Delegation of Decision-Making in Organizations

Margaret A. Meyer
Nuffield College and Department of Economics
Oxford University

2018
What determines the degree to which decision-making is centralized (concentrated in the hands of management) vs. delegated (to subordinates) in organizations?

Degree of delegation (decentralization) of decision-making has varied over time and varies across organizations and environments:

- Over 20th century, large corporations evolved (Chandler, 62; Williamson, 75) from U-form (“unitary”), in which employees grouped according to function, to M-form (“multi-divisional”), in which employees grouped according to products or processes.
  - With switch to M-form, more delegation to division mgrs; top mgt’s role: reallocation of cash flows via “internal capital market”.

- More recent increases in decent.: greater use of “profit/cost centers” within divisions; greater use of teams, coordinated laterally rather than vertically
  - Cross-sectional evidence in Acemoglu et al, 07: decentralization more likely for firms that are closer to the technological frontier, in more heterogeneous environments, and younger.
Why and when is delegation of decision-making valuable?

- Standard argument in management literature: Decision rights should be delegated to those managers who possess the relevant information.

- But what is the alternative to delegation?
  - Is it centralized decision-making by top management based only on top management's information, without any attempt to communicate with better-informed subordinates? If so, fairly easy to deduce why/when delegation would be valuable.
  - However, we should compare “delegation” with “centralization accompanied by communication”. 
“Delegation” vs. “centralization with communication”

Start by adopting the perspective of mechanism design:

If

- no limitations or costs of communication;
- no limitations or costs of processing of information;
- top management can commit to actions and payments in response to all possible reports by subordinates;

then

Generalized Revelation Principle (Myerson, 82): Any non-cooperative equilibrium outcome of an arbitrary organizational structure can be replicated by a centralized two-tier structure, in which each employee is given incentives to communicate truthfully all of his private information directly to top management.
Arbitrary Organizational Structure

Centralized Two-Tier Structure
“Delegation” vs. “centralization with communication”

So for delegation of decision-making to dominate centralization with communication, it must be that one or more of the following conditions apply:

- communication is costly (Melumad et al, JAE, 92, Rand, 95; Rotemberg, JEMS, 99; Mookherjee and Tsumagari, JPE, 14)

- processing of information is costly, so without delegation, top management faces “information overload” (Geanakoplos + Milgrom, JJIE, 91; McAfee + McMillan, JEMS, 95)

- management’s ability to write complete contracts is in some way limited (Aghion + Tirole, JPE, 97; Dessein, RESud, 02; Alonso et al, AER, 08)
Focus on two different analyses of “centralization with communication” vs. “delegation”

1. Centralization vs. delegation: the effect of communication costs
   • based on Melumad, Mookherjee, and Reichelstein (J. of Accounting and Econ., 1992; Rand, 1995) and Mookherjee and Tsumagari (JPE, 2014)

2. Centralization vs. delegation: the effect of limited commitment
   • based on Alonso, Dessein, and Matouschek (AER, 2008)
Centralization vs. delegation to a cost center: The effect of communication costs (Melumad, Mookherjee, and Reichelstein, 92, 95):

- Assume no limitations on P’s ability to commit to a mechanism
- $P$, $A_1$, $A_2$ risk-neutral
- $A_i$ produces output $a_i$ at cost $a_i\theta_i$; receives payment $x_i$
- $\theta_i$ privately observed by $A_i$ before contracting; $a_i\theta_i$ also private
- $\theta_1$, $\theta_2$ independent, continuously distributed; distribs. are common knowledge
- $A_1$ and $A_2$ have common, exogenous reservation utility of 0
- $P$ has benefit function $B(a_1, a_2)$; $P$ chooses a mechanism to minimize the total expected payment he must make to $A_1$ and $A_2$ to generate a benefit of at least $\bar{B}$, i.e. to induce an output pair $(a_1, a_2)$ such that $B(a_1, a_2) \geq \bar{B}$. 
Centralization vs. delegation to a cost center

Centralization:

1. P commits to a mechanism specifying production levels $a_1, a_2$ and transfers $x_1, x_2$ for agents, as functions of agents’ reports $\hat{\theta}_1, \hat{\theta}_2$ to $P$ about their costs $\theta_1, \theta_2$.

