costs model, m = (w,v) is an optimal investment-inducing mechanism if and only if w is the identity permutation, (i.e., lower cost individuals are assigned to higher levels of the hierarchy) and $v_j = c_j/(1-\alpha^{n-j+1})$.

According to Proposition 2 the principal can achieve minimal punishment in an investment-inducing mechanism only when he allocates the more competent individuals to higher-level positions. One should not be confused by the terminology leading to Proposition 2. We note that the intuitive property that it exposes applies without assuming any differences in terms of the complexity of the performance of activities at different levels of the hierarchy or in terms of their effect on the overall success of the project. Proposition 2 follows from the fact that the distortion factor $1/(1-\alpha^{n-j+1})$ increases with the rank, which implies that negative assortative assignment between cost and rank (i.e., lower-cost agents assigned to higher ranks) lowers the total sanctions.

We now turn to the second extension of our benchmark model by allowing the non-investment success probability to differ across players. We now assume that without investing player i manages his task with probability α_i. As before, we assume that investment guarantees that i's task terminates successfully with probability 1. Again, for simplicity we assume that $\alpha_i > \alpha_{i+1}$. We interpret α_i as an indicator for the importance of player i's task to the project as a whole. The values of α_i in the spectrum range from the duties of cleaner in the company (α close to 1) to the CEO's task of designing company strategy at the other extreme. If α_i is close to 1, then the outcome of i's task has little effect on the success probability of the whole project. Low probability α_i represents a task at key positions for which failure will have crucial implications for the project. To be able to isolate the effect of differential α_i we will assume here that the investment costs are identical across all players. The mechanism in this framework is a pair $m = (\theta,v)$ such that θ is a permutation on N specifying the allocation of tasks to different levels of the hierarchy and v is a vector of punishments. The same definition of optimal investment-inducing mechanisms that we discussed prior to Proposition 2 applies here. We can now show that:

Proposition 3: In the model with differential probabilities of success, $m = (\theta,v)$ is an optimal investment-inducing mechanism if and only if θ is the identity permutation (i.e., tasks with lower α are assigned to higher hierarchy levels) and $v_j = c/(1-\prod_{k=j}^{n} \alpha_k)$.

Proposition 3 asserts that in order to achieve minimal punishment in an investment-inducing mechanism the principal must allocate sensitive tasks to higher levels of the hierarchy. The intuition is most apparent when comparing the allocation of tasks suggested in the proposition to the inverse allocation, i.e., when less important tasks are allocated to higher levels of the hierarchy. When an agents shirks in the optimal allocation he triggers the shirking of other agents all of whom deal with more important tasks. In contrast, with the inverse allocation it triggers the shirking of agents dealing with less important tasks. Being less detrimental the latter creates more incentive to shirk and requires higher penalty to guarantee investment. For the more general case, consider two tasks A and B, where A has a higher probability of success under no investment. Suppose that the investment decision on task A is taken before the one on task B (i.e., assigned to a lower level of the hierarchy). Along the equilibrium path following non-investment by some player all subsequent players choose to shirk. If we exchange the tasks between the two individuals so that now task B (the more important one) is performed by the player lower in the hierarchy it will not affect the incentive of this player to invest because shirking will induce precisely the same set of players to shirk as before, yielding the same probability of failure of the project. In contrast, the incentive of the high-rank individual to invest is reduced requiring a higher penalty to induce this player to invest. Again, our model sustains this desirable property without assuming that higher level positions are occupied by more competent individuals. It follows merely from the information structure on which our model builds.

4. Reducing Punishment at the Cost of Lower Success Probability

The threat of punishment, as we already argued in the introduction, in addition to creating the incentives to invest, has a negative impact on corporate culture. If this effect aggravates, the principal may wish to reduce punishment below the level specified by the optimal incentive-inducing mechanism. This may be achieved only at the cost of reducing the probability of success to below 1, i.e., by allowing some of the players to shirk. In this section we will be concerned with mechanisms in which the principal is prepared to bear part of the risk. In this analysis we will return to the benchmark model in which the α_i's and the c_i are the same across players:

For a probability $0 \leq p \leq 1$ we say that a mechanism v is p-investment-inducing if every SPE of the game G (which depends on α and c) leads to the project's success with probability of at least p. More precisely, $\alpha^{s(d)} \geq p$ (recall from Section 2 that s(d) is the number of shirking individuals). A mechanism v is said to be optimal p-investment-inducing if it uses the smallest total punishment among all p-investment-inducing mechanisms.

The first thing we need to check is the way reducing punishment on one agent affects the investment incentive of other agents. If it reduces other agents' incentive to invest substantially it may be optimal to induce all agents to invest even if the principal is willing to tolerate a certain probability of failure. In fact, this is exactly the case when α is small enough.

Proposition 4: If α is small enough, then for any $\alpha^n < p < 1$ v is an optimal p-investment -inducing if and only if it is optimal investment-inducing (i.e., for p = 1).

Proposition 4 carries a concrete message. If α is small enough, i.e., if the successful outcome of each task strongly relies on investment, then the principal has to punish everybody even if he is willing to bear a large part of the risk. Allowing some individuals to shirk will drastically reduce the incentive of the rest to invest. This, in turn, will force the principal to threaten the rest of the people with a substantial extra punishment to the extent of making the total punishment greater than is required to induce all individuals to invest.

The situation becomes different, however, when α is close enough to 1. α close to 1 reflects situations in which the investment has a relatively low effect on the project's success. Particularly in these cases the principal may find it desirable to reduce total punishment at the cost of a small marginal sacrifice in the success probability.

For each $1 \leq k \leq n$, we will denote by p(k) the probability of success when exactly k individuals invest, i.e., $p(k) = \alpha^{n-k}$. It is clear that by reducing the total punishment the principal will affect the success probability in grids of p(k) k = 1, 2, ... , n as this probability is only affected by the number of investors. We will thus consider optimal investment inducing mechanisms with respect to these probabilities.

Proposition 5: Let α be close enough to 1 and $1 \leq k \leq n$. The mechanism v is an optimal p(k)-investment-inducing mechanism if and only if v is of the following form: the principal selects a group of k players $K = \{i_1, i_2, \ldots, i_k\}$ with $i_1 < i_2, \ldots < i_k$ (which will be induced to invest) . $v_{i_j} = c/(\alpha^{n-k} - \alpha^{n-j+1})$ and $v_r = 0$ for $r \in N \backslash K$. Furthermore, the total punishment in an optimal p(k) – investment-inducing mechanism is increasing with k.

