MECHANISM DESIGN WITH COSTLY COMMUNICATION: Implications for Decentralization[1]

Dilip Mookherjee and Masatoshi Tsumagari[2]

This Version: February 12, 2007

We develop a theory of mechanism design in a principal-multiagent setting with private information, where communication involves costly delay. The need to make production decisions within a time deadline prevents agents from communicating their entire private information to the principal, rendering revelation mechanisms infeasible. The mechanism design problem is formulated in a setting where production decisions are preceded by a multi-stage communication phase where agents and the principal exchange information. We examine trade-offs between centralization and decentralization of three components of the mechanism: contracting, communication and production decisions. Decentralization of contracting cannot dominate centralized contracting, but in some contexts can achieve equivalent profits for the principal. If cost hazard rates are linear, decentralization of production decisions and of communication strictly dominate centralization. These results apply even if communication is prone to exogenous errors or noise.

KEYWORDS: communication, mechanism design, decentralization, incentives, principal-agent, organizations

[1]We thank Ingela Alger, Richard Arnott, Ulf Lilienfeld, Bart Lipman, Michael Manove, Preston McAfee, Andy Newman, Roy Radner, Stefan Reichelstein and participants at Boston College and Brown University for useful comments.

[2]Boston University and Keio University, respectively

1 Introduction

It is well known that many basic questions concerning the organization of economic activity cannot be addressed by standard models of 'complete' contracts, despite incorporating problems of asymmetric information. These questions include separation of ownership from control, the relative value of centralized and decentralized decision-making, of different hierarchical forms with varying number of layers of decision-making and spans of control, the role of intermediaries in trade and finance, and so on. The key obstacle in the theory is the Revelation Principle, which states that the outcome of any organizational structure can be replicated by a degenerate two layer structure in which contracting, communication and all decisions are centralized: all control rights rest with the Principal that 'owns' the organization.

In its most general form (e.g., Myerson (1982)), the Revelation Principle is based on three sets of assumptions: (i) no restrictions on communication, contractual complexity, or computational ability; (ii) absence of collusion among agents, and (iii) ability of the Principal to commit to a mechanism. One of the most important reasons for pervasiveness of decentralized decision making are limits on communication and information processing capacities[3], which pertain to the first set of assumptions underlying the Revelation Principle.[4] Similar considerations have led to the 'incomplete contract' approach[5], where *ad hoc* restrictions are imposed on the contingencies that can be incorporated into contracts, or the extent of communication from agents to the Principal. Such restrictions are frequently justified by considerations of costs of complexity which typically remain unmodeled,[6] which has generated considerable controversy on the foundations of such an approach.[7] These considerations suggest the need for theories of optimal contracting subject to explicitly modeled costs of communication, information processing or com-

[3]See Hayek (1945), Marschak and Radner (1972), Mount and Reiter (1974, 1995), Radner (1992, 1993), and van Zandt (1996,1999)).

[4]Dropping (ii) concerning absence of collusion does not undermine the validity of the Principle, once collusion amongst agents is modeled explicitly in the form of sub-mechanisms (as in Laffont-Martimort (1997,2000) or Mookherjee-Tsumagari (2004)). Dropping assumption (iii) provides one approach to analyzing the costs and benefits of decentralization (Cremer (1995), Dessein (2002), Melumad-Mookherjee (1989), Poitevin (1995,2000)), an avenue we do not pursue here.

[5]See, e.g., Aghion and Tirole (1997).

[6]See Segal (1999) for a notable exception.

[7]See the March 1999 symposium in the *Review of Economic Studies*.

plexity which co-exist with strategic behavior and asymmetric information among agents.[8]

This paper develops a theory of limited capacity of agents and Principal to communicate, which prevents 'complete' contracts or revelation mechanisms from being implemented. In order to focus on communication constraints *per se*, we ignore issues of collusion among agents or limited commitment by the Principal. We follow Radner and van Zandt in postulating that production decisions are subject to time deadlines. However our approach differs from theirs in two essential respects: we incorporate incentive problems, and we focus on communicational constraints rather than limited computational abilities of agents. We assume that acts of writing and reading messages are time consuming, while other cognitive tasks (e.g., computation of optimal decisions) are not. In this respect the theory is similar to the Marschak-Radner (1972) theory of teams: agents are assumed to be rational, subject to 'technological' constraints on communication channels. Our theory may be viewed as an extension of Marschak-Radner team theory to incorporate incentive problems.

We postulate that a mechanism consists of three components or phases: a *contracting* phase, followed by a *communication* or *planning* phase, and finally a *decision-making* phase. Control rights over each of these components have to be allocated. Different organizational modes correspond to alternative combinations of control right assignments. We consider two different forms of contracting: centralized (where the Principal contracts with all agents personally), and decentralized (where the Principal delegates to some intermediate agents — 'managers' or 'prime contractors' – the right to contract with other agents — 'employees' or 'subcontractors'). Communication systems can likewise be centralized (where agents communicate only with the Principal, as in revelation mechanisms or tatonnement processes) or decentralized (where agents communicate directly with one another). Moreover, production decisions can be centralized (production targets decided by the Principal) or delegated to the agents (with appropriate performance-based incentive schemes).

Consequently, it is possible to mix and match components with differing

[8]Indeed, there are interesting theories of costly communication and of costly computation, such as Marschak and Radner (1972), Mount and Reiter (1974, 1995), Radner (1993), Bolton and Dewatripoint (1994), Marschak and Reichelstein (1995, 1998), van Zandt (1996,1999, 2003a,b), van Zandt and Radner (2001), but none of these simultaneously incorporate incentive problems.

degrees of decentralization. Centralized contracting is compatible with centralization or decentralization of either communication or production. One extreme is for all components to be centralized, as in a revelation mechanism. Alternatively, contracting and communication can be centralized while production decisions are decentralized (as in Lange-Lerner tatonnement mechanisms). Communication and production can both be decentralized, while contracting is centralized — as in firms where identification, discussion and solving of production problems are delegated to teams of workers for whom employment contracts are designed by a central headquarters. At the other extreme, all components of the mechanism can be decentralized: the Principal may contract with a single manager or prime contractor, and delegate to the latter the responsibility for contracting, communicating and decision making with respect to other employees or subcontractors.

Selection across these organizational alternatives is subject to a given technology of communication. The problem with developing a theory of mechanism design with costly communication concerns the specification of this technology. The few papers in the literature on this topic (incorporating incentive problems) make very restrictive and *ad hoc* assumptions concerning the set of feasible communication modes, e.g., consisting of one-shot communication mechanisms with restricted message space sizes, or a class of iterative mechanisms where agents communicate with a central agency. It is not clear the extent to which the results of these papers rely on the specific communication mechanisms considered. In this paper we adopt a more general class of dynamic communication mechanisms and obtain results that do not depend on the specific communication protocol. The class of communication systems we consider is fairly wide, including price or quantity-guided tatonnement processes, hierarchical networks, public meetings or 'notice-boards', and email systems.

We show that despite putting very little structure on the communication process (i.e., apart from the restriction that agents choose from a finite set of communication plans), some general results concerning optimal mechanism design can be derived. These results include the following:

(i) centralized contracting can replicate the outcomes of decentralized contracting;

(ii) if hazard rates of cost distributions are linear (which includes uniform and exponential distributions for cost shocks):

(a) incentive constraints pertaining to communication strategies do not bind, permitting the mechanism design problem to be formulated as a Marschak-Radner problem of selecting production assignments consistent with restrictions on the communication technology, to maximize expected 'virtual profits' of the Principal (where production costs are marked up by their corresponding inverse hazard rates to reflect informational rents of agents);

(b) decentralization of production decisions and of communication strictly dominate centralization;

(c) under additional assumptions concerning verifiability of subcontracting costs, decentralized contracting can implement an optimal allocation.

Result (i) implies that contrary to some claims in existing literature, costly communication does *not* provide a rationale for decentralized contracting. On the other hand, result (ii) shows that for a class of cost distributions, communication costs *do* provide a rationale for decentralization of communication and production decisions. Together, they imply that attention can be restricted to a class of mechanisms in which the Principal retains control over contracts, but decentralizes communication and production decisions to production agents — as in firms with worker teams described above. It is also possible for contracting to be decentralized without incurring any additional loss, if the Principal can additionally verify side payments and production assignments.[9] These results extend even if communication is prone to random errors, provided attention is restricted to a class of 'structured' communication mechanisms with a fixed number of rounds and assigned message spaces.

The paper is organized as follows. Section 2 describes relation to previous literature. Section 3 introduces the model, describing both centralized and decentralized contracting. Section 4 shows that decentralized contracting cannot achieve superior outcomes than centralized contracting. Section 5 thereafter provides a characterization of optimal allocations attainable via centralized contracts, both in general and in the presence of cost distributions with linear hazard rates. In the latter setting it establishes relative rankings of centralized and decentralized systems of production decisions and of com-

[9] Arguments in MMR (1995), elaborated further in Mookherjee (2006), show that these conditions are typically necessary as well for decentralization of contracting to entail no loss.

munication. Section 6 returns to consider decentralized contracting, and provides conditions for it to implement an optimal allocation. Section 7 then shows that the results extend to contexts where communication is subject to (exogenous) error or noise, in contrast to some claims in Williamson (1967). Finally, Section 8 concludes.

2 Related Literature

Previous theories of mechanism design with communication costs have exogenously constrained the size of message spaces (Green and Laffont (1987), Melumad, Mookherjee and Reichelstein (MMR, hereafter) (1992, 1997), and Laffont and Martimort (1998)).[10] We discuss the relation of this paper to MMR (1992), since the latter represented a more general version of the approach pursued in Green and Laffont (1987) and Laffont and Martimort (1998). MMR (1992) used a costly communication approach to develop a theory of the costs and benefits of delegation to managers of profit or cost centers within firms. Their formulation was based on the notion that centralized decision-making is necessarily less sensitive to private information possessed by agents, owing to their inability to communicate everything they know to the Principal. They additionally restricted attention to one-round communication between agents and the Principal with finite message spaces. On the other hand, no restriction was imposed on contracts designed by intermediaries for subordinates in a delegation mechanism. Decentralized contracting was shown to achieve strictly superior outcomes compared to centralized contracting. Central to this was the notion that communication constraints that restricted flexibility of centralized contracts did not similarly restrict subcontracts designed by intermediaries. Their approach is subject to the criticism that the offer of a subcontract is itself a form of communication between the intermediary and the subordinate, which should therefore be restricted in a manner similar to the restrictions imposed on centralized contracts. We show in this paper that once symmetric restrictions are imposed on contracts in both regimes, centralized contracts cannot be dominated by decentralized contracts.

This problem was addressed by Melumad, Mookherjee and Reichelstein in a subsequent paper (MMR 1997), where subcontracts designed by a manager

[10]Deneckere and Severinov (2003) develop a theory of mechanism design where truthful messages can be sent costlessly, while non-truthful messages may entail some cost.

or prime contractor were restricted in the same way that centralized contracts were. Symmetric restrictions were imposed on the number of contingencies (or message configurations) that production assignments for subordinates could be conditioned on. Yet other *ad hoc* features remained. Communication was restricted to a single round.[11] Even more important, perhaps, is that MMR (1992, 1997) restricted attention to two polar mechanisms: one in which contracting, communication and production decisions are all centralized, with another where all of these components are decentralized. Mixed modes were excluded, e.g. where contracting is centralized while production and communication is decentralized, or vice-versa. Many real-world firms exhibit such mixed modes: employees enter with contracts with the owner, but are organized into production teams or divisions that are awarded substantial autonomy over production and communication among one another. In this paper we impose no *ad hoc* restrictions on the range of possible communication systems or mechanisms, apart from the technological constraints that prevent agents instantaneously communicating to others all that they know.

3 Model

There is a Principal (P) and two agents 1 and 2. Agent $i = 1, 2$ produces a one-dimensional nonnegative real valued input q_i at cost $\theta_i q_i$, where θ_i is a real-valued parameter distributed over an interval $\Theta_i \equiv [\underline{\theta}_i, \bar{\theta}_i]$ according to a positive, continuously differentiable density function f_i and associated c.d.f. F_i. The distribution satisfies the standard monotone hazard condition that $\frac{F_i(\theta_i)}{f_i(\theta_i)}$ is nondecreasing, implying that the 'virtual cost' $v_i(\theta_i) \equiv \theta_i + \frac{F_i(\theta_i)}{f_i(\theta_i)}$ is strictly increasing. θ_1 and θ_2 are independently distributed, and these distributions F_1, F_2 are common knowledge among the three players.

The inputs of the two agents combine to produce a gross return $V(q_1, q_2)$ for P, whose net payoff is $V - t_1 - t_2$, where t_i denotes a transfer from P to i. The payoff of i is $t_i - \theta_i q_i$. All are risk-neutral and have autarkic payoffs of 0. We shall assume that V is twice continuously differentiable, strictly concave and satisfies Inada conditions. So the inputs of both agents are

[11] If agents are constrained on the amount of information they can communicate in any given round, there is a natural benefit from allowing multiple rounds of communication. In most real organizations, iterative rather than one-shot communication tends to be the norm.

essential in production. We shall also assume that the production function is non-separable: $V_{12} \neq 0$ for all q_1, q_2. This creates a need to coordinate production assignments between the agents.

3.1 Centralized Contracting

3.1.1 Timing

In centralized contracting, P selects a *communication protocol* (explained further below) at $t = -1$, and offers contracts to both agents. There is enough time between $t = -1$ and $t = 0$ for agents to read the offered contracts. P commits to the communication protocol and the contracts. The agents do not commit to their participation decisions until after they observe their private information.

At $t = 0$, each agent i privately observes the realization of θ_i, and independently decides whether to participate or opt out of the mechanism. If either agent opts out the game ends; otherwise they enter the planning or communication phase which lasts until $t = T$. The deadline T is exogenously given; alternatively it can be chosen by the organization designer to trade off the cost of delayed production with the benefit of added communication. Our results can be viewed as pertaining to choice among organizational modes for any given choice of T, with T chosen subsequently. Since the rankings we derive do not depend on the precise value of T, they apply also to a context where T is endogenously chosen.

At $t = T$, agent i selects production level q_i. This does not necessarily mean that i 'decides' q_i in any meaningful sense. As discussed further below, someone else may set a target for q_i — this could be a message sent to i by the target-setter during the communication phase — and the incentive scheme for i may effectively force i to meet this target. Finally payments are made according to the contracts signed.

3.1.2 Communication Protocols: Examples

Before providing a general treatment of communication protocols, it will help to provide examples of some specific protocols. For the time being we shall focus on a class of *structured communication protocols*, with a fixed number of rounds, a given sequence and form in which messages are sent and received. Section 5.3 will provide an underlying model of time taken to read and write messages which allows more flexible protocols.