2. $A_1$ and $A_2$ simultaneously choose reports $\hat{\theta}_1$ and $\hat{\theta}_2$ (respectively) to $P$.
   - Given $P$’s choice of mechanism, the reporting strategies of the agents must constitute a Bayesian Nash Eqm.

3. Production occurs and payments are made.
Centralization vs. delegation to a cost center

Delegation: P makes $A_1$ a “cost center”, in charge of sub-contracting with $A_2$

1. P offers $A_1$ a menu of compensation contracts.
   - Each contract corresponds to a different report by $A_1$ about $\theta_1$.
   - Each contract specifies how P’s payment $x_1$ to $A_1$ will depend on the payment $x_2$ that $A_1$ makes in sub-contracting with $A_2$.

2. $A_1$ sends a report $\hat{\theta}_1$ about $\theta_1$ to P, thus “choosing” one contract from the menu. P is committed to honor the contract $A_1$ chooses.

3. $A_1$, knowing his own cost $\theta_1$, sub-contracts with $A_2$.
   - $A_1$ offers $A_2$ a menu specifying how the production allocation and the payment from $A_1$ to $A_2$ will depend on $A_2$’s report $\hat{\theta}_2$ about his cost $\theta_2$.

4. $A_2$ chooses report $\hat{\theta}_2$.

5. Production occurs and payments are made.
Benchmark with no limitations on communication

With no limitations on communication, Myerson’s Generalized Revelation Principle implies that anything achievable with a cost center can be replicated under centralization.

But is there a loss from delegation to a cost center, compared to centralization? No.

**Proposition:** *With no limitations on communication, cost center and centralization perform equally well: the outcome under the optimal centralized mechanism can be replicated under an optimally designed cost center.*
Benchmark with no limitations on communication

Example: \( B(a_1, a_2) = a_1 + a_2; \quad \theta_1, \theta_2 \text{ i.i.d.} \)

Centralization: Optimal mechanism specifies that \( A_i \) with lower reported \( \theta_i \) produces \( a_i = \bar{B} \), while other agent produces nothing.

- \( P \) pays information rents to both agents to ensure truthful reporting is a BNE.

Delegation to a cost center: The potential problem is that \( A_1 \), in sub-contracting with \( A_2 \), will bias the production allocation towards himself because of the need to pay rents to \( A_2 \) to elicit \( A_2 \)'s information.

- But because \( P \) can observe payment \( x_2 \) from \( A_1 \) to \( A_2 \), \( P \) can eliminate this bias, by designing the menu of compensation contracts for \( A_1 \) so as to subsidize \( A_1 \)'s purchases of output from \( A_2 \).

- As a result, \( P \) can induce \( A_1 \) to choose same allocation as results under centralization (\( A_i \) with lower \( \theta_i \) produces \( \bar{B} \)), and \( P \) incurs same cost.
Allocation of production under both optimal centralization and optimal cost center when there are no limitations on communication
Centralization vs. delegation given limitations on communication

Modeling of limited communication: Assume agents can transmit only a coarse summary of private infor—though $\theta_i$ is a continuous variable, each agent can make only a finite number of possible reports. Here, assume only 2 possible reports.

- Note that $A_1$ is not assumed to be better than $P$ at understanding $A_2$’s infor.

- Delegation now brings a flexibility gain, since under delegation to a cost center, $A_1$ chooses subcontract with $A_2$ based on precise value of his own cost $\theta_1$, not on a coarse summary of it.