If the principal is willing to settle for a success probability p(k) lower than 1, then he can do so by allocating the responsibility among k individuals only. Proposition 5, in particular, says that the choice of these agents can be made arbitrarily, i.e., the total punishment will not depend on this choice. This is because the punishment imposed on each agent does not depend on his global position within the hierarchy but only on his level relative to the other individuals who share responsibility. This observation has interesting implications when we allow the cost of punishment for the principal to vary across different levels of the hierarchy. We have argued in the introduction that imposing a punishment on agents often becomes more costly as we move higher up in the hierarchy. This suggests the following simple modification of our benchmark model: for a level i in the hierarchy and a punishment v, let $g_i(v)$ be the cost of imposing punishment v at level i. We assume that $g_i(v) \geq g_j(v)$ for all v, whenever $i > j$. The definition of investment inducing mechanism is the same as before except that now the principal seeks to minimize $\Sigma_{i \in N}\, g_i(v_i)$. We refer to this version as the increasing-punishment-cost model.

Proposition 6: Let α be close enough to 1 and $1 \leq k \leq n$. In the model with increasing punishment costs $(v_1,...,v_n)$ is an optimal p(k)- investment-inducing mechanism if and only if $v_j = c/(\alpha^{n-k}-\alpha^{n-j+1})$ for j = (1, 2,..., k) and $v_j = 0$ for j = (k+1, ..., n).

Under the conditions specified in Proposition 6 optimal mechanisms are of the following form: first a level k is chosen (depending on the targeted probability of success). The principal imposes punishment only on agents of rank lower than k, but among these agents the punishment is going to be increasing in rank. This property of the mechanisms seems to be compatible with the stylized fact that the heaviest burden of responsibility often lies on the shoulders of middle ranks of the hierarchy. This lack of punishment for levels high up in the hierarchy may be interpreted as "scapegoating." However, it should be argued here that the tradeoff between the total punishment level and the security level of the project should depend on the importance of the project and its relation to other projects that are performed within the same organization. In fact each project can be thought of as one of several tasks performed simultaneously. Adopting the military analogy again, if the project's failure corresponds to the event of finding a bug in the soup at the kitchen of some military barrack, then it may require the unit cook to bear the consequences but it would be ridiculous to get rid of the general for that. On the other hand a defeat in a major battle may require calling to account higher-ranking persons in the hierarchy, perhaps even as far up as the top of the pyramid.

We summarize this section by pointing out that the security level p (i.e., the probability of success) that has been taken exogenously here can be determined endogenously by assuming that the principal attempts to maximize the expected net value of the project. Specifically, suppose that the gross value of the project is U and that the total punishment in the optimal p-INI mechanism is v(p). Then p is determined by maximizing; $pU + (1-p)v(p)$. But regardless of U and the resulting probability of success the optimal mechanism is as described in Proposition 6.

5. Architectures of Hierarchies

In our analysis of optimal INI mechanisms we have assumed a specific architecture for the hierarchy structure: namely, every two tasks are performed by individuals of different hierarchy levels. We will refer to this structure as the complete hierarchy structure. In this section we will be interested in comparing different hierarchy architectures by considering general (tree-type) hierarchy structure. One of the questions we wish to answer is the following: in his attempt to optimize the level of punishment in INI mechanisms, would the principal do better by choosing a hierarchy architecture which is different from the complete hierarchy? We will show that the answer to this question is negative, i.e., the optimal architecture from the point of view of investment incentive is the complete hierarchy.

To demonstrate the consequences of changing the hierarchy architecture on individuals' strategic considerations, let us examine the simplest case of two tasks. If the hierarchy is complete, then Proposition 1 asserts that the optimal total punishment that guarantees success with probability 1 is $c/(1-\alpha) + c/(1-\alpha^2)$. Suppose now, by contrast, that the two tasks are performed by individuals at the same hierarchy level, which means that none of the agents is informed about the investment decision of the other. A mechanism in this case is a normal form game in which the two individuals decide on investment simultaneously. To ensure that investment by both players forms an equilibrium in such a game the sanction v imposed on the players in case of failure should satisfy the equation $-c \geq -v(1-\alpha)$. So the total punishment required to get both players contributing in equilibrium is $2c/(1-\alpha)$, which is greater than the one required to sustain success with a complete hierarchy. Indeed, in order to get investment by all players sustained as a *unique* equilibrium we need yet a greater total punishment. This shows the superiority of the complete hierarchy in this example. We will generalize this observation now by considering general hierarchy structures represented by trees, which will formally be described by a partial order on the agents.

A general hierarchy structure (GHS) is a partial order "*h*" on the set of agents N. For two agents i and j we use i *h* j to denote that i supervises j (or i is the boss of j). To represent hierarchies the order must satisfy the following conditions:

1. *h* is anti-symmetric: i *h* j implies ¬[j *h* i] (meaning j is not the boss of i) .
2. *h* is anti-reflexive : ¬[i *h* i], and
3. *h* is transitive: i *h* j and j *h* k imply i *h* k.

Note that hierarchy structures that can be represented by a tree fall within this definition. However, our general definition of hierarchies allows for much more than trees. In particular, it allows agents to have more than a single *direct* supervisor.

For each $i \in N$ we denote by $B_i(h)$ the set of i's bosses according to the GHS *h*, i.e., $B_i = \{j; j\ h\ i\}$.

To be able to compare hierarchy structures we will define a partial order on structures that uses the sets B_i defined above[6]: Let *h* and *h*' be two GHS's. We say that the structure *h* is more hierarchical than *h*' if for all $i \in N$ $B_i(h') \subset B_i(h)$ with at least one strict inclusion. With respect to this partial order on hierarchies, the complete hierarchy in our benchmark model (i.e., 1 is the boss of 2, who is the boss of 3, etc.) is a maximal element while the structure in which all agents operate on the same hierarchy level is a minimal (least hierarchical) element. In Figure 1 we demonstrate three hierarchy structures. Note that h_1 and h_2 are incomparable in terms of the partial order defined above[7], but h_3 is more hierarchical than both h_1 and h_2 .