In a structured protocol, there is a fixed number of rounds of communication, denoted by K. The kth stage corresponds to time interval $[t_{k-1}, t_k)$, with $t_0 = 0$ and $t_K = T$. Messages are sent at time points $t_k, k = 0, \ldots, K$; in the intervening time intervals (e.g., (t_{k-1}, t_k)) messages received at the preceding time point (t_{k-1}) are read, and new messages are written, to be sent at the next time point (t_k). At each stage k there is a given *communication network*, represented by $\mathcal{S}_{ik}$, the set of agents that i sends messages to at t_{k-1}. Agent i selects a message m_{ijk} to be sent to each $j \in \mathcal{S}_{ik}$ from a finite message set $\mathcal{M}_{ijk}$. Let $\mathcal{R}_{ik}$ denote the corresponding set of agents that i receives messages from at t_{k-1}.

The underlying 'technology' determining time taken to read and write messages is not specified any further; it is assumed that the stages of the communication protocol are designed so that agents have enough time to read and write messages during the intervals separating different time points when messages are exchanged. The fact that reading and writing are time-consuming activities impose restrictions on the size of the message spaces. We also assume that apart from the time taken to read and write messages, these activities do not impose any additional costs on the agents, so there is no moral hazard problem associated with these activities. In particular, we assume in this setting that every message received by an agent is read by that agent.

The history of messages sent ($\{m_{ijk'}\}_{j \in \mathcal{S}_{ik'}, k' \leq k}$) and received by i ($\{m_{jik'}\}_{j \in \mathcal{R}_{ik'}, k' \leq k}$) until stage k is denoted by h_{ik}. A dynamic *communication plan* c_i of i is represented by a function specifying messages to be sent by i ($\{m_{ij,k+1}\}_{j \in \mathcal{S}_{i,k+1}}$) at any stage $k+1$, given an arbitrary history h_{ik} observed by i until the preceding stage.[12] The message history observed by i at the end of the last stage h_{iK} is denoted h_i.

A *centralized communication network* is one where agents do not directly communicate with one another: all communication is between P and the agents. This includes tatonnement processes, where P suggests a set of 'prices' at each round to all agents, to which each agent subsequently responds with a quantity offer (where possible 'price' and 'quantity' messages are restricted to a finite set). It also includes a two-stage quasi-revelation

[12] While it is tempting to use the term 'communication strategy' for this, we reserve this term for the mapping signifying choice of a dynamic communication plan by any given type of an agent.

mechanism, where agents make a report of their cost to P at the first round (from a finite set of possible cost reports), to which P responds with a quantity target and payment (each from a finite set) for each agent.

A *hierarchical communication network* is represented by a tree, where every agent i has a set of subordinates $\mathcal{B}_i$, and a *boss* $b(i)$, the unique agent j such that i is a subordinate of j. Then each of $\mathcal{S}_{ik}$ and $\mathcal{R}_{ik}$ are either $\mathcal{B}_i$ or $\{b(i)\}$. In our context, it corresponds to a three-tier hierarchy where one agent is an informational intermediary between the other agent and P.[13]

A *public communication network* is one where at any stage one or more agents sends a message to all others in the organization. Examples of this are public notice-boards on which agents can post messages at each stage, and public meetings in which at any given stage one agent 'speaks' and all others 'listen'.

In all the above examples, the number of stages and the size of the message spaces are predetermined in a way that provides each agent enough time between any two stages to read messages received at the previous stage, update their information, write messages to be sent at the next stage. In general agents will take more time to read more complicated messages, so the time between successive stages (and hence the number of stages that can be fit into the fixed planning horizon) will depend on the size of the message spaces, as well as the number of agents that any given agent communicates with.

Without an underlying model of time needed to read and write, trade-offs between alternative communication systems (involving different networks, message space sizes and number of stages) cannot be formulated. Nor can we allow more flexibility in the sequencing of communication: e.g., where one agent may decide to ask another agent a question, and the length of the reply can depend on the precise question asked besides the private information possessed by the respondent. Section 5.3 will provide a more detailed model of reading and writing times which will allow such flexibility, besides allowing us to consider trade-offs between different communication systems.

[13]With more agents and vertical tiers, an example is a hierarchical budgeting system of the kind analyzed in Mookherjee and Reichelstein (1997, 2001), where agents at the lowest layer make cost reports to their boss, followed by an aggregate cost report by the latter to their boss in turn. Once cost reports have flowed up the tree, this is followed by a percolation of output targets and payments (i.e., budgets) down successive layers of the hierarchy.

3.1.3 Communication Protocols: A General Treatment

The key restriction in our theory is that the set of communication plans available to any player is finite, so it is physically impossible for agents to communicate what they know to others within a finite deadline T. We represent the communication technology in centralized contracting by a set of possible communication protocols $\mathcal{P}_C$. In any protocol $p \in \mathcal{P}_C$, player i has a finite set $\mathcal{C}_i$ of possible *communication plans*. Any element c_i of $\mathcal{C}_i$ is a dynamic plan for writing and reading messages based on the 'history' observed by i. A vector of communication plans chosen by the different players (c_1, c_2, c_P) jointly determine the history observed by i upto the deadline T, denoted by $h_i(c_1, c_2, c_P)$.

Formally, a *communication protocol* is represented by the set of possible communication plans for each agent $\mathcal{C}_i, i = 1, 2$, a communication plan for P chosen from a feasible set $\mathcal{C}_P$, and a function describing the message history observed by each agent at time T, given by $h_i : \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{C}_P \to \mathcal{M}_i$ for each $i = 1, 2, P$, where $\mathcal{M}_i$ is a set of possible sequences of messages observed by i between $t = 0$ and $t = T$.

3.1.4 Contracts and Production Decisions

The history h_i of messages sent and read by agent i is assumed to be *ex post* verifiable by P, so transfers from P to i can be conditioned on h_i. The assumption that the Principal can verify which messages sent or read by any given agent can be relaxed in some directions. It is more natural to suppose that messages received by one agent from another are verifiable. For instance, if every message is copied to the Principal, or messages transmitted leave a trail that the Principal can explore *ex post*. If communication channels are noisy, or agents make random errors in writing or speaking, messages that were 'intended' to be sent by an agent can deviate from the messages received by others. Alternatively, with noise in communication channels, only messages received by others are verifiable, while sent messages are not. We show in Section 7 that our principal results carry over in the presence of such communication errors, in the context of structured communication protocols.

Moreover, agents may not read some of the messages they have received. It may be difficult for the Principal to verify whether an agent read any given message that they received. We can replace this by the assumption that receipt of any message has to be 'acknowledged' by the receiver before they can read it, and the Principal can verify whether a given message sent to an agent was acknowledged by that agent. In such a context, it is easy to verify that agents would not benefit from acknowledging messages that they do not intend to read. Hence verification of messages acknowledged would be tantamount to verification of the messages that have been read. We shall henceforth use the term 'read' or received' as equivalent shorthands for 'messages whose receipt has been acknowledged'.

More restrictive is the assumption that there is no scope for any 'private' communication between agents, that the Principal cannot observe. The Principal is also assumed to be able to verify production levels q_1, q_2. We do not allow for any form of collusion between the agents: extension of the model to incorporate collusion will have to be left for future research.

Given these assumptions, a centralized contract for i is represented by a transfer rule $t_i(q_i, q_j, h_i)$.

Note that restrictions on communication force production decisions to depend on 'coarse' messages about the state of the other agent: the production of agent i must depend on θ_j only through h_i. Specifically, agent i's production strategy is a function $\hat{q}_i(\theta_i, h_i(c_i(\theta_i), c_j(\theta_j), c_P))$. They can be fine-tuned to information about i's own cost θ_i, which constitutes the potential 'flexibility' advantage of decentralizing production decisions.

Formally, we shall say that the *production decision q_i is centralized* if it is measurable with respect to h_P, and *partially decentralized* if it is measurable with respect to $h_j, j \neq i$ but not with respect to h_P. In these cases, production decisions concerning q_i can be thought of as being made by P or by j respectively. In contrast, the production decision q_i is said to be *completely decentralized* if it is not the case that q_i is measurable with respect to h_P or $h_j, j \neq i$. The production decision of i cannot then be predicted by P or j at $t = T$ based on the messages they have sent or read. In other words it is not possible that agent i is assigned a production target by someone else, combined with an incentive scheme that forces i to abide by the target. Instead, i will make the production decision personally, a choice that will be influenced, though not completely determined, by the messages sent by others via the incentive scheme.

3.2 Decentralized Contracting

3.2.1 Timing

In decentralized contracting, P contracts with the manager (agent i), who subsequently contracts with the employee (agent j). The communication network is also hierarchical: the employee communicates with the manager, and the manager with the Principal. Contracts are offered at $t = -1$: P offers a contract for i and selects an 'upper-level' communication protocol between herself and the manager. The manager then offers a subcontract to j, and selects the 'lower level' communication protocol. The two networks are linked by the participation of the manager: messages sent by the manager on one network may be based on messages received in previous stages on the other network. For instance, the manager may receive a cost report from the employee, combine this with her own cost information to produce a summary cost report to P. Following this P may set an output target, with the manager subsequently allocating production responsibility between herself and the employee.

The rest is as under centralized contracting. At $t = 0$ agents i, j observe the realization of θ_i, θ_j respectively, and each independently decides whether or not to opt out. If neither opts out, they enter the communication phase from $t = 0$ until $t = T$. At T agents i, j decide q_i, q_j respectively, and payments are thereafter made as stipulated in the contracts.

3.2.2 Communication Protocol

The set of communication protocols consistent with decentralized contracting do not allow any direct communication between the employee j and the owner P. Each of them communicates only with the manager i. Moreover, P does not monitor communication between the manager and the employee, neither does the employee monitor communication between the manager and P.[14]

However, the manager may employ an information coordinator in his communication with the employee. Under centralized contracting, P can facilitate communication between the agents to allow for better coordination. If reading rather than writing messages is time consuming, agents may send certain messages to one another, and other messages to P, which are processed by P and 'summaries' are subsequently communicated back to the

[14]We use the term 'monitor' as a shorthand for 'observe' and/or 'ex post verify'.

agents. Such a coordination role can be played by some (nonstrategic) fourth party, or mechanical devices such as a computer program. In a decentralized setting, the lower level communication network involving the manager and the employee can also employ such a coordination device. Indeed, this is essential to maintain parity in the communicational constraints between centralized and decentralized mechanisms. Accordingly we shall assume that the manager and the employee can employ a coordination device called M with the same communicational and information processing capacity as P. In other words, the same message space can be assigned to M as was assigned to P under centralization.[15]

In similar vein, the 'upper level' network involving P and the manager may also engage an external coordinator or coordination device. Let the coordinators in the upper-level and lower-level networks be denoted by N and M respectively. Let the set of communication plans in the upper-level network of P, i, N be denoted by $\mathcal{C}_P^*, \mathcal{C}_i^U, \mathcal{C}_N$ respectively. And let the corresponding set for the lower-level network of i, j, M be denoted by $\mathcal{C}_i^L, \mathcal{C}_j, \mathcal{C}_M$.

Agent i is an intermediary, the sole link between the two networks. So the 'larger' communication plan of i includes plans for both communication on both networks: $c_i \equiv (c_i^U, c_i^L)$, as a function of messages previously sent and received on both networks. And the communication plans of members of either upper or lower network are conditioned on prior communication with i, so they depend indirectly on messages sent by participants in the other network. Hence histories observed by each member at $t = T$ depends on communication plans of all others: $h_l(c_P^*, c_i^U, c_N; c_i^L, c_j, c_M), l = P, i, N, j, M$.

Since P can commit to her strategy in advance, and N is a non-strategic player, they can be merged into a single non-strategic communication device, represented by a communication plan $c_P \equiv (c_P^*, c_N) \in \mathcal{C}_P \equiv \mathcal{C}_P^* \times \mathcal{C}_N$. To simplify notation, we therefore drop mention of N from now onwards.

Let $\mathcal{P}_D^l$ denote the set of communication protocols at level $l = 1, 2$ of the hierarchy. The Principal selects a protocol $p_1 \in \mathcal{P}_D^1$ at the upper layer $l = 1$. This determines the communicational responsibilities of the manager *vis-a-vis* the Principal, and constrains the protocols that the manager can choose from for the bottom layer $l = 2$. For instance, it should allow the manager enough time to be able to execute his communicational responsibilities on both networks. If P wants the manager to report on some cost estimate for

[15] Alternatively, P has the ability to appoint a coordination agent or device analogous to M in centralized contracting.

delivering V, for which prior communication with j is necessary, the protocol in the lower level network should ensure that the manager communicates with j before the time to report to P arrives. Given p_1, the subset of protocols that i can choose for the lower level network is a subset of $\mathcal{P}_D^2$, represented by a correspondence $\mathcal{P}^2(p_1) : \mathcal{P}_D^1 \rightrightarrows \mathcal{P}_D^2$. Hence $\mathcal{P}_D$, the set of communication protocols feasible in the decentralized contracting regime, is represented by $\mathcal{P}_D^1$ the set of protocols for the upper tier, along with the correspondence $\mathcal{P}^2(.)$.

The choice of communication protocols at the two tiers jointly determine an extensive form communication-cum-production subgame (i.e., sequence of moves and corresponding information sets for each player during the communication phase, followed by production choices at T). Despite the fact that P does not communicate directly with the employee, messages sent by the employee indirectly affect messages received by P since the former can affect messages sent by the manager to P.

Note that $\mathcal{P}_D \subset \mathcal{P}_C$, i.e., any protocol feasible under decentralized contracting is also feasible under centralized contracting, while the converse is not true. In centralized contracting, P has the option of selecting the same hierarchical communication protocol as in decentralized contracting. However, decentralized contracting must necessarily involve a hierarchical protocol, whereas centralized contracting is not so constrained.

Since $\mathcal{P}_D \subset \mathcal{P}_C$, we do not need any fresh notation for communication protocols in decentralized contracting, apart from noting that they form a strict subset of communication protocols in centralized contracting.[16]

3.2.3 Contracts and Production Decisions

Just as in centralized contracting, it is assumed that in decentralized contracting P can monitor the inputs supplied by either agent, as well as transfer payments made by the manager to her subordinates. In order to allow the manager to make certain kinds of commitments, we introduce a third party agent H to whom the manager may make (or receive) transfers.[17] This third

[16]This involves some abuse of notation, which turns out to be inessential. Strictly speaking, a communication network in decentralized contracting involves four participants $(P, 1, 2, M)$ while that in centralized contracting involves three participants $(P, 1, 2)$. So the same notation applies only when we merge the roles of the two 'non-strategic' participants P and M into a single non-strategic participant.

[17]This is elaborated further in Section 4.

agent plays no active role in the mechanism, and has to be assured of a non-negative expected transfer in order to participate. Let t_j, t_H denote transfers made by the manager i to j and H respectively, which are verifiable by P. The contract between P and the manager is therefore represented by the transfer rule $t_i = t_i(q_1, q_2, h_{Pi})$, where h_{Pi} denotes (h_P, h_i^U) the history of messages exchanged between P and i on the upper level network. Here h_P denotes the messages sent and received by P, and h_i^U denotes messages sent (or received) by i to (or from) P.