Proposition: With limited communication, delegation to a cost center strictly dominates centralization.
Centralization vs. delegation given limitations on communication

**Proposition:** With limited communication, delegation to a cost center strictly dominates centralization.

**Centralization:** Coarseness of communication makes it optimal for $P$ to communicate with agents **sequentially**, not simultaneously.

- The optimal way to ask $A_2$ to partition the values of $\theta_2$ depends on which of the two reports $A_1$ made about $\theta_1$.

**Delegation to a cost center:** The inefficiency in production decisions resulting from coarseness of communication is less than under centralization, because of the **flexibility gain**.
In striped areas, production is allocated inefficiently.
Implications of Melumad et al's analyses

- Delegation to a cost center is more efficient than centralized decision-making when two conditions hold:
  1. The cost center’s financial performance (here, $A_1$’s payment to $A_2$) can be monitored and contracted upon.
  2. The private information of the manager of the cost center (here, $A_1$’s cost type $\theta_1$) is difficult to fully transmit to top management.

- If under centralization, $P$ bases decisions only on public information (e.g., Acemoglu et al, 07), then of course attractiveness of delegation increases.

- For firms that are closer to the technological frontier or operating in more heterogeneous environments, private information of middle managers likely to be more substantial and complex $\Rightarrow$ decentralization accompanied by financial monitoring relatively more attractive.
Centralization vs. delegation: The effect of limited commitment

If top management’s ability to commit to how will respond to reports by subordinates is limited, then the Generalized Revelation Principle does not hold.

Many recent papers assume that although top management cannot commit to rules specifying actions and payments as functions of reports, it can commit to delegate decision rights. Then ask: When is such “formal delegation” (also called “contractual delegation”) of decision rights advantageous? Answers provided:

- Aghion and Tirole (JPE, 97): when it is crucial to provide subordinates with incentives to generate ideas and/or acquire information.

- Dessein (REStud, 02): when subordinate has private infor. about the efficient decision and principal cannot commit to a mechanism to elicit the infor.

- Alonso et al (AER, 08): in an envir. like Dessein (02) but with 2 agents (division mgrs) each privately informed about conditions in own division, as long as i) division mgrs not too biased towards their own division’s performance or ii) coordination of decisions across divisions is not too important.
Centralization vs. delegation when only decision rights are contractible (Alonso, Dessein, and Matouschek, AER, 08):

- Players: HQ and managers, $M_1$ and $M_2$, of ex ante symmetric divisions; all players maximize expected payoffs

- Decisions: production decisions $d_1$ and $d_2$

- Profit of division $i$, $\pi_i = -(d_i - \theta_i)^2 - \delta(d_i - d_j)^2$, where $\theta_i$ is privately observed by $M_i$, and $(\theta_1, \theta_2)$ are i.i.d., $\sim U[-s, s]$
  
  - Parameter $\delta \in [0, \infty)$ is common knowledge and measures importance of coordination (of $d_1$ and $d_2$) relative to adaptation (of $d_i$ to $\theta_i$)

- $M_i$'s payoff is $\lambda \pi_i + (1 - \lambda) \pi_j$
  
  - Parameter $\lambda \in [\frac{1}{2}, 1]$ is common knowledge and measures bias of each manager towards own division
  
  - Bias could depend on managers’ incentive schemes (exogenous here)

- HQ's payoff is $\pi_1 + \pi_2$
Possible allocations of decision rights in the ADM (08) model

**Centralization (C):** Each $M_i$ simultaneously makes a cheap-talk report about $\theta_i$ to HQ, and then HQ chooses both $d_1$ and $d_2$.

**Decentralization (D):** Each $M_i$ simultaneously makes a cheap-talk report about $\theta_i$ to $M_j$, and then each $M_i$ chooses $d_i$.

HQ chooses between C and D to maximize expectation of

\[ \pi_1 + \pi_2 = -(d_1 - \theta_1)^2 - (d_2 - \theta_2)^2 - 2\delta(d_1 - d_2)^2 \]

- NB: parties can commit only to who makes each decision, not to mechanisms specifying decisions and transfers as functions of reports.