As before, we assume that each agent observes the investment decision of his subordinates but does not observe the actions of the rest of the players. In a GHS, such information structures give rise to extensive form games with imperfect information. In fact, the same GHS can be represented by more than one extensive form game, but all the extensive form games of a GHS correspond to the same normal form game. It is therefore convenient to analyze the mechanisms in their normal form. Specifically, for each GHS *h* we consider the normal form game G(*h*) in which each strategy of player i is a mapping from the investment decisions of his subordinates to an action in {0,1}, and the payoff

---

[6] We note that an equivalent definition can be generated by referring to subordinates rather than bosses.

[7] $2 \in B_5(h_1)$ and $2 \notin B_5(h_2)$ so h_2 is not more hierarchical than h_1. On the other hand, $3 \in B_5(h_2)$ and $3 \notin$

function is given by the vector of sanctions and the cost of investment as before.

The normal form games representing general hierarchy structures may be rather complex and they may possess multiple equilibria. We shall therefore limit ourselves to comparing hierarchy structures in terms of the minimal total sanction necessary to induce all players to invest in at least one equilibrium (without attempting to compare sets of equilibria).

For a GHS *h* we denote by v(*h*) the minimal total sanction that is sufficient to induce all agents to invest in some equilibrium of G(*h*).

Proposition 7 asserts that more hierarchical organizations require lower sanctions to induce all agents to invest in equilibrium. Specifically,

Proposition 7: If *h* is more hierarchical than *h*' according to the partial order defined above, then v(*h*) < v(*h*').

Since the optimal mechanism for the complete hierarchy does not possess multiple equilibria we can compare this structure with the others also in terms of the total punishment in the optimal mechanism. Denoting by υ(h) the total punishment in the optimal mechanism for the hierarchy h we have:

Corollary: Let h* be the complete hierarchy structure, and let h be any other hierarchy structure. Then in optimal INI mechanisms of h and h* the total punishment υ(h) exceeds that of υ(h*).

More hierarchical architectures allow for more visibility within the organization in the sense that the investment decision of each agent is observed by more individuals. Proposition 7 says that in such organizations it is easier to induce individuals to invest. The intuition in quite clear. Since shirking by an agent triggers those observing him to shirk as well (in the optimal INI mechanism), there is a greater internal threat (in terms of the decline of the success probability) when agents are more visible. This allows the principal to sustain investment with smaller penalties. Proposition 7 can be interpreted

---

$B_5(h_1)$, meaning that h_1 is not more hierarchical than h_2.

beyond the scope of formal hierarchies (i.e., bosses and subordinates) by taking into account all barriers to visibility between agents, including for example the fact that, say, i has a visual contact to j's office but not vice versa. Interpreted thus, Proposition 7 might suggest why accommodating employees in a large hall that allows some mutual eye contact is superior to the solution of private offices.[8]

We emphasize that although Proposition 7 and its corollary assert the optimality of complete hierarchies in our model, there may be other considerations, outside our model, under which complete hierarchies may not be optimal. In particular, complete hierarchies are rather ineffective in terms of the time it takes for the organization to process information. See for example Keren and Levhari (1983), Radner (1993) and McAfee and McMillan (1995) for models of this sort.

6. General Success Functions

Our framework until now has been characterized by the fact that each agent is performing a different task (or responsible for a different decision) whose success probability depends only on his own investment decision without any externality effects. It turns out however that most of our results hold true in a much more general framework, which allows the probability of the project's success to depend on the investment decisions in a general fashion. Specifically, we consider a function p from $\{0,1\}^N$ to [0,1] that specifies the probability of success p(d) for every combination of actions d in $\{0,1\}^N$. Setting $S = \{j; d_j = 1\}$ we will view p as a function that associates a probability to each subset of N (i.e., the set of agents who exert effort). We impose that the success probability increases as more players exert effort, i.e., $T \subset S$ implies $p(T) < p(S)$. We also assume that the success function possesses increasing returns to scale, i.e., $p(S\cup\{i\}) - p(S) > p(T\cup\{i\}) - p(T)$, whenever $T \subset S$. This condition can be interpreted as complementarily between players in terms of the effect of their investment on the probability of success (see also

[8] We also note that the superiority of the complete hierarchy structure may be related to the fact that in some public good environments sequential mechanisms are superior to simultaneous mechanisms in implementing efficient outcomes (see for example Bag and Winter (1999)).

Segal (2000) who defines a similar property in the context of exclusive/inclusive contracts and in the context of natural monopolies in Winter (1994)). Note that our benchmark model of independent tasks satisfies this condition. It is also interesting to note that if we amend the benchmark model by requiring that investment increase the probability that the task succeeds from α to $\beta > \alpha$ (instead of increasing it to 1) then this model is a special case of the one described above. At the end of this section we will also briefly discuss the case of decreasing returns to scale.

As before, a mechanism specifies a vector of punishments $v = v_1, ..., v_n$ that are imposed if the project ends in failure. With c being the cost of effort the payoff function is now given by $f_i(d) = -c-v_i(1-p(S))$ if $d_i = 1$ and $f_i(d) = -v_i(1-p(S))$ if $d_i = 0$. Optimal mechanisms are now defined in the same manner as before.

With basically the same arguments used before we can now derive the following results:

Proposition 1*: A mechanism v is an optimal investment-inducing mechanism if and only if $v_j = c/(p(N)-p(S_j))$, where S_j are the set of players of lower rank than player j.

Furthermore, if the success probability has increasing returns to scale or constant returns to scale (i.e. if $p(S_{j+1}) - p(S_j)$ is either increasing in j or constant in j) then marginal punishment must increase with j, i.e., the punishment function is convex.

Proposition 2*: Suppose that i and j are symmetric with respect to p (i.e. $p(S\cup\{i\}) = p(S\cup\{j\})$) and $c_i > c_j$. Then m = (w,v) is an optimal investment-inducing mechanism if and only if j's rank is higher than i's and $v_j = c_j/(P(N)-p(S_j))$.