The subcontracts offered by the manager to the employee or third party agent specifies transfers t_j, t_H as a function of q_1, q_2 and messages exchanged on the lower level network $h_{12} \equiv (h_i^L, h_j^L, h_M)$. Here h_i^L and h_j^L denotes messages sent or received by i and j respectively among one another.

Production decisions q_i, q_j are made at $t = T$ by i, j respectively, based on their personal information at that point. The manager decides $\hat{q}_i(\theta_i, h_i)$ where $h_i \equiv (h_i^U, h_i^L)$, and the employee decides $\hat{q}_j(\theta_j, h_j)$. Production decisions may be centralized or decentralized, as in the centralized contracting regime. The same formal definitions of (completely, partially) decentralized and centralized production decisions apply here as in the centralized contracting regime.

In our formulation, subcontracts and communication protocol for the lower level network are designed by the manager at the *ex ante* stage. If they were designed instead at the *interim* stage, employees would need time to read the subcontract offered, which would cut into the time available for coordinating production plans. In that case, the set of possible subcontracts offered and accepted at the interim stage will belong to a finite set. Our formulation is equivalent to this: one can think of the subcontract offered *ex ante* as a finite menu of subcontracts offered at the interim stage, with subsequent communication between the agents between $t = 0$ and T serving to select one from the menu.

4 Allocations Attainable under Centralized and Decentralized Contracting

Recall that a given communication protocol specifies sets of feasible communication plans $\mathcal{C}_i, \mathcal{C}_j$ for the agents, as well as communication plans c_P or c_M for P and M respectively. Given the protocol, a *communication strategy* for agent $a = 1, 2$ is a mapping $c_a(\theta_a) : \Theta_a \to \mathcal{C}_a$. It specifies for any given type of the agent a dynamic communication plan, specifying messages to be sent at each round of the communication phase, as a function of θ_a and messages received from others at previous rounds. We shall use $\tilde{c}(.)$ to denote a vector of communication strategies for the different players. In equilibrium players will correctly anticipate the communication strategy used by others and use this to update their information after receiving messages from them.

A *production strategy* for agent a is a mapping $\hat{q}_a(\theta_a, h_a) : \Theta_a \times \mathcal{H}_a \to \Re_+$, representing the production decision made by type θ_a at T after observing message history h_a.

Given a particular set of contracts offered by P in centralized contracting and a given communication protocol $p \in \mathcal{P}_C$ (including communication strategy c_P^*), we have a well-defined Bayesian game of incomplete information.

An *allocation attainable under centralized contracting* is a state-contingent production and transfer rule $q_a(\theta_1, \theta_2), t_a(\theta_1, \theta_2), a = 1, 2$ such that there exists a communication protocol $p \in \mathcal{P}_C$, centralized contracts $t_a(q_1, q_2, h_a), a = 1, 2$, and an associated tuple of communication and production strategies $\tilde{c}(.) \equiv \{c_i(\theta_i), c_j(\theta_j), c_P\}$ and $\tilde{q}(.) \equiv \{\hat{q}_i(\theta_i, h_i), \hat{q}_j(\theta_j, h_j)\}$ which constitutes a Perfect Bayesian Equilibrium (PBE) of the corresponding subgame, such that for any state $\theta \equiv (\theta_1, \theta_2)$ and any $a = 1, 2$:

$$q_a(\theta) = \hat{q}_a(\theta_l, h_a(\tilde{c}(\theta))) \quad (1)$$

$$t_a(\theta) = t_a(\tilde{q}(\theta), h_a(\tilde{c}(\theta))) \quad (2)$$

Under decentralized contracting, a different Bayesian game is induced by choice of a contract for the manager and 'upper layer' communication protocol p_1 by P. Agent i, the manager, must select contracts t_j, t_H for agent j and the third party, a communication protocol $p_2 \in \mathcal{P}^2(p_1)$ (which includes a communication plan c_M for the coordinator M), and a communication strategy $c_i(\theta_i)$ to be executed by i during the communication phase. Production decisions and the strategies of agent j are similar to what they are under centralized contracting.

An *allocation attainable under decentralized contracting* is a state-contingent production and transfer rule $q_a(\theta), t_a(\theta), a = 1, 2$ such that there exists a contract $\hat{t}_i(q_1, q_2, t_j, t_H, h_{Pi})$ and communication protocol $p_1 \in \mathcal{P}_D^1$ (with communication plan $\hat{c}_P$) selected by P for the top tier, and a Perfect Bayesian Equilibrium (PBE) of the associated subcontracting subgame consisting of a subcontracts offered by i: $\hat{t}_j(q_1, q_2, h_{12}), \hat{t}_H(q_1, q_2, h_{12})$, a communication protocol

$p_2 \in \mathcal{P}^2(p_1)$, a tuple of communication strategies $\hat{c}(.) \equiv \{\hat{c}_i(\theta_i), \hat{c}_j(\theta_j), \hat{c}_M, \hat{c}_P\}$ and production strategies $\hat{q}(.) \equiv \{\hat{q}_i(\theta_i, h_i), \hat{q}_j(\theta_j, h_j)\}$, such that for any state $\theta \equiv (\theta_i, \theta_j)$:

$$
\begin{aligned}
q_a(\theta) &= \hat{q}_a(\theta_a, h_a(\hat{c}(\theta))), a = 1, 2 \\
t_i(\theta) &= \hat{t}_i(q_1(\theta), q_2(\theta), t_j(\theta), t_H(\theta), h_{Pi}(\hat{c}(\theta))) - t_j(\theta) - t_H(\theta) \\
t_a(\theta) &= \hat{t}_a(q_1(\theta), q_2(\theta), h_{12}(\hat{c}(\theta))), a = j, H
\end{aligned}
$$

Proposition 1 *Any allocation attainable under decentralized contracting can also be attained under centralized contracting.*

Proof. Consider an allocation $q_a(\theta), t_a(\theta), a = i, j$ attained by decentralized contracting with protocols p_1, p_2 at the two layers, transfer rules $\hat{t}_i, \hat{t}_j, \hat{t}_H$, communication and production strategies $\hat{c}, \hat{q}_1, \hat{q}_2$. Recall that the communication protocol $p \equiv (p_1, p_2)$ is feasible in centralized contracting. Recall also the assumption that P can verify all messages sent and received by all agents in centralized contracting. Therefore h_{12} is verifiable by P in centralized contracting. So the Principal can select the protocol p, and the following contracts:

$$
\begin{aligned}
t_j(q_1, q_2, h_{12}) &= \hat{t}_j(q_1, q_2, h_{12}) \\
t_i(q_1, q_2, h_{12}, h_{Pi}) &= \hat{t}_i(q_1, q_2, \hat{t}_j(q_1, q_2, h_{12}), \hat{t}_H(q_1, q_2, h_{12}), h_{Pi}) \\
&\quad -\hat{t}_j(q_1, q_2, h_{12}) - \hat{t}_H(q_1, q_2, h_{12})
\end{aligned}
$$

Then from $t = 0$ the continuation game involving choice of communication and production strategies by the agents is the same as under decentralized contracting, so the same strategies constitute a PBE of this game. ∎

The argument of Proposition 1 resembles that underlying the Revelation Principle: under identical communication and contracting constraints, centralized contracts can be designed by the Principal to duplicate any mechanism based on decentralized contracts. In this setting, the assumptions underlying the formulation of MMR (1992) cannot be justified. The key assumption in MMR is that the manager can offer a subcontract which is conditioned on the realization of θ_i, which cannot be duplicated with a centralized contract because that would require the manager to communicate θ_i

to P. The time constraints that prevent the manager from communicating θ_i in its entirety to P also prevent the manager from communicating a subcontract that is conditioned on θ_i to the employee. Since the same constraints on communication operate in both systems, whatever the manager communicates to the employee can be used by P to condition centralized contract offers to j. There is no scope for decentralized contracts to be better tuned to the manager's private information than centralized contracts.

5 Characterizing Optimal Allocations in Centralized Contracting

Proposition 1 pertains to allocation of control over contracting, but says nothing about control over production, or the design of communication. Having designed contracts, P needs to decide whether to retain control rights over production as well, or how to design a communication system. In this section we explore these issues. In order to do so we need to make some progress in characterizing optimal allocations. The following result shows how the standard characterization of feasible mechanisms with costless communication can be extended to this setting.

Proposition 2 *A production allocation $q \equiv \{q_i(\theta_i, \theta_j)\}_{i=1,2}$ is attainable under centralized contracting if and only there exists a communication protocol $p \in \mathcal{P}_C$ and associated functions $\hat{q}_i(\theta_i, h_i) : \Theta_i \times \mathcal{H}_i \to \Re_+$, $c_i(\theta_i) : \Theta_i \to \mathcal{C}_i$ and $U_i(h_i) : \mathcal{H}_i \to \Re_+$ such that for $i = 1, 2$:*

(i) $q_i(\theta_i, \theta_j) = \hat{q}_i(\theta_i, h_i(\tilde{c}(\theta)))$ *for all* θ

(ii) $\hat{q}_i(\theta_i, h_i)$ *is non-increasing in* θ_i *for any* h_i

(iii) $c_i(\theta_i)$ *solves*

$$\max_{c_i \in \mathcal{C}_i} E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i(c_i, c_j(\theta_j), c_P))dx + U_i(h_i(c_i, c_j(\theta_j), c_P))]$$

(iv) $E_{\theta_j}[U_i(h_i(c_i(\bar{\theta}_i), c_j(\theta_j), c_P))] \geq 0.$

Such an allocation generates the following expected profit for P:

$$E[V(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))) - v_i(\theta_i)\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))) - v_j(\theta_j)\hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))) \\ -U_i(h_i(c_i(\bar{\theta}_i), c_j(\theta_j), c_P)) - U_j(h_j(c_j(\bar{\theta}_j), c_i(\theta_i), c_P))] \quad (3)$$

Proof.

Necessity:

Suppose that $(t_i(\theta_i,\theta_j), q_i(\theta_i,\theta_j), \tilde{c}(\theta))$ is the outcome of a PBE of some centralized contract mechanism. Consider any history h_i observed by type θ_i of agent $i(=1,2)$ at T, and let $\hat{q}_i(\theta_i, h_i)$ denote the production choice of i at that information set. Then (i) follows from the definition of the equilibrium outcome of this PBE.

Define $\hat{t}_i(q_i, h_i)$ to be the transfer expected by i following choice of output q_i, with expectation taken with respect to the posterior beliefs of i at that information set. Specifically, it is the expected value of $t_i(q_i, \hat{q}_j(\theta_j, h_j), h_i)$ where expectation is taken with respect to posterior beliefs of i over (θ_j, h_j) after having observed h_i. (It can be verified that conditional on h_i these beliefs do not depend on θ_i or the communication strategy chosen by i: this is shown explicitly in the more complicated environment of noisy communication in Section 7.)

Sequential rationality of PBE strategies implies that $\hat{q}_i(\theta_i, h_i)$ maximizes $\hat{t}_i(q_i, h_i) - \theta_i q_i$ for every θ_i and h_i. This implies (ii); moreover upon defining

$$U_i(h_i) = \hat{t}_i(\hat{q}_i(\bar{\theta}_i, h_i), h_i) - \bar{\theta}_i \hat{q}_i(\bar{\theta}_i, h_i)$$

the envelope theorem implies that

$$\hat{t}_i(\hat{q}_i(\theta_i, h_i), h_i) - \theta_i \hat{q}_i(\theta_i, h_i) = \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i)dx + U_i(h_i). \tag{4}$$

Sequential rationality with respect to choice of communication plans then implies condition (iii) and with respect to participation decisions implies condition (iv). This implies that θ_i solves

$$\max_{\theta_i'} E_{\theta_j}[\hat{t}_i(\hat{q}_i(\theta_i', h_i(\tilde{c}(\theta_i', \theta_j))), h_i(\tilde{c}(\theta_i', \theta_j))) - \theta_i \hat{q}_i(\theta_i', h_i(\tilde{c}(\theta_i', \theta_j)))].$$

Hence the expected transfer is represented by

$$E_{\theta_j}[\theta_i \hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))) + \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i(\tilde{c}(x, \theta_j)))dx + U_i(h_i(\tilde{c}(\bar{\theta}_i, \theta_j)))].$$

Then the expected profit of P reduces to (3) via standard arguments.

Sufficiency:

Construct centralized contracts as follows: we claim there is a PBE of the corresponding game which results in the given allocation. Given the function $\hat{q}_i(\theta_i, h_i)$, define $\hat{\theta}_i(q_i, h_i)$ as follows:

$$\hat{\theta}_i(q_i, h_i) \equiv \sup_{\theta_i}\{\theta_i \mid \hat{q}_i(\theta_i, h_i) \geq q_i\}.$$

Also define the following transfer functions stipulated in the centralized contract:

$$t_i(q_i, h_i) = \hat{\theta}_i(q_i, h_i)q_i + \int_{\hat{\theta}_i(q_i,h_i)}^{\bar{\theta}_i} \hat{q}_i(x, h_i)dx + U_i(h_i).$$

By construction, then, the following is true for any θ_i, h_i:

$$t_i(\hat{q}_i(\theta_i, h_i), h_i) - \theta_i\hat{q}_i(\theta_i, h_i) = \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i)dx + U_i(h_i).$$

Combined with (ii) this implies that $t_i(q_i, h_i) - \theta_i q_i$ is maximized at $\hat{q}_i(\theta_i, h_i)$. Then taking the strategy of the other agent $\hat{q}_j(\theta_j, h_j)$ and $c_j(\theta_j)$ as given, agent i's interim payoff as a function of a given communication plan c_i is

$$E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i(c_i, c_j(\theta_j), c_P))dx + U_i(h_i(c_i, c_j(\theta_j), c_P))].$$

By (iii), this is maximized at $c_i(\theta_i)$. With passive beliefs for off-equilibrium path histories (i.e., with agents not revising their prior beliefs when they observe off-equilibrium path messages), pursuing the continuation of $c_i(\theta_i)$ from any date t between 0 and T is optimal. For if this is not the case and there exists a deviation from some date t history onwards, the plan $c_i(\theta_i)$ could not have been sequentially rational at $t = 0$. A better plan at $t = 0$ would have been to select the deviating continuation strategy from date t history onwards, combined with the continuation strategy stipulated by $c_i(\theta_i)$ from all other date-history pairs. Finally, by (iv) it is sequentially rational for i to participate. Hence there exists a PBE which results in the given allocation. ∎

Proposition 2 represents the constraints imposed by limited communication and incentive compatibility. Constraint (i) shows how the former limits coordination of production decisions: a given agent's production can be based only on h_i, a coarse signal of the cost state of the other agent. This limits the flexibility of production assignments. Constraint (iii) represents the incentive constraint with respect to choice of communication strategies by agents.