- **Centralization performs better w.r.t. coordination:** yields lower expected value of $(d_1 - d_2)^2$.

- **Decentralization performs better w.r.t. adaptation:** yields lower expected value of $(d_1 - \theta_1)^2 + (d_2 - \theta_2)^2$. 
Possible allocations of decision rights in the ADM (08) model

Centralization (C): Each $M_i$ simultaneously makes a cheap-talk report about $\theta_i$ to HQ, and then HQ chooses both $d_1$ and $d_2$.

Decentralization (D): Each $M_i$ simultaneously makes a cheap-talk report about $\theta_i$ to $M_j$, and then each $M_i$ chooses $d_i$.

HQ chooses between C and D to maximize expected value of $\pi_1 + \pi_2$

- Benefits of C from HQ’s perspective:
  1. decision-making is unbiased (with D, managers underweight coordination)
  2. equilibrium quality of communication is better (less coarse)

- Benefit of D from HQ’s perspective: flexibility gain since each $M_i$ bases choice of $d_i$ on precise value of own $\theta_i$

- Contrast Melumad et al model: There, in D, transfers to $A_1$ can eliminate that agent’s bias. And coarseness of communication is exogenous and equal in C and D. So only flexibility gain matters for the comparison btw. C and D.
Equilibrium communication under C and D

Under both C and D, each $M_i$ has incentive to exaggerate magnitude of $\theta_i$.

- **HQ**: $-(d_1 - \theta_1)^2 - (d_2 - \theta_2)^2 - 2\delta(d_1 - d_2)^2$
- **M1**: $-\lambda(d_1 - \theta_1)^2 - (1 - \lambda)(d_2 - \theta_2)^2 - \delta(d_1 - d_2)^2$
- **M2**: $-(1 - \lambda)(d_1 - \theta_1)^2 - \lambda(d_2 - \theta_2)^2 - \delta(d_1 - d_2)^2$

- Under C, from $M_i$’s perspective, HQ puts too much weight on coordination and too little on adaptation of $d_i$ to $\theta_i$ ($\frac{2\delta}{\lambda} \geq \frac{\delta}{1-\lambda}$), so $|d_i|$ is not sensitive enough to $|\theta_i|$. Hence, $M_i$ has incentive to exaggerate $|\theta_i|$ when communicating to HQ.

- Under D, from $M_i$’s perspective, $M_j$ puts too little weight on coordination and too much on adaptation of $d_j$ to $\theta_j$ ($\frac{\delta}{\lambda} \leq \frac{\delta}{1-\lambda}$), so $|d_j|$ is not sensitive enough to $|\theta_i|$. Hence, $M_i$ has incentive to exaggerate $|\theta_i|$ when communicating to $M_j$.

- **NB**: Since $\theta_1, \theta_2$ are independent, $E(\theta_j|\theta_i) = 0$ for all $\theta_i$. 

Equilibrium communication under C and D

Under both C and D, each $M_i$ has incentive to exaggerate magnitude of $\theta_i$.

- Stronger incentives to exaggerate $\implies$ coarser communication in eqm.
- All communication eqa are interval eqa, with size of intervals $\uparrow$ in $|\theta_i|$. 
  - If and only if $\theta_i = 0$, $M_i$ (sender) has no incentive to misreport.
  - Divergence in preferences btw. $M_i$ (sender) and $HQ/M_j$ (receiver) increases in $|\theta_i|$. Hence communication is coarser the larger is $|\theta_i|$.
  - Because, for $\theta_i = 0$, incentives of $M_i$ (sender) and $HQ/M_j$ (receiver) are perfectly aligned, there is no upper limit on number, $N$, messages that can arise in eqm. ADM focus on the most informative eqm ($N \to \infty$).