For Proposition 3 we note that with general success functions we do not have a structure in which each individual is dealing with a different task. However, we can define a partial order over the importance of positions in the organization. Specifically, we will say that i's position is more important than j's position if by investing i is more capable than j of increasing the success probability regardless of the configuration of investment by other players. More formally, when $p(S\cup\{i\}) > p(S\cup\{j\})$ for all $S \subseteq N\backslash\{i,j\}$. Suppose now that in addition to fixing the sanctions a mechanism has to specify the allocation of positions to levels of the hierarchy; we then obtain the following parallel to Proposition 3:

Proposition 3*: Assume that the order of importance of positions is complete so player i's position is more important than player i+1's. (i= 1,2,...,n-1). Then (θ ,v) is an optimal investment-inducing mechanism if and only if θ is the identity permutation, i.e., it assigns higher positions to higher levels of the hierarchy and $v_j = c/(p(N)-p(S_j))$.

The proof of Proposition 3* is given in the Appendix.

Propositions 4, 5 and 6 which deal with INI mechanisms in which the planner bears part of the risk of failure cannot be extended in a straightforward way because their statement relies on the value of the parameter α which does not exist in the general case. Nevertheless, we can identify sufficient conditions on the function p, which imply that the optimal INI mechanism is of the same nature as hat described in Proposition 5 and Proposition 6.

We say that p is symmetric if p(S) depends only on the number of agents in S. For each $1 \le k \le n$, p(k) denotes the probability of failure when exactly k players invest. For each $1\le k \le n$ and $m \le k$ we denote $D(k,m) = p(k) - p(k-m)$. Our condition of increasing returns to scale implies that D(k,m) is increasing in k. But for Proposition 5* we will need D(k,m) to increase moderately enough. Specifically, denote $\Delta = \max_{k,m,r} D(k+r,m) - D(k,m)$, where k,m,r satisfy $1\le k \le n$, $m \le k$ and $k+r \le n$.

Denote $d_+ = p(n) - p(0)$ and $d_- = \min_k [p(k+1) - p(k)] = p(1) - p(0)$. For Proposition 5* we require that returns to scale increase moderately (or alternatively that p is close enough to a constant returns to scale function). Intuitively, with returns to scale increasing sharply exemption from sanctions due to failure becomes very costly and will place a substantial load on those who share responsibility. We formulate “moderate” increasing returns to scale by setting a bound on Δ. Specifically, we require that

$\Delta < d_-^2/nd_+$.

Proposition 5*: Suppose that p is symmetric and that $\Delta < d_-^2/nd_+$. Then, for $1 \le k \le n$ the mechanism v is an optimal q(k)-investment-inducing mechanism if and only if v is of the following form: the principal selects a group of k players $K = \{i_1, i_2, \ldots, i_k\}$ with $i_1 < i_2, \ldots < i_k$ (which will be induced to invest) . $v_{ij} = c/(p(k)-p(j-1))$ and $v_r = 0$ for $r \in N \backslash K$. Furthermore the total punishment in an optimal q(k) -investment-inducing mechanism is increasing with k.

The proof of Proposition 5* is given in the Appendix. The parallel of Proposition 6 now follows immediately:

Proposition 6*: Suppose that p is symmetric and that $\Delta < d_-^2/nd_+$. Let $1 \le k \le n$. In the model with increasing punishment costs, $v = (v_1,...,v_n)$ is an optimal q(k)- investment inducing mechanism that induces k players to invest if and only if $v_j = c/(p(k)-p(j-1))$ for j = (1,2,...,k) and $v_j = 0$ for j = (k+1,...,n).

Finally, using the monotonicity property of success functions, arguments similar to those in the proof of Proposition 7 now yield:

Proposition 7*: Consider any success probability p and let π^* be the complete hierarchy structure. Let π be any other hierarchy structure. Then in optimal INI mechanisms of π and π^* the total punishment $v(\pi)$ exceeds that of $v(\pi^*)$.

We note that without increasing returns to scale the results may be different. Consider the following example in which the marginal success probability declines as more players invest: There are three agents with a symmetric success function given by p(0) = 0, p(1) = 1/3, p(2) = ½, and p(3) = 7/12. Thus the first investor increases the success probability by 1/3, the second by 1/6, and the third by 1/12. Because of the decreasing returns players' incentives to invest decline as more players contribute. For example, if 1 and 2 invest, then player 3 will invest only if his sanction is at least 12c, while he can be induced to contribute at a sanction of 3c if 1 and 2 shirk. Consequently, the optimal mechanism imposes a uniform sanction of 12c on all players. Under this mechanism each agent chooses to invest even if he is the only one investing.

Note that with decreasing returns the hierarchy architecture has no effect on the sanctions imposed by the optimal mechanism. To induce all players to invest, sanctions have to be at least 12c on each player even when all players operate on the same level of the hierarchy, i.e., when they make the investment decision simultaneously. This observation, which can be generalized for decreasing returns to scale technologies, has the following interesting testable implication. When the technology has increasing returns to scale the hierarchy structure matters. And, at least as far as the incentive to invest is concerned, more hierarchical structures are more advantageous. In contrast, when returns to scale are decreasing this advantage vanishes, and all hierarchy structures are equally effective in inducing agents to invest.

7. Concluding Remarks

1. We have used a mechanism design approach to address the issue of optimal allocation of responsibility in organizations. As we mentioned earlier, our model lends itself to a dual interpretation: first one in the context of hierarchies by assuming that bosses can observe the investment decision of their subordinates; second in a more general context where the asymmetric information among the agents stems from other features of the organization, e.g., the assembly line interpretation. In many situations these two interpretations coincide. Indeed, while we sometimes tend to think of bosses as acting before their subordinates, if we think of tasks in terms of reaching a decision, the order of moves is obviously the other way around. Consider for example a firm whose project involves reaching a decision about whether to launch a new product or not. The natural process of such decision making involves information flowing from lower levels of the hierarchy upwards. The marketing department will have to conduct a market survey in

order to estimate the prospects of the product to sell well, and the R&D department will have to determine whether the technology is mature enough to allow for a reliable usage of the product. Only after aggregating the recommendations of all subordinates will the CEO ready to come up with the final decision, i.e., to launch or not to launch.

2. We have assumed an information structure under which a boss can monitor all his subordinates. However, we would have obtained the same results if we had assumed that a boss can only monitor the behavior of his immediate subordinate. This would of course mean that the mechanism involves an extensive form game with imperfect information (i.e., with non-trivial information sets) but subgame perfection applies here as well and the results are the same. Along the equilibrium path of the optimal mechanism here, a player decides to shirk if and only if his immediate subordinate shirks. This is enough to create the domino effect that induces all subsequent players to shirk, and hence the necessary deterrence against shirking.