The mechanism design problem can now be posed as follows: select communication protocol $p \in \mathcal{P}_C$, and associated production allocations $(\hat{q}_i(\theta_i, h_i), \hat{q}_j(\theta_j, h_j))$, communication strategies $\tilde{c}(\theta) \equiv (c_i(\theta_i), c_j(\theta_j), c_P)$ to

$$\max E[V(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta_i, \theta_j))), \hat{q}_j(\theta_i, h_j(\tilde{c}(\theta_i, \theta_j)))) - v_i(\theta_i)\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta_i, \theta_j))) - v_j(\theta_j)\hat{q}_j(\theta_j, h_j(\tilde{c}(\theta_i, \theta_j)))] \quad (5)$$

subject to

(i) $\hat{q}_i(\theta_i, h_i)$ is non-increasing in θ_i

(ii) $c_i(\theta_i)$ solves

$$\max_{c_i \in \mathcal{C}_i} E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_i(c_i, c_j(\theta_j), c_P))dx + U_i(h_i(c_i, c_j(\theta_j), c_P))]$$

for a set of functions $U_i(h_i), i = 1, 2$ satisfying the restriction

$$E_{\theta_j}[U_i(h_i(c_i(\bar{\theta}_i), c_j(\theta_j), c_P))] = 0.$$

Constraint (ii) pertaining to incentive compatibility of the communication strategies is complicated; further insight into the structure of the problem is possible when this does not bind. We now provide a distributional condition which ensures this.

5.1 Exponentially Distributed Costs

Consider the case where θ_i has an exponential distribution: $F_i(\theta_i) = [\frac{\theta_i - \underline{\theta}_i}{\bar{\theta}_i - \underline{\theta}_i}]^{1/\lambda_i}$ for some $\lambda_i > 0$. This includes uniform distributions as a special case.

Proposition 3 *Suppose costs of both agents are exponentially distributed. Then both the monotonicity constraint (i) and the incentive constraint (ii) pertaining to communication strategies in the mechanism design problem are not binding, i.e., an optimal production allocation can be represented as the solution to the unconstrained (team-theoretic) problem of choosing a communication protocol* $p \in \mathcal{P}_c$*, production and communication strategies* $\{q_i(\theta_i, h_i), c_i(\theta_i)\}_{i=1,2}, c_P$ *to maximize expected profits (5) of the Principal.*

Proof. Consider any communication protocol $p \in \mathcal{P}_c$. In the unconstrained problem, the optimal allocation $(q_i^R(\theta_i,\theta_j), c_i^R(\theta_i))$ satisfies the the following properties:

(i) $q_i^R(\theta_i,\theta_j) = \hat{q}_i^R(\theta_i, h_i(\tilde{c}^R(\theta)))$.

(ii) $\hat{q}_i^R(\theta_i, h_i(c_i, c_j^R(\theta_j), c_P^R))$ maximizes

$$E_{\theta_j}[V(q_i, \hat{q}_j^R(\theta_j, h_j(c_i, c_j^R(\theta_j), c_P^R))) \mid h_i(c_i, c_j^R(\theta_j), c_P^R) = h_i] - v_i(\theta_i)q_i$$

(iii) $c_i^R(\theta_i)$ solves $\max_{c_i \in \mathcal{C}_i} E_{\theta_j}[\pi(\theta_i, h_i(c_i, c_j^R(\theta_j), c_P^R))]$ where

$$\pi(\theta_i, h_i) \equiv E_{\theta_j}[V(\hat{q}_i^R(\theta_i,h_i), \hat{q}_j^R(\theta_j, h_j(c_i, c_j^R(\theta_j), c_P^R)))$$
$$-v_i(\theta_i)\hat{q}_i^R(\theta_i,h_i) - v_j(\theta_j)\hat{q}_j^R(\theta_j, h_j(c_i, c_j^R(\theta_j), c_P^R)) \mid h_i(c_i, c_j^R(\theta_j), c_P^R) = h_i].$$

(ii) implies that $\hat{q}_i^R(\theta_i, h_i)$ is non-increasing in θ_i, since $v_i(\theta_i)$ is increasing in θ_i. Note that π depends on c_i only through its effect on h_i, since h_j does not depend on c_i taking h_i, c_j and c_P as given. By the Envelope theorem, (iii) is equivalent to

$$c_i^R(\theta_i) \in \arg\max_{c_i \in \mathcal{C}_i} E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i^R(x, h_i(c_i, c_j^R(\theta_j), c_P^R))v_i'(x)dx + \pi(\bar{\theta}_i, h_i(c_i, c_j^R(\theta_j), c_P^R))] \tag{6}$$

For it to be attainable,

$$c_i^R(\theta_i) \in \arg\max_{c_i \in \mathcal{C}_i} E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i^R(x, h_i(c_i, c_j^R(\theta_j), c_P^R))dx + U_i(h_i(c_i, c_j^R(\theta_j), c_P^R))] \tag{7}$$

and

$$E_{\theta_j}[U_i(h_i(c_i^R(\bar{\theta}_i), c_j^R(\theta_j), c_P^R)] = 0,$$

must be satisfied for some set of functions $U_i(h_i)$.

When θ_i is distributed exponentially, $v_i'(\theta_i) = 1 + \lambda_i$ for all θ_i. Define $U_i(h_i)$ as

$$U_i(h_i) \equiv [\pi(\bar{\theta}_i, h_i) - E_{\theta_j}[\pi(\bar{\theta}_i, h_i(c_i^R(\bar{\theta}_i), c_j^R(\theta_j), c_P^R))]]/(1+\lambda_i).$$

It is evident that

$$E_{\theta_j}[U_i(h_i(c_i^R(\bar{\theta}_i), c_j^R(\theta_j), c_P^R)] = 0$$

and

$$
\begin{aligned}
& E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i^R(x, h_i(c_i, c_j^R(\theta_j), c_P^R))dx + U_i(h_i(c_i, c_j^R(\theta_j), c_P^R))] \\
= \; & 1/(1+\lambda_i)[E_{\theta_j}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i^R(x, h_i(c_i, c_j^R(\theta_j), c_P^R))v_i'(x)dx + \pi(\bar{\theta}_i, h_i(c_i, c_j^R(\theta_j), c_P^R)) \\
- \; & \pi(\bar{\theta}_i, h_i(c_i^R(\bar{\theta}_i), c_j^R(\theta_j), c_P^R))]]
\end{aligned}
$$

is maximized at $c_i = c_i^R(\theta_i)$. ∎

With exponential distributions, the marginal virtual cost $v_i'(\theta_i)$ is constant. This enables the Principal to costlessly solve the incentive problem with respect to the communication strategy choice by agents. Each agent seeks to maximize his own information rents, as represented by the objective function (7). But the Principal's objective is (6), which treats these information rents as costs incurred rather than benefits received. This divergence of objectives is the source of the incentive problem, which needs to be addressed by suitable choice of transfers. The payments for given output levels need to adjusted to reduce the discrepancy between the agents' and Principal's objective functions expressed as function of the production levels. The precise adjustment factor generally depends on the agent's true cost realization. Elicitation of this information is not possible owing to the constraints on communication. If however the required adjustment factor is independent of the agent's true type, which happens to be the case if costs are exponentially distributed, convergence of their objectives can be secured without any communication between agents and P.

5.2 Decentralized versus Centralization of Production Decisions

We now show that with exponential cost shocks, any optimal mechanism requires production decisions to be completely decentralized. The intuitive reason is that the incentives of agents and the Principal are sufficiently aligned that production decisions can be delegated to agents in order to achieve greater flexibility: this discretionary power is profitable for the agents, and therefore also for the Principal.

Proposition 4 *Suppose costs are exponentially distributed. Then in any optimal mechanism, production decisions must be completely decentralized.*

Proof. The proof of this relies on the following Lemma.

Lemma 1 *In the optimal mechanism, the i's optimal communication strategy $c_i^*(\theta_i)$ is almost everywhere locally constant.*

Proof.

Step 1

In what follows fix the optimal communication plan for P and suppress it in the notation, while focusing on optimal choice of communication plan by each agent. With $h_i = h_i(c_i, c_j)$, the optimal production $\hat{q}_i^*(\theta_i, h_i)$ and communication plan $c_i^*(\theta_i)$ for agent i satisfy

$$\begin{aligned} \hat{q}_i^*(\theta_i, h_i) &= \arg\max_{q_i} E_{\theta_j}[V(q_i, \hat{q}_j^*(\theta_j, h_j(c_i, c_j^*(\theta_j)))) \mid h_i(c_i, c_j^*(\theta_j)) = h_i] \\ &- v_i(\theta_i) q_i \end{aligned}$$

$$\begin{aligned} c_i^*(\theta_i) &\in \arg\max_{c_i \in \mathcal{C}_i} \pi_i(\theta_i, c_i) \equiv E_{\theta_j}[V(\hat{q}_i^*(\theta_i, h_i(c_i, c_j^*(\theta_j)), \hat{q}_j^*(\theta_j, h_j(c_i, c_j^*(\theta_j)))) \\ &- v_i(\theta_i)\hat{q}_i^*(\theta_i, h_i(c_i, c_j^*(\theta_j))) - v_j(\theta_j)\hat{q}_j^*(\theta_j, h_j(c_i, c_j^*(\theta_j)))] \end{aligned}$$

Since $v_i(\theta_i)$ is continuous, the Maximum Theorem implies that $\pi(\theta_i, c_i)$ is a continuous function of θ_i, for any c_i .

Step 2: *If $\pi(\theta_i, c_i)$ is continuous in θ_i for every c_i and the number of elements in $\mathcal{C}_i$ is finite, there exists $c_i^*(\theta_i)$ so that*

$$c_i^*(\theta_i) \in \arg\max_{c_i \in \mathcal{C}_i} \pi(\theta_i, c_i)$$

and $c_i^*(\theta_i)$ is almost everywhere locally constant.

Suppose not. Then there exists a non-degenerate interval $(\theta_i^*, \theta_i^{**})$ over which the optimal communication strategy cannot be selected to be a constant strategy for any subinterval. In other words, for any θ_i in this interval, if $\hat{c}_i(\theta_i) \in \arg\max_{c_i \in \mathcal{C}_i} \pi(\theta_i, c_i)$, then in any neighborhood of θ_i there exists $\theta_i^{'}$, and an alternative communication plan $c_i' \in \mathcal{C}_i$ such that

$$\pi(\theta_i, \hat{c}_i(\theta_i)) \geq \pi(\theta_i, c_i^{'})$$

and

$$\pi(\theta_i^{'}, c_i^{'}) > \pi(\theta_i^{'}, \hat{c}_i(\theta_i)).$$

Now choose arbitrary $\theta_i^0 \in (\theta_i^*, \theta_i^{**})$. $B(\theta_i^0)$ and $C_i(\theta_i^0)$ are constructed as follows:

- $C_i(\theta_i^0)$ is defined as $\{c_i \in C_i \mid \pi(\theta_i^0, \hat{c}_i(\theta_i^0)) = \pi(\theta_i^0, c_i)\}$
- In the case that $\bar{C}_i(\theta_i^0) \equiv C_i \backslash C_i(\theta_i^0)$ is not empty: Since $\pi(\theta_i, c_i)$ is continuous for θ_i, for $c_i \in \bar{C}_i(\theta_i^0)$, there exists neighborhood $B(\theta_i^0, c_i)$ of θ_i^0 so that $\pi(\theta_i^{'}, \hat{c}_i(\theta_i^0)) > \pi(\theta_i^{'}, c_i)$ for any $\theta_i^{'} \in B(\theta_i^0, c_i)$. Since there are a finite number of elements in $\bar{C}_i(\theta_i^0)$, $\Pr(\cap_{c_i \in \bar{C}_i(\theta_i^0)} B(\theta_i^0, c_i)) > 0$. Define $B(\theta_i^0) \equiv \cap_{c_i \in \bar{C}_i(\theta_i^0)} B(\theta_i^0, c_i)$. Then for any $\theta_i^{'} \in B(\theta_i^0)$, $\pi(\theta_i^{'}, c_i^{'}) < \pi(\theta_i^{'}, \hat{c}_i(\theta_i^0))$ for any $c_i^{'} \in \bar{C}_i(\theta_i^0)$.
- In the case that $\bar{C}_i(\theta_i^0) \equiv C_i \backslash C_i(\theta_i^0)$ is empty: $B(\theta_i^0)$ is chosen as an arbitrary neighborhood of θ_i^0.

By hypothesis, there exists $\theta_i^1 \in B(\theta_i^0)$ and $c_i^{'}$ so that

$$\pi(\theta_i^1, c_i^{'}) > \pi(\theta_i^1, \hat{c}_i(\theta_i^0)).$$

From the construction of $B(\theta_i^0)$, $c_i^{'} \in C_i(\theta_i^0)$. Next construct $C_i(\theta_i^1)$ and $B(\theta_i^1)$ from θ_i^1 using the same procedure as in the construction of $C_i(\theta_i^0)$ and $B(\theta_i^0)$ from θ_i^0. Since $C_i(\theta_i^1) \subset C_i(\theta_i^0)$ and $C_i(\theta_i^1)$ does not include $\hat{c}_i(\theta_i^0)$, the number of elements is strictly smaller in $C_i(\theta_i^1)$.

In a manner similar to the choice of θ_i^1 given θ_i^0, we can choose $\theta_i^2 \in B(\theta_i^1)$ and construct $C_i(\theta_i^2)$ and $B(\theta_i^2)$. This procedure can be repeated until the number of elements in $C_i(\theta_i^k)$ becomes one. Then since $\hat{c}_i(\theta_i)$ is constant for $\theta_i \in B(\theta_i^k)$, we obtain a contradiction, thus completing the proof of Lemma 1. ∎

Return now to the proof of Proposition 4. Lemma 1 implies that there exist $\bar{c}_i \in \mathcal{C}_i$ and non-degenerate interval $[\theta_i^{'}, \theta_i^{''}] \subset \{\theta_i \mid c_i^*(\theta_i) = \bar{c}_i\}$. $q_i^*(\theta_i, \theta_j) = \hat{q}_i^*(\theta_i, h_i(\bar{c}_i, c_j^*(\theta_j), c_P^*))$ is strictly decreasing in θ_i on $[\theta_i^{'}, \theta_i^{''}]$, since this solves

$$\max_{q_i} E_{\theta_j}[V(q_i, \hat{q}_j^*(\theta_j, h_j(\bar{c}_i, c_j^*(\theta_j), c_P^*))) \mid h_i(\bar{c}_i, c_j^*(\theta_j), c_P^*) = h_i] - v_i(\theta_i) q_i.$$

On the other hand, $h_j = h_j(\bar{c}_i, c_j^*(\theta_j), c_P^*)$ and $h_P = h_P(\bar{c}_i, c_j^*(\theta_j), c_P^*)$ are independent of θ_i on $[\theta_i^{'}, \theta_i^{''}]$. This implies that q_i^* is not measurable with respect to h_j and h_P. ∎

5.3 Centralization versus Decentralization of Communication

We have seen above that with exponential costs, incentives can be aligned so that greater exchange of information among agents is profitable for the Principal, and agents are willing to follow the Principal's instructions concerning choice of communication strategy. If one communication protocol p allows agents to obtain finer information (in the sense of Blackwell) than another protocol p', then p will be more profitable.