* NB: this eqm does not correspond to full revelation; its informativeness depends on $\lambda$ and $\delta$. 
Comparing equilibrium quality of communication under C vs D

Under both C and D, each $M_i$ has incentive to exaggerate magnitude of $\theta_i$. Stronger incentives to exaggerate $\Rightarrow$ coarser communication in eqm.

Quality of eqm communication is worse (coarser) under D than under C, because preferences of $M_1$ and $M_2$ (who communicate under D) are less aligned than are preferences of $M_i$ and HQ (who communicate under C).

- Under C, as $\delta \uparrow$, quality of eqm communication worsens: HQ makes $d_i$ less sensitive to beliefs about $|\theta_i|$, so $M_i$’s incentives to exaggerate $|\theta_i| \uparrow$.

- Under D, as $\delta \uparrow$, quality of eqm communication improves: $M_j$ makes $d_j$ more sensitive to beliefs about $|\theta_i|$, so $M_i$’s incentives to exaggerate $|\theta_i| \downarrow$.

- Eqm commun. becomes equally good under D and C as $\delta \to \infty$ (or $\lambda \to \frac{1}{2}$).
**HQ’s choice between centralization and decentralization**

**Proposition:** For HQ to prefer centralization, coordination must be sufficiently important (\(\delta\) sufficiently large) and division managers must be sufficiently biased towards their own divisions (\(\lambda\) sufficiently large).

- Surprisingly, if \(\lambda \leq \bar{\lambda}\), then even as \(\delta\) gets very large, \(C\) never dominates \(D\).
  - “Bias” in decision-making and coarser communication under \(D\) are outweighed by \(D\)’s flexibility gain.
If top mgt can delegate decision rights and also restrict the set of decisions from which subordinates may choose, what restrictions should it impose?

- Szalay (REStud, 05): Suppose it is crucial to provide A with incentives to acquire information. Then even if P's and A's preferences over decisions, for given information, are fully aligned, it can be desirable for P to force A to choose between extreme options.

- Armstrong and Vickers (Etrica, 10): Suppose attributes of projects are two-dimensional, P and A value different attributes, $\theta_p$ and $\theta_a$, resp., and only A can observe how many projects are available and their attributes.
  - Suppose P can commit to a rule specifying whether a project proposed by A will be accepted, as a function of $(\theta_p, \theta_a)$.  
  - Optimal rule specifies that the larger is $\theta_a$, the larger is the minimum acceptable level of $\theta_p$.  

Is “formal delegation” of decision-making authority feasible/credible?

- Milgrom and Roberts (90, 92) and Meyer, Milgrom, and Roberts (JEMS, 92) argue that top management always has discretionary authority to intervene in the activities of subunits or subordinates.
  - Subordinates know this and will engage in “influence activities” to alter top management’s decisions.
  - A *credible* commitment not to intervene may require making a subunit a legally independent entity.

- Similarly, Baker, Gibbons, and Murphy (JLEO, 99) argue: “decision rights in organizations are not contractible; the boss can always overturn a subordinate’s decision, so formal authority resides only at the top.”
  - They ask: Can a promise to delegate authority can be *credible*? And if so, under what conditions?
**BGM (JLEO, 99): Can self-enforcing contracts support informal delegation of decision rights by a boss (P) to a subordinate (A)?**

- In each period of $\infty$-horizon relationship, A privately chooses effort on finding a good project. A’s and P’s preferences over projects are not perfectly aligned.

- Each period, A can propose to P a project he has found. P can observe whether or not the project is in the firm’s interest and then decide whether to accept A’s recommendation.

- P’s short-run incentive is to reject projects not in the firm’s interest, but if he is expected to do this, A’s incentives to find projects will be weaker.

- When will eqm behavior involve P “rubber stamping” A’s proposals, i.e. when will A possess “informal authority” in eqm?

- Key conclusions of BGM (99): If future value of relationship is large enough, subordinate can possess informal authority. But even if informal delegation of authority would be efficient, it is not necessarily achievable, because boss may have incentives to renege on the informal agreement.