3. A fundamental assumption in such a framework is that the rules of the game or the mechanism is commonly known by all the agents. But how should we interpret this assumption in real life? Should we think of such mechanisms as taking the form of clear legislation specifying the consequences of failure in governmental organizations? Do we have to interpret them as clauses in contracts between companies and executives specifying conditions under which dismissal can take place? Not necessarily! If we interpret the sanctions v_j as the expected punishment imposed on level j, then we can think of a mechanism as representing the organizational *culture* (or the corporate culture) concerning the allocation of responsibility. Following a failure, the imposition of sanctions doesn't have to be deterministic. Different levels of the hierarchy may experience sanctions with different severity but also with different likelihood. The organizational culture is optimal if it induces expected punishment which is compatible with our optimal mechanism. This interpretation is of course more adequate for organizations with sufficiently long history that allows agents to learn the culture through the recollection of past events.

Appendix

Proof of Proposition 1: First note that v_j as specified in Proposition 1 satisfy the following equation:

$$-c = -v_j (1-\alpha^{n-j+1}) \qquad (1)$$

The LHS of (1) corresponds to the payoff for player j when he and all other players contribute, while the RHS corresponds to j's payoff when all players preceding by j contribute and all his followers shirk. Thus under the assumption that all his followers will contribute when j himself contributes, j is indifferent between contributing and shirking. Further, we argue that in an SPE if player j has to act after some players have chosen not to contribute, then player j will choose $s_j = 0$. Indeed, let r be the number of players preceding j who did not contribute; then from (1) and the fact that $\alpha < 1$ we can obtain

$$-c - v_j(1-\alpha^{r+n-j}) < -v_j(1- \alpha^{n-j+1+r}) \qquad (2)$$

Using (2) inductively we see that its LHS stands for player j's payoff when he contributes while the RHS stands for his payoff when he shirks. This means that the following behavior forms an SPE: at any subgame after which one or more players have chosen to shirk all subsequent players will shirk as well, and at any subgame following contributions by all acting players all remaining players will contribute as well. Note also that if we slightly increase the punishments above v_j then the corresponding game has a unique SPE which leads to investment by all player (with the LHS of (1) becoming greater than the RHS). To show that v is an optimal INI mechanism we will show that any punishment vector $v^* = (v_1^*,...,v_n^*)$ with $v_j^* < v_j$ for some j, the corresponding game has no SPE with investment by all agents. Indeed, assume by way of contradiction that such vector v* exists and consider the SPE at which it yields investment by all players. Assume that j is the player with the largest index for which $v_j^* < v_j$. Consider a deviation by player j at which he shirks instead of investing. Let k be the number of players who shirk along the off equilibrium path following j's deviation. The player j's payoff is $-v_j^*(1-\alpha^k)$. By equation (1), the fact that $k \leq n-j+1$, and our assumption on v_j^* we have:

$$-v_j^*(1-\alpha^k) > -v_j(1-\alpha^k) \geq -v_j(1-\alpha^{n-j+1}) = -c.$$

This shows that j's deviation is profitable and establishes the contradiction.

Finally, note that with the optimal mechanism v players are indifferent between investing and shirking but any arbitrarily small increase in the punishment breaks this indifference and yields a unique SPE (see Footnote 5).

Proof of Proposition 2: For a fixed permutation w assigning agents into levels of the hierarchy an optimal INI mechanism satisfies $v_j = c_{w(j)} /(1-\alpha^{n-j+1})$ for j = 1, 2,..., n, where $c_{w(j)}$ is the investment cost of the individual assigned to level j of the hierarchy. This follows from the same arguments used in the proof of Proposition 1 by replacing c with c_j in equations (1) and (2) of that proof. To determine the optimal assignment, consider the one that corresponds to the identity permutation. Let j denote the level of the hierarchy so we have $c_i > c_j$ whenever j > i. Set $b_j = 1/(1-\alpha^{n-k+1})$ for k = 1, 2,..., n and note that for j > i $b_j > b_i$. Setting $b_j = b_i + \delta$ and $c_i = c_j + \varepsilon$ we get that the total punishment for i and j in the assignment w is $b_jc_j + b_ic_i = (b_i+\delta)c_j + (c_j + \varepsilon)b_i = 2b_ic_j + \delta c_j + \varepsilon b_i$.

Consider now the permutation w’ in which i and j exchange positions. The total punishment for i and j will now be $(b_i+\delta)(c_j + \varepsilon) + b_ic_j = 2b_ic_j + \delta c_j + \varepsilon b_i + \delta\varepsilon$. Since the punishments for all other players have not changed by moving from w to w’ we conclude that w’ requires an excess total punishment of $\delta\varepsilon$. Consider now an arbitrary permutation w* which is different from w. By successive binary exchange of positions of the sort performed above we can move in stages from w* to w reducing the total cost at each stage of this process. This implies that w is the optimal permutation.

Proof of Proposition 3: For a given permutation θ of tasks into different levels of the hierarchy equation (1) is generalized to be $-c = -v_j (1-\alpha_{\theta(j)} \alpha_{\theta(j+1)},..., \alpha_{\theta(n)})$, where $\alpha_{\theta(j)}$ is the parameter of the task θ(j) assigned to level j of the hierarchy. Using the same arguments as in Proposition 1 we get that $v_j = c/(1-\alpha_{\theta(j)} \alpha_{\theta(j+1)},..., \alpha_{\theta(n)})$ for the optimal INI for a given θ. Consider now the identity permutation. For each i,j with j > i we have $\alpha_j < \alpha_i$. For this permutation $v_j = c/(1-\alpha_j \alpha_{j+1},..., \alpha_n)$ and $v_i = c/(1-\alpha_i\alpha_{i+1},...,\alpha_{j-1}\alpha_j,..., \alpha_n)$. Consider now a permutation θ’ in which we exchange the positions of the tasks performed at levels i and j, i.e., the task that was performed at level i will be performed now at level j and vice versa. The optimal punishment at level i is now

$v_i' = c/(1-\alpha_j\alpha_{i+1},...,\alpha_{j-1}\alpha_i,..., \alpha_n)$, which is the same as v_i. However, the punishment for level j is $v_j' = c/(1-\alpha_i\ \alpha_{j+1},..., \alpha_n)$ which is greater than v_j since $\alpha_i > \alpha_j$. So the total punishment under the identity permutation is lower. We can now apply this argument successively, as we did in Proposition 2, to establish that the identity permutation is indeed the optimal one.