Centralized communication systems restrict all communication between agents to be chanelled through the Principal (or a central coordination device designed by P). The latter forms a bottleneck with regard to the flow of information: for agent i to communicate with agent j, first i has to send a message to P, which P can relay or process and send to j. Allowing direct communication among agents should permit speedier and greater information exchange. For this reason centralized communication systems will be dominated by decentralized ones.

We now explore this idea formally. In order to compare different communication protocols we have to explicitly specify feasibility constraints, which requires an underlying model of the time it takes an agent to read and write messages. The class of communication protocols considered here includes the structured protocols described in Section 3, as well as a more general and flexible set of protocols. It requires us to make explicit more detailed features of the communication technology, which we model in the fashion of email exchange systems (though it includes structured protocols as a special category as well).

Assume that information is exchanged in some common language in binary form between $t = 0$ and T. Let $\mathcal{L}$ denote $\{0, 1\}^\infty$, and $l(\theta_i) \in \mathcal{L}$ the representation of the real number θ_i as a countably infinite string of 0's and 1's. It takes time for a player to write a message in (or translate into) this language, and for other players to read messages written by others.

Messages are sent in binary form, as a finite string of 0's and 1's. We can break this down into each constituent 'bit' of information, consisting of a single binary 0-1 message. Each message is thus a string of bits. Each bit takes some time to write for the sender, and some time for the receiver to read. Time is measured in discrete intervals, each of which is assumed small enough that the time taken to read or write a single bit of information is a positive integer (number of intervals) for every agent. Let the time taken by agent i to write a single bit be denoted by R_i, and to read a single bit be W_i, where R_i, W_i are both positive integers.

In this setting, if agent i were to start writing a message consisting of k bits at time point t, agent i would be occupied from time t until $(t + kW_i)$ in writing the message. It could be sent at any time point s at or after $(t+kW_i)$. Messages take no time in reaching the receivers. So agent j would receive it at s, the same instant that it was sent by i. Agent j would know at s that a message has been received from i, as well as the length (k) of this message. We assume that observing the list (and length) of messages received from others at any point of time is not itself time-consuming. Moreover, we make the simplifying assumption that when agents start to read or write a given message they do not have the option of interrupting this activity in-between (e.g., in the light of new information, such as the arrival of new messages, or upon reading the contents of the first part of a message, the receiver cannot decide to stop reading any further).

At any point of time t, agent j would thus know the list and length of messages received from others at all prior times $t' \leq t$. Of these, some have already been read, and others are unread. The history observed by j upto t consists of contents of messages read by j until t, and the list and length of hitherto unread messages received by j until t. We denote this history by $h_j(t)$. The corresponding set of all possible histories until t is denoted $\mathcal{H}_j(t)$.

The communication protocol specified by the mechanism has a given communication network, defining for each agent i the set of others $\mathcal{S}_i$ in the organization to whom i can send messages, and those $\mathcal{R}_i$ from whom i can receive messages. We abstract from the costs of copying or transmitting messages: only reading and writing activities are time-consuming. A message once written by i can thus be sent to any subset of $\mathcal{S}_i$.[18] A *centralized communication network* is one where every agent $i \neq P$ communicates only with the Principal, i.e., $\mathcal{S}_i = \mathcal{R}_i = \{P\}$.[19] A communication network is said to be *decentralized* if it is not centralized, and is said to be *completely decentralized* if neither the Principal nor any external coordinator participates in the network.

The structured communication protocols considered earlier form a particular class of protocols, with additional restrictions: messages may be sent

[18] This permits different messages to be sent to different recipients.

[19] It is easy to see that our results will extend to networks where everyone communicates instead with an external coordinator appointed by the Principal.

only at designated time points, with intervening intervals between successive time points long enough (relative to the length of allowed messages and reading/writing abilities of all agents) to allow each agent enough time to read the messages recently received, and then write new messages to be sent at the next time point. Given the reading and writing abilities of agents, this imposes an upper bound on the size of the message spaces at each round. Other protocols need not impose the requirement that messages be exchanged by all agents in the organization at specific time points; nor do they restrict the size of messages at any given point. Agents can then decide more flexibly on the timing and length of their messages: e.g., depending on the content of a message received an agent may decide on the length and detail of their reply.[20]

In a general protocol, a *communication plan for i* is described as follows. Consider any time point t at which i is 'free', i.e., i is not in the middle of reading or writing a message that was started previously at $t' < t$. Given the history $h_i(t)$, agent i will decide either to start writing a fresh message to a subset of $\mathcal{S}_i$, or start reading one of the hitherto unread messages received by i at or before t, or do nothing. Let $\mathcal{M}_i^R(t)$ and $\mathcal{M}_i^U(t)$ denote the list of messages read and unread respectively by i until t. The communication plan thus specifies a plan to read or write a new message: $r_{it}(h_i(t)) \in \{\emptyset \cup \mathcal{M}_i^U(t)\}$, $w_{it}(h_i(t)) \in \{\emptyset \cup \mathcal{N}_i(t)\}$, where $\mathcal{N}_i(t)$ denotes the set of messages that could be written in the time remaining (i.e., between t and T) and sent to a subset of $\mathcal{S}_i$. These reading and writing plans must satisfy the restriction that r_{it} and w_{it} cannot both be non-empty at the same time.

In any given protocol, let $\mathcal{C}_i$ denote the set of all possible communication plans for i. A vector of communication plans for the different players $\tilde{c} \equiv \{c_1, c_2, c_P\}$ induces a personal history $h_i(t|\tilde{c})$ for every player and every date t.

Note in this setting that agents are allowed to learn both from the contents of messages they have received and read so far, as well as the list and length of messages that they have been sent to them and have not been read. In particular, some information can be 'costlessly' communicated by the act of **not** sending a message: if j has not received a message from i at some time point t, this communicates some information to j at or after t. The act of

[20] In a structured protocol the interval between time points would need to be expanded for all agents, to allow any one agent more time to write more detailed messages; such constraints do not arise in unstructured protocols.

not sending a message involves no time delays for either the potential sender and receiver.[21] This feature plays a role in the proof below that decentralized communication networks strictly dominate centralized ones, though the proof of weak domination does not.

Say that one communication strategy vector is *at least as informative* (resp. *more informative*) than another if the former induces an information partition for either agent which is at least as fine in the Blackwell sense as the information partition induced by the latter, in (almost) all states (resp. is strictly finer for a set of states with positive probability). This implies that the production allocation can be unambiguously 'more flexible', and higher profits can be achieved in the team-theoretic problem of maximizing expected profits (5). Hence a Blackwell improvement in information of agents will allow P to benefit if costs are exponentially distributed.

Proposition 5 *Consider the class of communication protocols described in this subsection. Assume that:*

(i) there exists an agent i who is at least as efficient in reading and writing as P: i.e, $R_i \leq R_P$ and $W_i \leq W_P$, and

(ii) costs are exponentially distributed.

Then given any centralized communication network, there exists a completely decentralized network which generates a strictly higher expected profit for the Principal.

We sketch the proof here. Given an equilibrium communication strategy vector in any centralized network, construct the following decentralized network and corresponding communication strategy vector. Select any agent i whose reading and writing abilities are as good as P's and construct the following decentralized network where i replaces P as the central node. Every other agent j will communicate directly with i instead of P: in all other respects j's communication strategy remains the same. At the same time, the new nodal agent i will duplicate P's communication strategy with respect to $j \neq i$ in the centralized network. Since $R_i \leq R_P, W_i \leq W_P$, this communication strategy is feasible for i.

[21]In similar fashion, some information can be communicated by a message even though its contents have not been read by the receiver, since the receiver learns something simply from the fact that a message of such a length was sent by the sender. Such a message however involves time delays in their writing by the sender.

Then $j \neq i$ will observe exactly the same history in every state, since she writes the same messages as before, and reads the same messages from i that she read previously from P. Hence j's information partition will be the same as in the centralized communication system. On the other hand, i has at least as much information about θ_j than in the centralized system, since he reads all messages written by j (rather than messages written by P based on the latter's reading of the messages written by j and i). The information partitions attained in the decentralized network are now at least as fine as they were in the centralized network, allowing at least the same expected profit to be attained.

We now explain how a further variation in communication plans can be constructed in the decentralized network in order to yield a strict improvement. Note that in the original set of communication plans employed in the centralized network, there exists a time interval of length at least $W_i + W_P$ at the beginning of the communication phase during which j receives no message from P. The reason is that it takes P at least W_P units of time to write a message. Moreover, messages from P to j convey information concerning i's state, so must be based in turn on prior messages received by P from i, which must take the latter at least W_i time to write.

Following the switch to the communication plans in the decentralized network in which i mimics the role of P in communicating with j, note that i no longer has to spend time writing messages to P which are then relayed to j. Hence at least W_i units of time are freed up for i at the beginning of the communication phase, since i no longer has to write to P. During this time i can send an additional message to j.

The specific deviation in i's communication plan is constructed as follows. Lemma 1 showed that in any optimal mechanism, i's communication strategy is locally constant with probability one. Select a nondegenerate interval (θ_i, θ_i') over which i's communication strategy is locally constant. Let agent i augment his assigned communication strategy (which replicated P's strategy with respect to $j \neq i$) in the following way. Select an arbitrary $\tilde{\theta}_i \in (\theta_i, \theta_i')$. Let i send an additional binary message ($= 1$, say) to j in the first W_i time units of the communication phase if and only if $\theta_i > \tilde{\theta}_i$. If this condition is not satisfied, i does not send any message during this time, and continues with the communication plan that mimics P's communication plan with j.

This will enable agent j to learn whether $\theta_i > \tilde{\theta}_i$, information that was not communicated to j in the centralized network. Note that it takes j no additional reading time to learn this information, since it is communicated simply upon observing whether i sent a message to j at time W_i following the start of the communication phase. Hence j is not required to select a different communication plan in order to obtain this additional information, which can be utilized to construct a more flexible production allocation compared with the centralized system.[22]

6 Optimality of Decentralized Contracting

In this section we show that with exponentially distributed costs, decentralized contracting can implement the same expected payoff for P as centralized contracting. This shows that the corresponding result of MMR (1992, 1995) extends to the current context, under similar assumptions concerning verifiability of production assignments and subcontracting cost incurred by the intermediary. The idea is similar: P can delegate subcontracting and production decisions to i with a linear 'profit-center' contract, where profits are computed as revenues (V) minus a (subsidized) adjustment for an accounting measure of cost (($t_j + t_H$) plus a fixed per-unit reimbursement of cost of input q_i delivered by the manager evaluated at the lowest possible cost $\underline{\theta}_i$).

However, there are some additional complications in the current setting. In the absence of communication delays, P can ask i to submit a report of his own cost at the time of contracting, and condition the subsidy on subcontracting on this report. This is necessary to overcome the problem of 'double marginalization of rents' arising from the monopsony power of i over the subcontractor. Subsidizing subcontract costs neutralizes the inclination of i to maximize his own rents at the expense of the subcontractor by outsourcing too little to the latter. The exact subsidy rate generally depends on the true cost realization of i. If communication delays prevent full revelation of this cost, it may be impossible to overcome the problem of double marginalization of rents completely.

However, when costs are exponentially distributed, the optimal subsidy rate is independent of i's true cost. This obviates the need for any cost report to be made by i at the time of contracting with P: a linear profit center contract can cause the contractor to internalize P's objectives perfectly.

[22] Agent j now learns whether agent i's cost lies in $(\theta_i, \tilde{\theta}_i)$ or $(\tilde{\theta}_i, \theta_i')$, across which the conditional expectation of i's virtual cost is different. This can be used to create distinct production assignments corresponding to these different events, something that was not possible in the centralized network.

Proposition 6 *If costs are exponentially distributed, decentralized contracting (with either agent selected as the manager) can attain the same expected profit for P as centralized contracting.*

Proof. Suppose that P selects agent i as manager and offers the following contract to i:

$$t_i = \frac{1}{\lambda_i + 1}[V(q_i, q_j) + \lambda_i(t_j + t_H) + \lambda_i \underline{\theta}_i q_i - \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))].$$

where[23]

$$\begin{aligned} & \pi(\theta_i, c_i) \\ = \; & E_{\theta_j}[V(\hat{q}_i^*(\theta_i, h_i(c_i, c_j^*(\theta_j), c_P^*)), \hat{q}_j^*(\theta_j, h_j(c_i, c_j^*(\theta_j), c_P^*))) \\ - \; & v_i(\theta_i)\hat{q}_i^*(\theta_i, h_i(c_i, c_j^*(\theta_j), c_P^*)) - v_j(\theta_j)\hat{q}_j^*(\theta_j, h_j(c_i, c_j^*(\theta_j), c_P^*))] \end{aligned}$$

and $(\hat{q}_i^*(\theta_i, h_i), \hat{q}_j^*(\theta_j, h_j), \tilde{c}^*(\theta))$ denotes the optimal allocation in the centralized system. In this system, i is not required to communicate anything to P.

Taking this contract as given, i's payoff equals

$$t_i - \theta_i q_i - t_j - t_H = \frac{1}{\lambda_i + 1}[V(q_i, q_j) - v_i(\theta_i)q_i - t_j - t_H - \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))]$$

where we use the fact that with exponentially distributed cost, $v_i(\theta_i) = (1 + \lambda_i)\theta_i - \lambda_i \underline{\theta}_i$. With the objective of maximizing the expected value of this payoff, i designs $t_j(q_i, q_j, h)$, $t_H(q_i, q_j, h)$ and a communication protocol for the network involving i, j and the coordination device M. Since i does not have to send any message to P in the upper-level network, every communication protocol feasible under centralization is also feasible here.

Given any communication protocol, arguments analogous to those in Proposition 2 establish that $(q_i(\theta_i, \theta_j), q_j(\theta_i, \theta_j), \tilde{c}(\theta))$ is implementable in this decentralized system if and only if there exists $t_j(q_i, q_j, h)$ and $t_H(q_i, q_j, h)$ so that

(i) $q_j(\theta_i, \theta_j) = \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))$ where

$$\begin{aligned} \hat{q}_j(\theta_j, h_j) = \arg\max_{q_j} & E_{\theta_i}[t_j(\hat{q}_i(\theta_i, h_i(c_i(\theta_i), c_j, c_M)), q_j, h(c_i(\theta_i), c_j, c_M)) \\ & \mid h_j(c_i(\theta_i), c_j, c_M) = h_j] - \theta_j q_j \end{aligned}$$

[23] Since the constructed mechanism will be chosen so as to generate an allocation identical to that attained by an optimal centralized contract, we shall use the same notation for histories used in the latter, and use agent-specific subscripts.

(ii) $q_i(\theta_i, \theta_j) = \hat{q}_i(\theta_i, h_i(\tilde{c}(\theta)))$ where

$$\begin{aligned}
& \hat{q}_i(\theta_i, h_i) \\
= \; & \arg\max_{q_i} E_{\theta_j}[V(q_i, \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M))) \\
- \; & t_j(q_i, \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M)), h(c_i, c_j(\theta_j), c_M)) \\
- \; & t_H(q_i, \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M)), h(c_i, c_j(\theta_j), c_M)) \mid h_i(c_i, c_j(\theta_j), c_M) = h_i] \\
- \; & v_i(\theta_i) q_i
\end{aligned}$$

(iii) $c_j(\theta_j)$ maximizes

$$\begin{aligned}
& E_{\theta_i}[t_j(\hat{q}_i(\theta_i, h_i(c_i(\theta_i), c_j, c_M)), \hat{q}_j(\theta_j, h_j(c_i(\theta_i), c_j, c_M)), h(c_i(\theta_i), c_j, c_M)) \\
& -\theta_j \hat{q}_j(\theta_j, h_j(c_i(\theta_i), c_j, c_M))]
\end{aligned}$$

(iv) $c_i(\theta_i)$ maximizes

$$\begin{aligned}
& E_{\theta_j}[V(\hat{q}_i(\theta_i, h_i(c_i, c_j(\theta_j), c_M)), \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M))) \\
- \; & t_j(\hat{q}_i(\theta_i, h_i(c_i, c_j(\theta_j), c_M)), \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M)), h(c_i, c_j(\theta_j), c_M)) \\
- \; & t_H(\hat{q}_i(\theta_i, h_i(c_i, c_j(\theta_j), c_M)), \hat{q}_j(\theta_j, h_j(c_i, c_j(\theta_j), c_M)), h(c_i, c_j(\theta_j), c_M)) \\
- \; & v_i(\theta_i) \hat{q}_i(\theta_i, h_i(c_i, c_j(\theta_j), c_M))]
\end{aligned}$$

(v) The participation constraint for j:

$$\begin{aligned}
& E_{\theta_i}[t_j(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta))) \\
& -\theta_j \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))] \geq 0
\end{aligned}$$

(vi)The participation constraint for i:

$$\begin{aligned}
& E_{\theta_j}[V(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))) \\
- \; & t_j(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta))) \\
- \; & t_H(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta))) \\
- \; & v_i(\theta_i) \hat{q}_i(\theta_i, h_i(\tilde{c}(\theta)))] \geq \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))
\end{aligned}$$

(vii) The participation constraint for the third party:

$$E_{\theta_i, \theta_j}[t_H(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta)))] \geq 0$$

From (i),(iii) and (v),

$$\begin{aligned} & E_{\theta_i}[t_j(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta))) - \theta_j \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))] \\ = \; & E_{\theta_i}[\int_{\theta_j}^{\bar{\theta}_j} \hat{q}_j(x, h_j(c_i(\theta_i), c_j(x), c_M))dx + U_j(h_j(c_i(\theta_i), c_j(\bar{\theta}_j), c_M))] \end{aligned}$$

and

$$E_{\theta_i}[U_j(h_j(c_i(\theta_i), c_j(\bar{\theta}_j), c_M))] \geq 0$$

Then i's ex-ante payoff equals

$$\begin{aligned} & \frac{1}{\lambda_i + 1} E[V(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))) \\ - \; & t_j(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta))) - v_i(\theta_i)\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))) \\ - \; & \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i)) \\ - \; & t_H(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta)))] \\ = \; & \frac{1}{\lambda_i + 1} E[V(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta)))) \\ - \; & v_j(\theta_j)\hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))) - v_i(\theta_i)\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))) \\ - \; & \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i)) - U_j(h_j(c_i(\theta_i), c_j(\bar{\theta}_j), c_M)) \\ - \; & t_H(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta)))] \end{aligned}$$

An upper bound for this expected payoff is given by that corresponding to the optimal allocation under centralized contracting:$\hat{q}_i^*(\theta_i, h_i)$, $\tilde{c}^*(\theta)$ and

$$E_{\theta_i}[U_j(h_j(c_i(\theta_i), c_j(\bar{\theta}_j), c_M))] = 0.$$

with

$$E[t_H(\hat{q}_i(\theta_i, h_i(\tilde{c}(\theta))), \hat{q}_j(\theta_j, h_j(\tilde{c}(\theta))), h(\tilde{c}(\theta)))] = 0.$$

This upper bound to the expected payoff of i equals

$$\begin{aligned} & \frac{1}{\lambda_i + 1} E[V(\hat{q}_i^*(\theta_i, h_i(\tilde{c}^*(\theta))), \hat{q}_j^*(\theta_j, h_j(\tilde{c}^*(\theta)))) \\ - \; & v_j(\theta_j)\hat{q}_j^*(\theta_j, h_j(\tilde{c}^*(\theta))) - v_i(\theta_i)\hat{q}_i^*(\theta_i, h_i(\tilde{c}^*(\theta))) - \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))] \\ = \; & \frac{1}{\lambda_i + 1} E[\pi(\theta_i, c_i^*(\theta_i)) - \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))] \\ = \; & E[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i^*(x, h_i(c_i^*(x), c_j^*(\theta_j), c_M^*))dx]. \end{aligned}$$

This upper bound can be achieved by i by selecting the same communication protocol as under centralized contracting (with the coordination role of P replaced by M with $c_M = c_P^*$), and the following subcontracts. Since $\hat{q}_j^*(\theta_j, h_j)$ is non-increasing in θ_j, define $\theta_j(q_j, h_j)$ and $t_j(q_j, h_j)$ as follows.

$$\theta_j(q_j, h_j) \equiv \sup_{\theta_j}\{\theta_j \mid \hat{q}_j^*(\theta_j, h_j) \geq q_j\}$$

and

$$t_j(q_j, h_j) \equiv \theta_j(q_j, h_j)q_j + \int_{\theta_j(q_j,h_j)}^{\bar{\theta}_j} \hat{q}_j(x, h_j)dx + U_j(h_j)$$

where $U_j(h_j)$ is chosen exactly as in the optimal centralized contract. Similarly define $\theta_i(q_i, h_i)$ and $s_i(q_i, h_i)$ as follows.

$$\theta_i(q_i, h_i) \equiv \sup_{\theta_i}\{\theta_i \mid \hat{q}_i^*(\theta_i, h_i) \geq q_i\}$$

and

$$s_i(q_i, h_i) \equiv \theta_i(q_i, h_i)q_i + \int_{\theta_i(q_i,h_i)}^{\bar{\theta}_i} \hat{q}_i(x, h_i)dx + U_i(h_i)$$

where $U_i(h_i)$ is chosen as in the optimal centralized contract.

Let i offer $t_j(q_j, h_j)$ to j and select payments

$$t_H(q_i, q_j, h_i, h_j) = V(q_i, q_j) - t_j(q_j, h_j) - (1+\lambda_i)s_i(q_i, h_i) + \lambda_i\underline{\theta}_i q_i - \pi(\bar{\theta}_i, c_i^*(\bar{\theta}_i))$$

to a third party. Then i's payoff at $t = T$ following history h_i will equal

$$s_i(q_i, h_i) - \theta_i q_i,$$

and j's payoff following observed history h_j at T will be

$$t_j(q_j, h_j) - \theta_j q_j.$$

With these contracts, continuation payoffs for i and j from $t = 0$ onwards will be exactly the same as in the optimal centralized contract, so the same participation, communication and production decisions constitute a PBE of the continuation game from $t = 0$ onwards. Moreover, by construction

$$E[t_H(\hat{q}_i^*(\theta_i, h_i(\tilde{c}^*(\theta))), \hat{q}_j^*(\theta_j, h_j(\tilde{c}^*(\theta))), h_i(\tilde{c}^*(\theta)), h_j(\tilde{c}^*(\theta)))] = 0,$$

implying that it will be optimal for concerned third parties to participate. Hence the upper bound payoff for i will be implemented. At the same time P's ex-ante payoff is

$$\begin{aligned} & E[V(\hat{q}_i^*(\theta_i, h_i(\tilde{c}^*(\theta))), \hat{q}_j^*(\theta_j, h_j(\tilde{c}^*(\theta)))) \\ - & v_i(\theta_i)\hat{q}_i^*(\theta_i, h_i(\tilde{c}^*(\theta))) - v_j(\theta_j)\hat{q}_j^*(\theta_j, h_j(\tilde{c}^*(\theta)))] \end{aligned}$$

since

$$\begin{aligned} & V(q_i, q_j) - t_i(q_i, q_j, t_j(q_j, h_j), t_H(q_i, q_j, h_i, h_j)) \\ = & V(q_i, q_j) - s_i(q_i, h_i) - t_j(q_j, h_j) - t_H(q_i, q_j, h_i, h_j). \end{aligned}$$

∎

One difference from the construction in MMR(1992) is that payments to a third party are used here. This owes to the fact that subcontracts must be designed *ex ante* rather than at the *interim* stage, owing to the communicational restrictions that apply to subcontract design. This gives rise to a potential incentive problem akin to an informed principal problem: the manager will select the communication strategy at the *interim* stage after learning her own private information. The *ex ante* optimal communication strategy for the manager may not be sequentially rational. Payments to the third party represent a commitment device that enable this incentive problem to be overcome. In the proof above, third party payments are designed so that the *ex post* payoffs of the manager are exactly as in the centralized contract setting, where the appropriate incentive constraints for the appropriate communication strategies are satisfied.

7 Extension to Case of Noisy Communication

The preceding results were based on the assumption of absence of any errors in communication — that messages are received in exactly the same form that were sent. An alternative source of friction in communication could result from random deviations between messages received and sent owing to 'noise' in communication channels, or intended to be sent e.g., if there are random errors in speaking or writing. In that event, one would typically expect that messages received by any agent are verifiable, whilst the messages sent (or intended to be sent) are not. The significance of communication errors in organizational design has been stressed by Williamson (1967), who argued that errors in 'serial reproduction' of information that travels through many layers is an important source of 'loss of control' in vertical hierarchies.

This section extends the preceding model to incorporate random (exogenous) errors in communication, and shows all the previously mentioned results continue to apply, provided the communication protocols are structured (i.e., in the form of a fixed number of rounds of communication, with assigned message spaces and communication networks at each round).[24]

In particular, our result concerning equivalence of decentralized and centralized contracting contrasts with Williamson's assertion that communication errors give rise to 'vertical control loss' in multi-tier hierarchies. The reason is simple: in our model production decisions are decentralized also, so information does not need to travel across more than one layer before they are incorporated in decisions.[25]

With random errors in communication that cause received messages to deviate from the messages that were sent (or intended to be sent), it is reasonable then to suppose that only received messages are *ex post* verifiable, while messages actually sent or intended to be sent are not. (In what follows, we shall use 'messages sent' to represent either of these possibilities, since they give rise to the same theory). This introduces an element of privacy of messages — only a sender knows the message that was sent. The history of messages h_i observed by i now includes messages sent by i, apart from messages received by any $j \in \mathcal{S}_i$. Only the latter are *ex post* verifiable.

Provided we restrict attention to the structured communication protocols described in Section 3, and provided communication errors follow an exogenous distribution, we now explain how our preceding results extend to this setting.

Focus on the centralized contracting system, where the set of participants involved in the communication protocol comprise P and the two production agents. Let m_{ijk} denote a message sent by i to j at round k of the communi-

[24]We do not, however, consider the possibility of strategic interference by agents with respect to each other's communication channels, a context in which Legros and Newman (1999, 2002) address a number of interesting questions.

[25]This raises an interesting question for future research: whether communication 'noise' degrades the performance of multi-tier hierarchies in more general settings (e.g., more complex hierarchies with more than one branch, or where some 'public good' decisions need to be made by top-level managers).

cation protocol, and $\tilde{m}_{ijk} = \phi(m_{ijk}, \eta_{ijk})$ the corresponding message received by j, where η_{ijk} denotes the communication error associated with this message. The realization of $\tilde{m}_{ijk}$ is known by j and i, and is *ex post* verifiable by contract enforcers. Whereas only i knows the message m_{ijk} that was originally sent or intended to be sent. The history observed by i until round k is denoted by $h_{ik} = (h_{ik}^N, h_{ik}^V)$ where $h_{ik}^N = \{m_{ij\tau}\}_{j \neq i, \tau=1,..,K}$ is non-verifiable and $h_{ik}^V = \{\tilde{m}_{ij\tau}, \tilde{m}_{ji\tau}\}_{j \neq i, \tau=1,..,K}$ is verifiable. A communication plan c_i selects messages m_{ijk} from a finite message set $\mathcal{M}_{ijk}$ to be sent at any round k to others in the same communication network as i, as a function of $h_{i,k-1}$. Let η denote the vector of errors $\{\eta_{ijk}\}_{j \neq i, k=1,...,K}$. Then the history observed by i until round k can be represented as $h_{ik} = h_{ik}(c_i, c_j, c_P, \eta)$, $h_{ik}^N = h_{ik}^N(c_i, c_j, c_P, \eta)$, $h_{ik}^V = h_{ik}^V(c_i, c_j, c_P, \eta)$. Finally use h_k^V to denote $(h_{ik}^V, h_{jk}^V, h_{Pk}^V)$, the set of verifiable messages in round k, which can also be expressed as a function of (c_i, c_j, c_P, η). The errors η follows some given distribution which is common knowledge.

The definition of centralized and decentralized contracting games and allocations attainable by them can now be extended in a straightforward manner. The description of the state space is extended to include the realization of cost shocks (θ_1, θ_2) as well as the communication error vector η, which have exogenous distributions that are common knowledge among the agents and P. Allocations are now contingent on both cost shocks and communication errors.

Contractual payments t_a to agent a can be conditioned only on the verifiable messages a receives or others receive from a, summarized by h_a^V, which depends on the chosen communication plans c_i, c_j, c_P as well as the error η. Production decisions made by agent a can be based in addition on messages a actually sent, which are non-verifiable. Despite this discrepancy between the basis of contracts and production decisions, a characterization of allocations attainable by centralized contracting analogous to Proposition 2 continues to obtain. This is presented in the Appendix. Since all other extensions are fairly routine, we do not present them in the paper.

8 Concluding Comments

This paper developed a theory of mechanism design for a production team in a context where agents and Principal have limited capacity to communicate with one another. Its main purpose was to show how the Marschak-Radner view that the value of decentralization derives from limitations on ability of agents to communicate, applies in a setting with an incentive problem. As is well-known, in a context of unlimited communication and commitment ability of the Principal, attention can be focused on revelation mechanisms every aspect of which — contracting, production decisions and communication — are centralized. This flies in the face of pervasiveness of delegation of decision-making from owners of firms to managers, or customers to primary contractors or trading intermediaries. Previous attempts to adapt mechanism design theory to contexts of limited communication in order to explain the value of decentralized mechanisms were based on a number of *ad hoc* assumptions. These included unnatural restrictions on centralized contracts, and on communication systems.

The approach of this paper avoided imposing such restrictions on the class of mechanisms, apart from restricting (possibly dynamic) communication plans to a finite set. An analogue of the Revelation Principle was obtained: attention can be restricted to centralized contracting mechanisms, though production and communication systems cannot be centralized in general. This helps clarify the nature of the precise argument for decentralization, i.e., that it pertains to production decisions and communication, rather than contracting rights.

If messages received by agents are *ex post* verifiable, a general characterization of attainable allocations in terms of incentive and communication constraints was obtained. The communication constraints are of two kinds: restricted coordination of production assignments, and incentive constraints with respect to communication. In general these constraints are complicated.