Proof of Proposition 4: Consider a mechanism v for which SPE leads to success with probability $\alpha^n < p < 1$. It must be the case that along the equilibrium path a group K of players choose to shirk and $k = |K| < n$. Let j be some player in N\K, and consider j's decision node in the game. If player j contributes the continuation path will lead to success with probability α^k, and j's expected payoff is $-v_j(1 -\alpha^k) - c$. If on the other hand j shirks the success probability is bounded below by α^n and his expected payoff is bounded below by $-v_j (1 -\alpha^n)$. Hence, in order to induce j to contribute we must have $-v_j(1 -\alpha^k) - c \geq -v_j (1 -\alpha^n)$, or $v_j \geq c/(\alpha^k-\alpha^n)$. But as α goes to zero this sends j's punishment to infinity. Thus for small enough α j's punishment will exceed the total punishment required to induce all players to contribute in an optimal INI mechanism, which completes the argument.

Proof of Proposition 5: Consider the following mechanism: Choose an arbitrary set of k players whose places in the order of moves are $i_1, i_2,..., i_k$ respectively and set $v_{ij} = c/(\alpha^{n-k}-\alpha^{n-j+1})$ and $v_r = 0$ for $r\in$ N\K (the set K refers to the agents at levels $i_1, i_2,..., i_k$). We will first argue that the equilibrium of such a mechanism results in the players in K contributing and the players in N\K shirking. First note that for players i_j in K, v_{ij} solves the following equation:

- $-c -v_{ij}(1-\alpha^{n-k}) = -v_{ij}(1-\alpha^{n-j+1})$ (3)

The LHS of (3) is the expected payoff to i_j if he contributes and all the remaining players in K contribute as well. The RHS is i_j's payoff if he shirks and all members of K appearing after him in the order shirk as well. Thus player i_j is indifferent between contributing and shirking under the conditions satisfied above. Furthermore,

- $-c -v_{ij}(1-\alpha^{n-k+1}) < -v_{ij}(1-\alpha^{n-j+2})$, (4) which imply that if one or more of i_j's predecessors in K shirks i_j prefers to shirk as well. Applying (3) and (4) with backward induction implies that the equilibrium path of the game yields all players in K contributing and all players in N\K shirking. Note that in the mechanisms (plural because the choice of the set K is arbitrary) described above the punishment for a player depends on his position within the hierarchy only through his relative position among the agents who are assigned to contribute. Furthermore, among all mechanisms that induce exactly k agents to contribute the ones described above are optimal. This uses the same argument for optimality as in the proof of Proposition 1.

Let g(k,α) denote the total punishment of inducing k players to contribute in an optimal mechanism described above for a given α. To complete the proof it is sufficient to show that if α is close enough to 1 g(k,α) < g(k+r,α) whenever 1≤ k ≤n-1, 2 ≤ k+r ≤ n and r ≥1. Indeed, setting m = n-(k+r), where m is the number of players that shirk, we have

$$g(k+r,a) = \underbrace{0+0,...,0}_{m} + \frac{c}{a^m - a^{m+1}} + \frac{c}{a^m - a^{m+2}} + ... + \frac{c}{a^m - a^n} \text{ and}$$

$$g(k,a) = \underbrace{0+0,...,0}_{m+r} \frac{c}{a^{m+r} - a^{m+r+1}} + \frac{c}{a^{m+r} - a^{m+r+2}} + ... + \frac{c}{a^{m+r} - a^n}.$$

Thus

$$(1-a)g(k+r,a) = \underbrace{0+0,...,0}_{m} + \frac{c}{a^m} + \frac{c}{a^m(1+a)} + ... + \frac{c}{a^m(1+a+a^2+,,,a^{n-m-1})} \text{ and}$$

$$(1-a)g(k,a) = \underbrace{0+0,...,0}_{m+r} \frac{c}{a^{m+r}} + \frac{c}{a^{m+r}(1+a)} + ... + \frac{c}{a^{m+r}(1+a+a^2+,,,a^{n-m-r-1})}.$$

Now when α converges to 1, (1-α)g(k+r, α) approaches c(1+1/2 +,...,1/(n-m)), while (1-α)g(k, α) approaches c(1+1/2+,...,1/(n-m-r)), which is smaller. This implies that for α close enough to 1 g(k, α) < g(k+r, α), which is what we need to complete the proof.

Proof of Proposition 6: By Proposition 5 for a given choice of k players $i_1, i_2, ..., i_k$ which are induced to contribute the optimal mechanism induces a cost of $\sum_{j=1}^{k} g_{ij}(v_{ij})$, where v_{ij} are as given in Proposition 5. Since the cost of punishment increases with the hierarchy level optimality implies that $i_1, i_2, ..., i_k$ equal 1, 2,3,..., k respectively.

Proof of Proposition 7: For $i \in N$ let $b_i(h) = |B_i(h)| + 1$, and $v_i^*(h) = c/(1 - \alpha^{b_i(h)})$. We show that $v^*(h) = \sum_{i \in N} v_i^*(h)$ is the minimal total sanction necessary to induce all players to invest in an equilibrium of G(h). This will complete the proof because by assumption for some i $B_i(h) \backslash B_i(h') \neq \emptyset$, yielding $v_i^*(h) < v_i^*(h')$ and $v_i^*(h) \leq v_i^*(h')$ for the rest of the players. Let $S_i(h) = \{j; i\ h\ j\}$ be the set of i's subordinates and consider the strategy combination σ in which each player invests if and only if all his subordinates invest. Under $v_i^*(h)$, it can be verified that σ is an equilibrium of G(h). Indeed, suppose by way of contradiction that player i can increase his payoff by deviating. There are two types of deviations from the specified strategy. First, i may shirk when all $S_i(h)$ invest. In this case, given the strategies of the other players, the set of players that will shirk is $B_i(h) \cup \{i\}$, and the probability of failure is $(1- \alpha^{b_i(h)})$, making i indifferent to investment. Second, player i may deviate by investing even when he observes some players in S(i) shirking. Indeed, if he does so he pays in expectation at least $v_i^*(1- \alpha) + c$. On the other hand he pays $(1-\alpha^{b_i(h)})v_i^*$ if he doesn't invest,[9] which is less. We now need to show that there exists no vector of sanctions v such that for some i in N $v_i < v_i^*$ and such that for some equilibrium of G(h) (under the mechanism v) all players invest. By way of contradiction let us suppose that such v exists and consider i with $v_i < v_i^*$. Let σ^* be the equilibrium leading to investment by all players under v. σ_i^* specifies i to invest if all $j \in S_i(h)$ invest. Suppose that player i deviates from σ_i^* by shirking even when all players in $S_i(h)$ invest. (Note that i's strategy is allowed to make i's action contingent only on the actions of players in $S_i(h)$.) Without specifying the strategies of the rest of the players, the worst case scenario for player i is that such deviation will lead all players in $B_i(h)$ to shirk as well (it may indeed happen that only a subset of $B_i(h)$ will shirk but no player outside $B_i(h)$ will shirk because the strategies of these players are not measurable with respect to i's action). This means that an upper bound on i's punishment following such a deviation is $v_i(1- \alpha^{b_i(h)})$, which by assumption is less than c. Hence, the deviation is profitable,

[9] Note that this is where the transitivity of the supervision order enters. By shirking, i induces his immediate bosses to shirk, which induce their immediate bosses to shirk, etc. The transitivity condition guarantees that all these agents are in the set B_i(h).

yielding the necessary contradiction.

Proof of Proposition 3*: Normalizing by setting c = 1 and P(N) = 1 player i's payment for a fixed order w is $\frac{1}{1-P(S_i^w)}$, where S_i^w is the set of players appearing before player i in the order w.

Claim: Let i and j be two adjacent players in the order w and assume that i appears before j. Denote by w' the order in which i and j switch positions. Let f(w) and f(w') be the total punishment imposed by the principal with respect to the fixed orders w and w'. If j's position is more important than i's , then f(w)≤ f(w').

Proof: Write f(w) = $\sum_{k\in S_i^w}\frac{1}{1-P(S_k^w)}+\frac{1}{1-P(S_i^w)}+\frac{1}{1-P(S_j^w)}+\sum_{k\in N\setminus[S_j^w\cup\{j\}]}\frac{1}{1-P(S_k^w)}$

and f(w') = $\sum_{k\in S_j^{w'}}\frac{1}{1-P(S_k^{w'})}+\frac{1}{1-P(S_j^{w'})}+\frac{1}{1-P(S_i^{w'})}+\sum_{k\in N\setminus[S_i^{w'}\cup\{i\}]}\frac{1}{1-P(S_k^{w'})}$

but $\sum_{k\in S_i^w}\frac{1}{1-P(S_k^w)}=\sum_{k\in S_j^{w'}}\frac{1}{1-P(S_k^{w'})}$, $\sum_{k\in N\setminus[S_j^w\cup\{j\}]}\frac{1}{1-P(S_k^w)} = \sum_{k\in N\setminus[S_i^{w'}\cup\{i\}]}\frac{1}{1-P(S_k^{w'})}$ and

$\frac{1}{1-P(S_i^w)}=\frac{1}{1-P(S_j^{w'})}$. Moreover, since j's position is assumed to be more important than i's we have

$P(S_j^w) = P(S_i^w\cup\{i\})\le P(S_i^w\cup\{j\}) = P(S_i^{w'})$.

We thus have f(w) - f(w') = $\frac{1}{1-P(S_j^w)}-\frac{1}{1-P(S_i^{w'})}\le 0$.

Suppose now that with respect to the order 1, 2, 3,…, n players' positions decrease in importance, i.e., player 1's position is the most important and n's is the least. Let w be an order different from the order w* = (n,n-1,n-2,…,1). There exists a finite sequence $w_1, w_2, \ldots, w_k$ with w_1 = w and w_k = w* such that for each $1 \le r \le k$ there exists a pair of players i, j such that: (1) i's position is more important than j's, (2) i appears before j in the order w_r, (3) the two players appear in the reverse order in w_{r+1}, and (4) all other players are ordered in the same way in w_r and w_{r+1}. By the above claim we have $f(w_r) \geq f(w_{r+1})$ for all r, which means that $f(w) \geq f(w^*)$ and hence w* is the optimal order.

Proof of Proposition 5*: As in Proposition 5 consider the mechanism in which k players are induced to invest. Specifically, for players in K = {1, 2, 3, ..., k} whose places in the order of moves are $i_1, i_2, ..., i_k$ respectively we set

$v_{ij} = c/(p(k) - p(j-1))$ and $v_r = 0$ for $r \in N \backslash K$. A similar argument as in Proposition 5 shows that in equilibrium of this mechanism all the players in K invest and the players in N\K shirk. This follows from the fact that v_{ij} solve the following equations:

- $-c - v_{ij}(1-p(k)) = -v_{ij}(1-p(j-1))$ (3*)

The LHS of (3*) is the expected payoff to i_j if he contributes and all the remaining players in K contribute as well. The RHS is i_j's payoff if he shirks and all members of K appearing after him in the order shirk as well. Thus player i_j is indifferent between contributing and shirking under the conditions satisfied above. Furthermore, due to p's increasing returns to scale

- $-c - v_{ij}(1-p(k-1)) < -v_{ij}(1- p(j-2))$, (4*)

which imply that if one or more of i_j's predecessors in K shirks i_j prefers to shirk as well. The rest of the argument follows as in Proposition 5.

We will now show that under our condition the total punishment as specified in v_{ij} above is increasing with the number of players that are induced to invest, i.e., with the size of the set K.

Let g(k) denote the total punishment required to induce k players to contribute.

$g(k) = \sum_{r=1}^{k} \frac{c}{p(k) - p(k-r)}$, while $g(k+1) = \sum_{r=1}^{k+1} \frac{c}{p(k+1) - p(k-r+1)}$.

Hence, g(k+1) - g(k) =

$$\sum_{r=1}^{k} \frac{c(p(k) - p(k-r)) - c(p(k+1) - p(k-r+1))}{(p(k) - p(k-r))(p(k+1) - p(k-r+1))} + \frac{c}{p(k+1) - p(0)} \text{ or}$$

$$g(k+1) - g(k) \geq \sum_{r=1}^{k} \frac{-c\Delta}{(p(k)-p(k-r))(p(k+1)-p(k-r+1))} + \frac{c}{p(k+1)-p(0)}.$$

Using the definitions of d_+ and d_- we have,

$g(k+1) - g(k) \geq -kc\Delta/d_-^2 + c/d_+$. And the RHS of the last inequality is positive due to our condition which implies that $g(k+1) > g(k)$.