In order to obtain more specific results we considered the case where cost shocks are exponentially distributed. Here we could show that the incentive constraints with respect to communication do not bind, as incentive schemes can be designed to align the objectives of the Principal and the agents. Then the only constraints represented by limited communication pertain to limited coordination of production assignments: the more information can be brought to bear on production decisions, the better for the Principal. Hence it is essential to let each agent decide his or her own production, after exchanging messages with others: centralized decision-making about production in the form of 'targets' or 'orders from above' are suboptimal. Moreover, under some weak assumptions on the communication technology, decentralized communication dominate centralized communication. At the same time contracting can be decentralized without any loss of profit, provided the Principal can verify transfers and production assignments between agents. Hence decentralization of operating procedures — communication and production — is essential for optimality. Decentralization of contracting is not essential, but is consistent with optimality under appropriate monitoring capacity of the Principal.

Many open questions remain. To what extent can the results be extended to contexts where cost distributions are not restricted to the exponential family? Can there be a trade-off between centralization and decentralization of production decisions? How can this approach be extended to incorporate more complicated multi-layered organizations?

Yet other more general questions concerning organizational design can also be examined. Under what conditions does the presence of a third party facilitate communication and coordination among production agents? This would provide insight into the role of managers who do not participate in any production activities, whose only role is to process information communicated by production agents and help formulate production plans. In the model presented here, such third party 'coordinators' would have no room for strategic behavior, owing to the assumption of absence of collusion. If the model were extended to accommodate collusion, it would lead to a theory of hierarchies where intermediaries not directly involved in production play a coordinating role and behave strategically. Another question pertains to the effects of changing communication technology on organizational design. Comparative statics of such a model with respect to information technology could generate predictions that could be tested against empirical patterns of how these have been changing in recent times (a brief overview of which is provided in Mookherjee (2006)).

References

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

Bolton, P. and M. Dewatripoint (1994), "The Firm as a Communication Network", *Quarterly Journal of Economics*, 109(4), 809-839.

Cremer, J. (1995), "Arm's Length Relationships," *Quarterly Journal of Economics*, 110(2), 275-95.

Deneckere R. and Severinov S. (2003), "Mechanism Design and Communication Costs," mimeo, Department of Economics, University of Wisconsin and Fuqua School of Business, Duke University.

Dessein, W. (2002), "Authority and Communication in Organizations", *Review of Economic Studies*, 69(4), 811-38

Green, J. and J. Laffont (1987), "Limited Communication and Incentive Compatibility", *Information, Incentives, and Economic Mechanisms: Essays in honor of Leonid Hurwicz*, T. Groves, Radner, R. and Reiter, S. (ed.), Minneapolis: University of Minnesota Press

Hayek, F.A. (1945), "The Use of Knowledge in Society", *American Economic Review*, 35(4), 519-530.

Laffont, J. and D. Martimort (1997), "Collusion under Asymmetric Information," *Econometrica*, 65(4), 875-911.

——————— (1998), "Collusion and Delegation," *Rand Journal of Economics* 29(2), 280-305.

——————— (2000), "Mechanism Design with Collusion and Correlation," *Econometrica*, 68(2), 309-42.

Legros, P. and A. Newman (1999), "Interference : Contracts and Authority with Insecure Communication". Discussion Paper, University Libre de Brussels, July 1999.

——————— (2002), "Courts, Contracts and Interference", *European Economic Review*, 46(4/5), 733-44.

Marschak, J. and R. Radner (1972), *Economic Theory of Teams*, New Haven: Yale University Press.

Marschak, T. and S. Reichelstein (1995), "Communication Requirements for Individual Agents in Networks and Hierarchies," *The Economics of Informational Decentralization: Complexity, Efficiency and Stability*, J. Ledyard (ed.), Boston: Kluwer Academic Publishers.

——— (1998), "Network Mechanisms, Informational Efficiency and Hierarchies," *Journal of Economic Theory*, 79(1), 106-41.

Melumad, N. and D. Mookherjee (1989), "Delegation as Commitment: The Case of Income Tax Audits," *Rand Journal of Economics*, 20(2), 139-63.

Melumad, N., D. Mookherjee and S. Reichelstein (1992), "A Theory of Responsibility Centers," *Journal of Accounting and Economics*, 15, 445-484.

——— (1997), "Contract Complexity, Incentives and the Value of Delegation," *Journal of Economics and Management Strategy*, 6(2), 257-289.

Mookherjee, D. and S. Reichelstein (1997), "Budgeting and Hierarchical Control," *Journal of Accounting Research,* 35(2), 129-55.

Mookherjee, D. and S. Reichelstein (2001), "Incentives and Coordination in Hierarchies," *Advances in Theoretical Economics,* 1(1), .

Mookherjee, D. and M. Tsumagari (2004), "The Organization of Supply Networks: Effects of Delegation and Intermediation," *Econometrica*, 72(4), 1179-1220.

Mookherjee, D. (2006), "Decentralization, Hierarchies and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, forthcoming.

Mount, K. and S. Reiter (1974), "The Informational Size of Message Spaces," *Journal of Economic Theory*, 8, 161–191.

Mount, K. and S. Reiter (1995), "A Theory of Computing with Human Agents," working paper, Department of Economics, Northwestern University.

Myerson, R. (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10(1), 67-81.

Poitevin, M. (1995), “Contract Renegotiation and Organizational Design,” mimeo, CIRANO Working Paper No. 95-3, Montreal.

Poitevin, M. (2000), “Can the Theory of Incentives Explain Decentralization?,” *Canadian Journal of Economics,* 33(4), 878-906.

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

———- (1993),“The Organization of Decentralized Information Processing,” *Econometrica*, 61(5), 1109-1146.

Review of Economic Studies (1999), Special Symposium Issue on Contracts, January 1999, 1–217.

Segal, I. (1999), “A Theory of Incomplete Contracts,” *Review of Economic Studies*, 66, 57–82.

Williamson, O. (1967), “Hierarchical Control and Optimum Firm Size,” *Journal of Political Economy*,75(2), 123–138.

van Zandt, T. (1996), “Decentralized Information Processing in the Theory of Organizations,” in *Contemporary Economic Development Reviewed, Volume 4: The Enterprise and its Environment*, Murat Sertel (Ed.), London: Macmillan Press.

——— (1999), “Real-Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents,” *Review of Economic Studies*, 66,(3), 633-58.

———- (2003a), ”Real-time Hierarchical Resource Allocation,” INSEAD.

———- (2003b), ”Real-time Hierarchical Resource Allocation with Quadratic Costs,” INSEAD.

van Zandt, T. and R. Radner (2001), ”Real-time decentralized information processing and returns to scale, *Economic Theory*, 17, 497-544.

Appendix: Extension to Noisy Communication

The key step in the extension is the following Lemma. This Lemma establishes that *conditional on the messages received by j from i* which **is** known by i and is verifiable, messages actually sent by i do not affect the posterior beliefs subsequently held by i concerning the type of j. The reason is that j's communication plan selects messages conditioned on messages received earlier by j. Knowledge of the latter allows i to update her beliefs about the type of agent j: the messages actually sent by i are redundant for this purpose. In other words, the verifiable history h_i^V is a sufficient statistic with respect to the non-verifiable history h_i^N for i's posterior beliefs over θ_j and η.
Recall that communication is structured into rounds $k = 1, 2, \ldots, K$ with $t_0 = 0, t_K = T$. Consequently the history observed by i at the end of the communication phase can equivalently be denoted by h_i or h_{iK}.

Lemma 2 *Let $\eta_{(i,\cdot)}$ denote errors in messages sent by i, and $\eta_{(-i,\cdot)}$ the errors in messages sent by all others.*

(i) Choose arbitrary c_i and h_i^V with the property that

$$\{(c_j, c_P, \eta) \mid h_i^V(c_i, c_j, c_P, \eta) = h_i^V\} \neq \phi.$$

Then $h_i^N(c_i, c_j, c_P, \eta)$ is constant on

$$\{(c_j, c_P, \eta) \mid h_i^V(c_i, c_j, c_P, \eta) = h_i^V\}.$$

(ii) Choose arbitrary $(c_j, c_P, \eta_{(-i,\cdot)})$ and h_i^V so that

$$\{(c_i, \eta_{(i,\cdot)}) \mid h_i^V(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)}) = h_i^V\} \neq \phi.$$

Then

$$h_j(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)})$$
$$h_P(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)})$$

and

$$h^V(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)})$$

are constant on $\{(c_i, \eta_{(i,\cdot)}) \mid h_i^V(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)}) = h_i^V\}$.

(iii) Define $\Omega(h_i^V)$ *by*

$$\Omega(h_i^V) \equiv \{(c_i, \eta_{(i,\cdot)}) \mid \exists (c_j, c_P, \eta_{(-i,\cdot)}) \text{ s.t. } h_i^V(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h_i^V\}.$$

Then

$$\{(c_j, c_P, \eta_{(-i,\cdot)}) \mid h_i^V(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h_i^V\}$$

is independent of $(c_i, \eta_{(i,\cdot)})$ *on* $\Omega(h_i^V)$.

Proof of (i)

With $m_{ijk} = m_{ijk}(c_i, h_{i,k-1})$, h_i^N is represented by a function of c_i and $h_i = (h_{iK}^N, h_{iK}^V)$, which means that $h_{iK}^N(c_i, c_j, c_P, \eta)$ is constant, taking c_i and h_{iK}^V as given.

Proof of (ii)

We prove the statement by induction. Choose arbitrary $(c_j, c_P, \eta_{(-i,\cdot)})$ and h_{iK}^V so that

$$\{(c_i, \eta_{(i,\cdot)}) \mid h_{iK}^V(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)}) = h_{iK}^V\} \neq \phi$$

and

$$(c_i, \eta_{(i,\cdot)}), (c_i', \eta_{(i,\cdot)}') \in \{(c_i, \eta_{(i,\cdot)}) \mid h_{iK}^V(c_i, c_j, c_P, \eta_{(-i,\cdot)}, \eta_{(i,\cdot)}) = h_{iK}^V\}$$

so that $(c_i, \eta_{(i,\cdot)}) \neq (c_i', \eta_{(i,\cdot)}')$.

Step 1

At $k = 1$,

$$\begin{aligned} & h_{j1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \ & \{\{m_{ja1}(c_j), \phi(m_{ja1}(c_j), \eta_{ja1})\}_{a \neq j}, \phi(m_{ij1}(c_i), \eta_{ij1}), \phi(m_{Pj1}(c_P), \eta_{Pj1})\} \end{aligned}$$

and

$$\begin{aligned} & h_{j1}(c_i', c_j, c_P, \eta_{(i,\cdot)}', \eta_{(-i,\cdot)}) \\ = \ & \{\{m_{ja1}(c_j), \phi(m_{ja1}(c_j), \eta_{ja1})\}_{a \neq j}, \phi(m_{ij1}(c_i'), \eta_{ij1}'), \phi(m_{Pj1}(c_P), \eta_{Pj1})\}. \end{aligned}$$

Since $h^V_{i1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}$,

$$\phi(m_{ij1}(c_i), \eta_{ij1}) = \phi(m_{ij1}(c^{'}_i), \eta^{'}_{ij1}),$$

implying $h_{j1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h_{j1}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)})$. With a similar procedure, we can show that

$$h_{P1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h_{P1}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

Similarly,

$$\begin{aligned} & h^V_1(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \ & \{\{\phi(m_{ia1}(c_i), \eta_{ia1})\}_{a\neq i}, \{\tilde{m}_{ja1}\}_{a\neq j}, \{\tilde{m}_{Pa1}\}_{a\neq P}\} \end{aligned}$$

and

$$\begin{aligned} & h^V_1(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \ & \{\{\phi(m_{ia1}(c^{'}_i), \eta^{'}_{ia1})\}_{a\neq i}, \{\tilde{m}_{ja1}\}_{a\neq j}, \{\tilde{m}_{Pa1}\}_{a\neq P}\}. \end{aligned}$$

With $\{\phi(m_{ia1}(c_i), \eta_{ia1})\}_{a\neq i} = \{\phi(m_{ia1}(c^{'}_i), \eta^{'}_{ia1})\}_{a\neq i}$ implied by $h^V_{i1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}$,

$$h^V_1(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_1(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

Step 2

Define $h_{(a,k-1)}$ and $h^{'}_{(a,k-1)}$ for $a \in \{i, j, P\}$ by

$$h_{(a,k-1)} \equiv h_{(a,k-1)}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)})$$

and

$$h^{'}_{(a,k-1)} \equiv h_{(a,k-1)}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

Suppose that

$$h_{(j,k-1)} = h^{'}_{(j,k-1)},$$
$$h_{(P,k-1)} = h^{'}_{(P,k-1)}$$

and

$$h^V_{k-1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{k-1}(c^{'}_i, c_j, c_P, \eta^{'}_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

These imply $m_{jak}(c_j, h_{(j,k-1)}) = m_{jak}(c_j, h'_{(j,k-1)})$ and $m_{Pak}(c_P, h_{(P,k-1)}) = m_{Pak}(c_P, h'_{(P,k-1)})$. With

$$h^V_{ik}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{ik}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}),$$

$$\phi(m_{ia}(c_i, h_{(i,k-1)}), \eta_{iak}) = \phi(m_{ia}(c'_i, h'_{(i,k-1)}), \eta'_{iak}).$$

Then it is easy to check that

$$h_{ak}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h_{ak}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

for $a \in \{j, P\}$ and

$$h^V_k(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_k(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}).$$

Proof of (iii)

Choose arbitrary h^V_{iK} and $(c_i, \eta_{(i,\cdot)}), (c'_i, \eta'_{(i,\cdot)}) \in \Omega(h^V_{iK})$ so that $(c_i, \eta_{(i,\cdot)}) \neq (c'_i, \eta'_{(i,\cdot)})$. We show that $h^V_{iK}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{iK}$ implies $h^V_{iK}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(i,\cdot)}) = h^V_{iK}$. By the definition of $\Omega(h^V_{iK})$, there exists $(c'_j, c'_P, \eta'_{(-i,\cdot)})$ so that $h^V_{iK}(c'_i, c'_j, c'_P, \eta'_{(i,\cdot)}, \eta'_{(-i,\cdot)}) = h^V_i$.

Step 1

Our proof follows induction. At $k = 1$,

$$\begin{aligned} & h^V_{i1}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \;& (\{\phi(m_{ia1}(c'_i), \eta'_{ia1})\}_{a \neq i}, \phi(m_{ji1}(c_j), \eta_{ji1}), \phi(m_{Pi1}(c_P), \eta_{Pi1})). \end{aligned}$$

Since $h^V_{i1}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}(c'_i, c'_j, c'_P, \eta'_{(i,\cdot)}, \eta'_{(-i,\cdot)})$, $\phi(m_{ia1}(c'_i), \eta'_{ia1}) = \phi(m_{ia1}(c_i), \eta_{ia1})$ for any $a \neq i$. This means that $h^V_{i1}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{i1}$.