References

Agihion, P. and J. Tirole (1997) "Formal and Real Authority in Organizations," *Journal of Political Economy* 105, 1-29.

Bag, P.K, and E. Winter (1999) "Simple Subscription Mechanisms for the Production of Public Goods," *Journal of Economic Theory* 87, 72-97.

Hart, O. and J. Moore (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy* 98, 1119-1157.

Hart, O and J. Moore (1999) "On the Design of Hierarchies: Coordination versus Specialization," Discussion paper 1880, Harvard Institute of Economic Research.

Holmstrom, B. (1982) "Moral Hazard in Teams," *Bell Journal of Economics* 13, 324-40.

James, D. and M. Soref (1981) "Profit Constraints on Managerial Autonomy: Managerial Theory and the Unmaking of the Corporate President," *American Sociological Review* 46, 1-18.

Keren, M. and D. Levhari (1983) "The Internal Organization of the Firm and the Shape of Average Costs," *Bell Journal of Economics* 41, 474-486.

Kremer, M. (1993) "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics* 108, 551-575.

McAfee, R.P. and J. McMillan, (1995) "Organizational Diseconomies of Scale," *Journalof Economics and Management Strategy* 4, 399-426.

Radner, R. (1992) "Hierarchy, The Economics of Managing" *Journal of Economic Literature* 30, 1382-1415.

Radner, R. (1993) "The Organization of Decentralized Information Processing," *Econometrica* 61,1109-1146.

Rosen, S. (1986) "Prizes and Incentives in Elimination Tournaments," *American Economic Review* 76, 701-715.

Sah, R.K. and Stiglitz (1986) "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review* 76, 716-25.

Segal, I. (2000) "Collusion, Exclusion and Inclusion in Random Order Bargaining,", Mimo.

Shleifer, A. and R. W. Vishny (1997) "A survey of Corporate Governance," *Journal of Finance* 52, 737-783.

Spence, A.M (1974) "Market Signaling" Cambridge, MA: Harvard University Press.

The Economist: "Firing the Boss" Oct 30, 1999.

Sobel, J. "How to Count to One Thousand" *Economic Journal* 102, 1-8.

Van Zandt, T. (1998) "The Scheduling and Organization of Periodic Associative Computation: Efficient Networks," *Economic Design* 3, 93-127.

Van Zandt, T. (1999) "Real Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents," *Review of Economic Studies* 66, 633-658.

Winter, E. (1994) "Non-Cooperative Bargaining in Natural Monopolies," *Journal of Economic Theory* 64, 202-220.

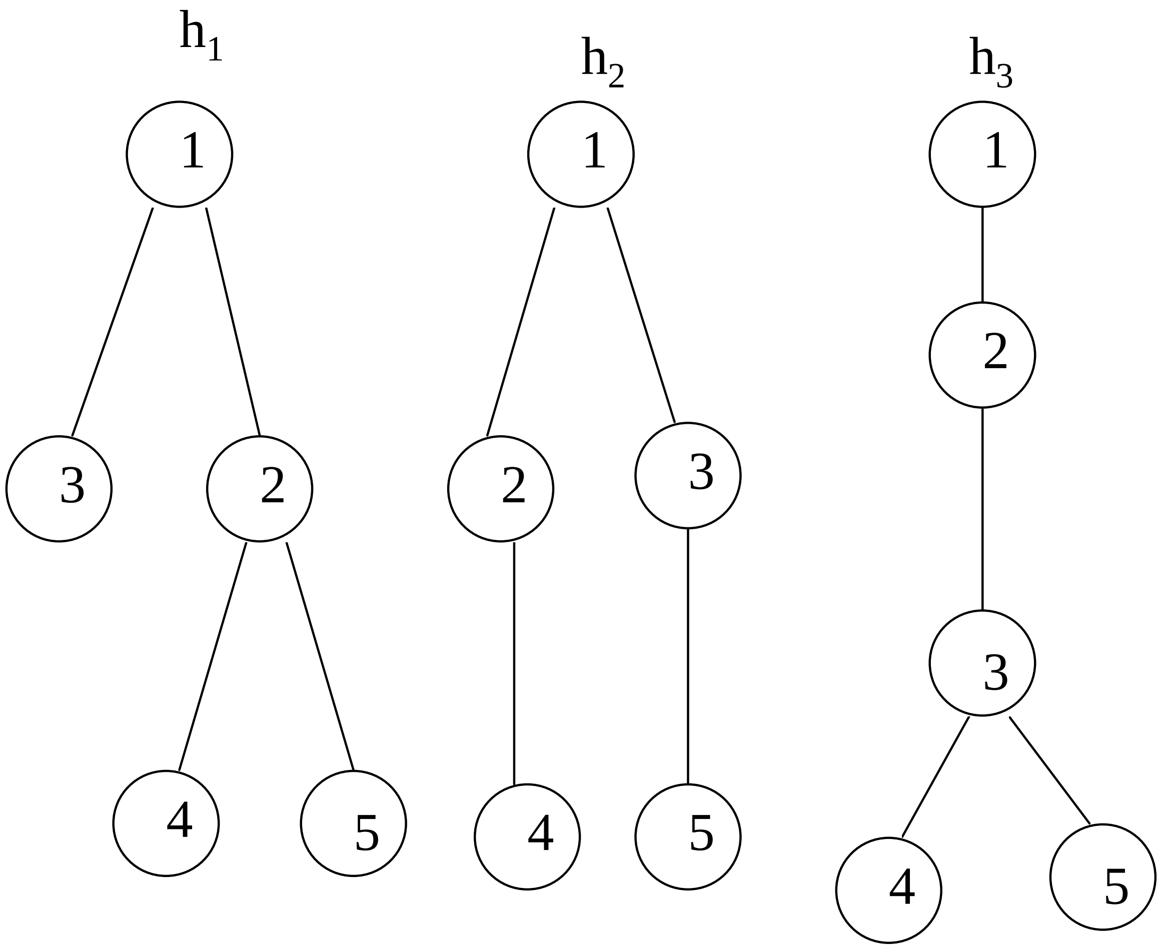

Figure 1