Step 2

For $a \in \{i, j, P\}$, define

$$h_{(a,k-1)} \equiv h_{(a,k-1)}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)})$$

$$h'_{(a,k-1)} \equiv h_{(a,k-1)}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)})$$
$$h''_{(a,k-1)} \equiv h_{(a,k-1)}(c'_i, c'_j, c'_P, \eta'_{(i,\cdot)}, \eta'_{(-i,\cdot)})$$

Suppose that $h^V_{(i,k-1)}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{(i,k-1)}$. Then

$$\begin{aligned} & h^V_{(i,k)}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \; & \{h^V_{(i,k-1)}, \phi(m_{iak}(c'_i, h'_{(i,k-1)}), \eta'_{iak}), \phi(m_{aik}(c_a, h'_{(a,k-1)}), \eta_{aik})\}_{a\neq i}. \end{aligned}$$

On the other hand,

$$\begin{aligned} & h^V_{(i,k)} = h^V_{(i,k)}(c_i, c_j, c_P, \eta_{(i,\cdot)}, \eta_{(-i,\cdot)}) \\ = \; & \{h^V_{(i,k-1)}, \phi(m_{iak}(c_i, h_{(i,k-1)}), \eta_{iak}), \phi(m_{aik}(c_a, h_{(a,k-1)}), \eta_{aik})\}_{a\neq i} \end{aligned}$$

and

$$\begin{aligned} & h^V_{(i,k)} = h^V_{(i,k)}(c'_i, c'_j, c'_P, \eta'_{(i,\cdot)}, \eta'_{(-i,\cdot)}) \\ = \; & \{h^V_{(i,k-1)}, \phi(m_{iak}(c'_i, h''_{(i,k-1)}), \eta'_{iak}), \phi(m_{aik}(c'_a, h''_{(a,k-1)}), \eta'_{aik})\}_{a\neq i}. \end{aligned}$$

This means that

$$\phi(m_{iak}(c_i, h_{(i,k-1)}), \eta_{iak}) = \phi(m_{iak}(c'_i, h''_{(i,k-1)}), \eta'_{iak})$$

and

$$\phi(m_{aik}(c_a, h_{(a,k-1)}), \eta_{aik}) = \phi(m_{aik}(c'_a, h''_{(a,k-1)}), \eta'_{aik}).$$

¿From part (ii) of the Lemma, for $a \neq i$, $h_{(a,k-1)} = h'_{(a,k-1)}$, taking $(c_j, c_P, \eta_{(-i,\cdot)})$ and $h^V_{(i,k-1)}$ as given, implying

$$\phi(m_{aik}(c_a, h_{(a,k-1)}), \eta_{aik}) = \phi(m_{aik}(c_a, h'_{(a,k-1)}), \eta_{aik}).$$

¿From part (i) of the Lemma, taking c'_i and $h^V_{(i,k-1)}$ as given, $h'_{(i,k-1)} = h''_{(i,k-1)}$, implying

$$\phi(m_{iak}(c_i, h_{(i,k-1)}), \eta_{iak}) = \phi(m_{iak}(c'_i, h''_{(i,k-1)}), \eta'_{iak}) = \phi(m_{iak}(c'_i, h'_{(i,k-1)}), \eta'_{iak})$$

It follows that

$$h^V_{(i,k)}(c'_i, c_j, c_P, \eta'_{(i,\cdot)}, \eta_{(-i,\cdot)}) = h^V_{(i,k)}.$$

This Lemma allows us to characterize the allocations attainable under centralized contracting in terms of incentives facing each agent separately. The expected transfer that agent i will receive upon selecting some production level q_i in any mechanism can be expressed as a function of the verifiable messages h_{iK}^V alone, since i can form her beliefs about agent j's type based on the latter. Hence the following extension of Proposition 2 obtains.

Proposition 7 *When communication is organized in a finite number of rounds $k = 1, \ldots, K$ and is subject to random errors represented by the error vector $\eta = \{\eta_{ijk}\}_{j \neq i, k=1,\ldots,K}$ with a given (exogenous) distribution, a production allocation $q \equiv \{q_i(\theta_i, \theta_j, \eta)\}_{i=1,2}$ is attainable under centralized contracting if and only if there exists a communication protocol $p \in \mathcal{P}_C$ and associated functions $\hat{q}_i(\theta_i, h_{iK}^V) : \Theta_i \times \mathcal{H}_i \to \Re_+$, $c_i(\theta_i) : \Theta_i \to \mathcal{C}_i$ and $U_i(h_{iK}^V) : \mathcal{H}_i \to \Re_+$ such that for $i = 1, 2$:*

(i) $q_i(\theta_i, \theta_j, \eta) = \hat{q}_i(\theta_i, h_{iK}^V(\tilde{c}(\theta), \eta))$

(ii) $\hat{q}_i(\theta_i, h_{iK}^V)$ *is non-increasing in* θ_i*, for every* h_{iK}^V

(iii)

$$c_i(\theta_i) = \arg\max E_{\theta_j,\eta}[U_i(h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta)) + \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(x, h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta))dx]$$

and

$$E_{\theta_j,\eta}[U_i(h_{iK}^V(c_i(\bar{\theta}_i), c_j(\theta_j), c_P, \eta))] \geq 0.$$

The resulting ex-ante payoff for P is

$$\begin{aligned} & E[V(\hat{q}_i(\theta_i, h_{iK}^V(\tilde{c}(\theta), \eta)), \hat{q}_j(\theta_j, h_{iK}^V(\tilde{c}(\theta), \eta))) \\ - & \ v_i(\theta_i)\hat{q}_i(\theta_i, h_{iK}^V(\tilde{c}(\theta), \eta)) - v_j(\theta_j)\hat{q}_j(\theta_j, h_{iK}^V(\tilde{c}(\theta), \eta)) \\ - & \ U_i(h_{iK}^V(c_i(\bar{\theta}_i), c_j(\theta_j), c_P, \eta)) - U_j(h_{jK}^V(c_j(\bar{\theta}_j), c_i(\theta_i), c_P, \eta))]. \end{aligned}$$

Step 1

If an allocation $(t, q, c) \equiv \{t_i(\theta_i, \theta_j, \eta), q_i(\theta_i, \theta_j, \eta), \tilde{c}(\theta)\}$ is implementable, there exist functions $\hat{t}_i(q_i, q_j, h_K^V)$ such that

(a) $t_i(\theta_i, \theta_j, \eta) = \hat{t}_i(q_i(\theta_i, \theta_j, \eta), q_j(\theta_i, \theta_j, \eta), h_K^V(\tilde{c}(\theta), \eta))$

(b) $q_i(\theta_i, \theta_j, \eta) = \hat{q}_i(\theta_i, c_i(\theta_i), h_{iK}(\tilde{c}(\theta_i, \theta_j), \eta))$.

(c) $\hat{q}_i(\theta_i, c_i, h_{iK})$ maximizes

$$\begin{aligned}&E_{\theta_j,\eta}[\hat{t}_i(q_i, \hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i, c_j(\theta_j), c_P, \eta)), h_K^V(c_i, c_j(\theta_j), c_P, \eta))\\&\mid h_{iK}(c_i, c_j(\theta_j), c_P, \eta) = h_{iK}, c_i] - \theta_i q_i\end{aligned}$$

(d) $c_i(\theta_i)$ maximizes

$$\begin{aligned}&E_{\theta_j,\eta}[\hat{t}_i(\hat{q}_i(\theta_i, c_i, h_{iK}(c_i, c_j(\theta_j), c_P, \eta)), \hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i, c_j(\theta_j), c_P, \eta)),\\&h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta)) - \theta_i \hat{q}_i(\theta_i, c_i, h_{iK}(c_i, c_j(\theta_j), c_P, \eta))]\end{aligned}$$

(e) Participation constraint:

$$\begin{aligned}&E_{\theta_j,\eta}[\hat{t}_i(\hat{q}_i(\theta_i, c_i(\theta_i), h_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta)),\\&\hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta)),\\&h_K^V(c_i(\theta_i), c_j(\theta_j), c_P, \eta)) - \theta_i \hat{q}_i(\theta_i, c_i(\theta_i), h_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta))] \geq 0\end{aligned}$$

Step 2:Proof of Necessity

¿From part (i) of the Lemma, since h_{iK}^N is constant taking c_i and h_{iK}^V as given,

$$\begin{aligned}&E_{\theta_j,\eta}[\hat{t}_i(q_i, \hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i, c_j(\theta_j), c_P, \eta)), h_K^V(c_i, c_j(\theta_j), c_P, \eta))\\&\mid h_{iK}(c_i, c_j(\theta_j), c_P, \eta) = h_{iK}, c_i]\\=\;&E_{\theta_j,\eta}[\hat{t}_i(q_i, \hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i, c_j(\theta_j), c_P, \eta)), h_K^V(c_i, c_j(\theta_j), c_P, \eta))\\&\mid h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta) = h_{iK}^V, c_i]\end{aligned}$$

¿From part (ii) of the Lemma, $h_{jK}(c_i, c_j(\theta_j), c_P, \eta)$ and $h_K^V(c_i, c_j(\theta_j), c_P, \eta)$ are constant taking $(c_j(\theta_j), c_P, \eta_{(-i,\cdot)})$ and h_{iK}^V as given. From part (iii) of the Lemma,

$$\{(\theta_j, \eta_{(-i,\cdot)}) \mid h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta) = h_{iK}^V\}$$

is independent of $(c_i, \eta_{(i,\cdot)})$. This implies that

$$\begin{aligned}&E_{\theta_j,\eta}[\hat{t}_i(q_i, \hat{q}_j(\theta_j, c_j(\theta_j), h_{jK}(c_i, c_j(\theta_j), c_P, \eta)), h^{VK}(c_i, c_j(\theta_j), c_P, \eta)\\&\mid h^V_{iK}(c_i, c_j(\theta_j), c_P, \eta) = h^V_{iK}, c_i]\end{aligned}$$

can be represented as a function of (q_i, h^V_{iK}). Let this function be denoted by $T_i(q_i, h^V_{iK})$. Then $\hat{q}_i(\theta_i, h^V_{iK})$ must maximize $T_i(q_i, h^V_{iK}) - \theta_i q_i$, and therefore must be non-increasing in θ_i. This implies (i) and (ii). From (d) and (e),

$$\begin{aligned}&c_i(\theta_i) = \arg\max_{c_i} E_{\theta_j,\eta}[T_i(\hat{q}_i(\theta_i, h^V_{iK}(c_i, c_j(\theta_j), c_P, \eta)), h^V_{iK}(c_i, c_j(\theta_j), c_P, \eta))\\ - \;&\theta_i \hat{q}_i(\theta_i, h^V_{iK}(c_i, c_j(\theta_j), c_P, \eta))]\end{aligned}$$

$$\begin{aligned}&E_{\theta_j,\eta}[T_i(\hat{q}_i(\theta_i, h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta)), h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta))\\ - \;&\theta_i \hat{q}_i(\theta_i, h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta))] \geq 0\end{aligned}$$

It is easy to show that these imply (iii).

Step 3: P's payoff

From step 2, P's ex-ante payoff which implements $(q_i(\theta_i, \theta_j, \eta), \tilde{c}(\theta))$ is

$$\begin{aligned}&E_{(\theta_i,\eta)}[V(\hat{q}_i(\theta_i, h^V_{iK}(\tilde{c}(\theta), \eta)), \hat{q}_j(\theta_j, h^V_{jK}(\tilde{c}(\theta), \eta)))\\ - \;&T_i(\hat{q}_i(\theta_i, h^V_{iK}(\tilde{c}(\theta), \eta)), h^V_{iK}(\tilde{c}(\theta), \eta)) - T_j(\hat{q}_j(\theta_j, h^V_{jK}(\tilde{c}(\theta), \eta)), h^V_{jK}(\tilde{c}(\theta), \eta))]\end{aligned}$$

From step 2,

$$\begin{aligned}&E_{\theta_j,\eta}[T_i(\hat{q}_i(\theta_i, h^V_{iK}(\tilde{c}(\theta), \eta)), h^V_{iK}(\tilde{c}(\theta), \eta)) - \theta_i \hat{q}_i(\theta_i, h^V_{iK}(\tilde{c}(\theta), \eta))]\\ \geq \;&E_{\theta_j,\eta}[T_i(\hat{q}_i(\theta_i', h^V_{iK}(\tilde{c}(\theta_i', \theta_j), \eta)), h^V_{iK}(\tilde{c}(\theta_i', \theta_j), \eta)) - \theta_i \hat{q}_i(\theta_i', h^V_{iK}(\tilde{c}(\theta_i', \theta_j), \eta))]\end{aligned}$$

for any θ_i'. This implies that

$$\begin{aligned}&E_{\theta_j,\eta}[T_i(\hat{q}_i(\theta_i, h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta)), h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta))]\\ = \;&E_{\theta_j,\eta}[\theta_i \hat{q}_i(\theta_i, h^V_{iK}(c_i(\theta_i), c_j(\theta_j), c_P, \eta)) + \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(t, h^V_{iK}(c_i(t), c_j(\theta_j), c_P, \eta))dt]\\ + \;&U_i(h^V_i(c_i(\bar{\theta}_i), c_j(\theta_j), c_P, \eta))\end{aligned}$$

This generates the expression for the *ex ante* payoff of P stated in the proposition.

Step 4: Proof of Sufficiency

For arbitrary $\hat{q}_i(\theta_i, h_{iK}^V)$ which is non-increasing in θ_i, define $\theta_i(q_i, h_{iK}^V)$ by

$$\theta_i(q_i, h_{iK}^V) \equiv \sup_{\theta_i}\{\hat{q}_i(\theta_i, h_{iK}^V) \geq q_i\}$$

Also define the following transfer functions stipulated in the centralized contract:

$$t_i(q_i, h_{iK}^V) = \theta_i(q_i, h_{iK}^V)q_i + \int_{\theta_i(q_i, h_{iK}^V)}^{\bar{\theta}_i} \hat{q}_i(t, h_{iK}^V)dt + U_i(h_{iK}^V).$$

By construction, then, the following is true for any (θ_i, h_{iK}^V):

$$t_i(\hat{q}_i(\theta_i, h_{iK}^V), h_{iK}^V) - \theta_i\hat{q}_i(\theta_i, h_{iK}^V) = \int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(t, h_{iK}^V)dt + U_i(h_{iK}^V).$$

This implies that $t_i(q_i, h_{iK}^V) - \theta_i q_i$ is maximized at $\hat{q}_i(\theta_i, h_{iK}^V)$. Then taking the strategy of the other agent $\hat{q}_j(\theta_j, h_{jK}^V)$ and $c_j(\theta_j)$ as given, agent i's interim payoff as a function of a given communication plan c_i is

$$E_{(\theta_j,\eta)}[\int_{\theta_i}^{\bar{\theta}_i} \hat{q}_i(t, h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta))dt + U_i(h_{iK}^V(c_i, c_j(\theta_j), c_P, \eta))].$$

By (iii), this is maximized at $c_i(\theta_i)$. The rest of the argument is the same as used for Proposition 2.