Delegation of Decision-Making in Organizations

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

2020

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.
  - Switch to M-form was accompanied by more delegation to division mgrs; top mgt's role became focused on reallocation of cash flows via "internal capital market".

- More recent increases in decentralization: greater use of "profit/cost centers" w/i 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:

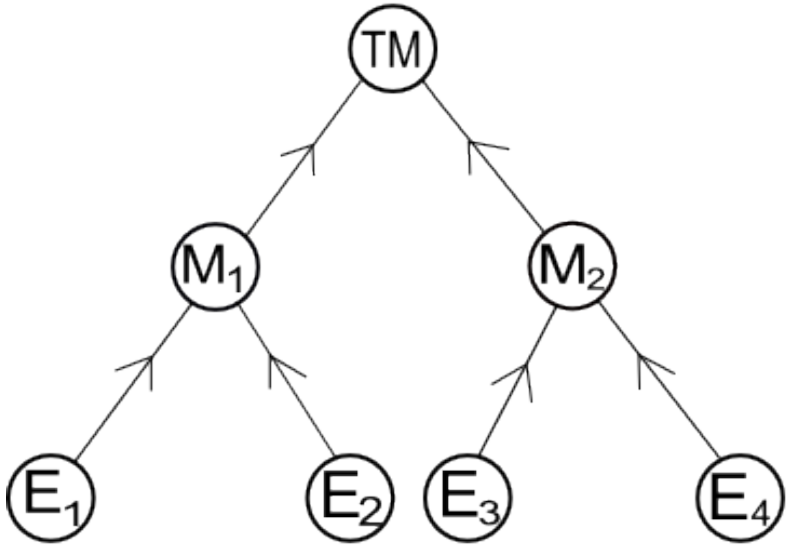
If

- 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;

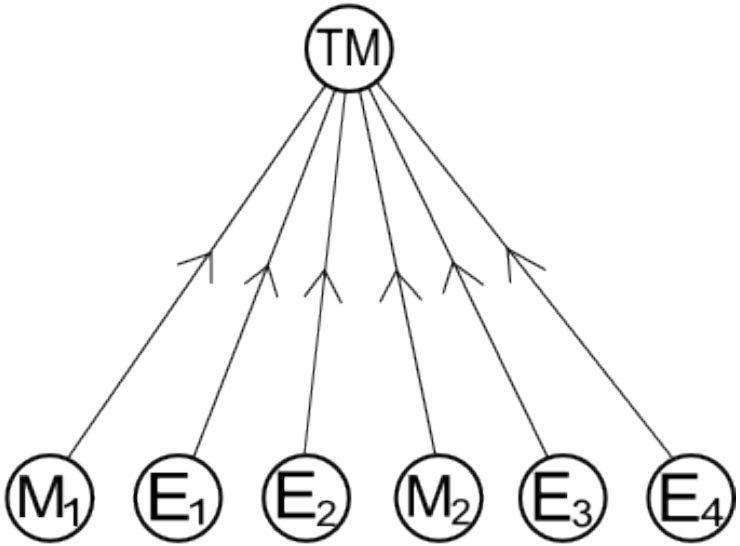
then

Generalized Revelation Principle (Myerson, *J. Math. Econ.*, 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, *REStud*, 02; Alonso et al, *AER*, 08; Rantakari, *REStud*, 08)

This lecture will focus on two different analyses of "centralization with comunication" vs. "delegation"

1. Centralization vs. delegation: the effect of communication costs
    - based on Melumad, Mookherjee, and Reichelstein (*J. of Accting 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)
    - Rantakari (*REStud, 2008*) analyzes a closely related model

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
- θ_i privately observed by A_i before contracting; $a_i\theta_i$ also private, e.g. because may include effort costs
- θ_1, θ_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

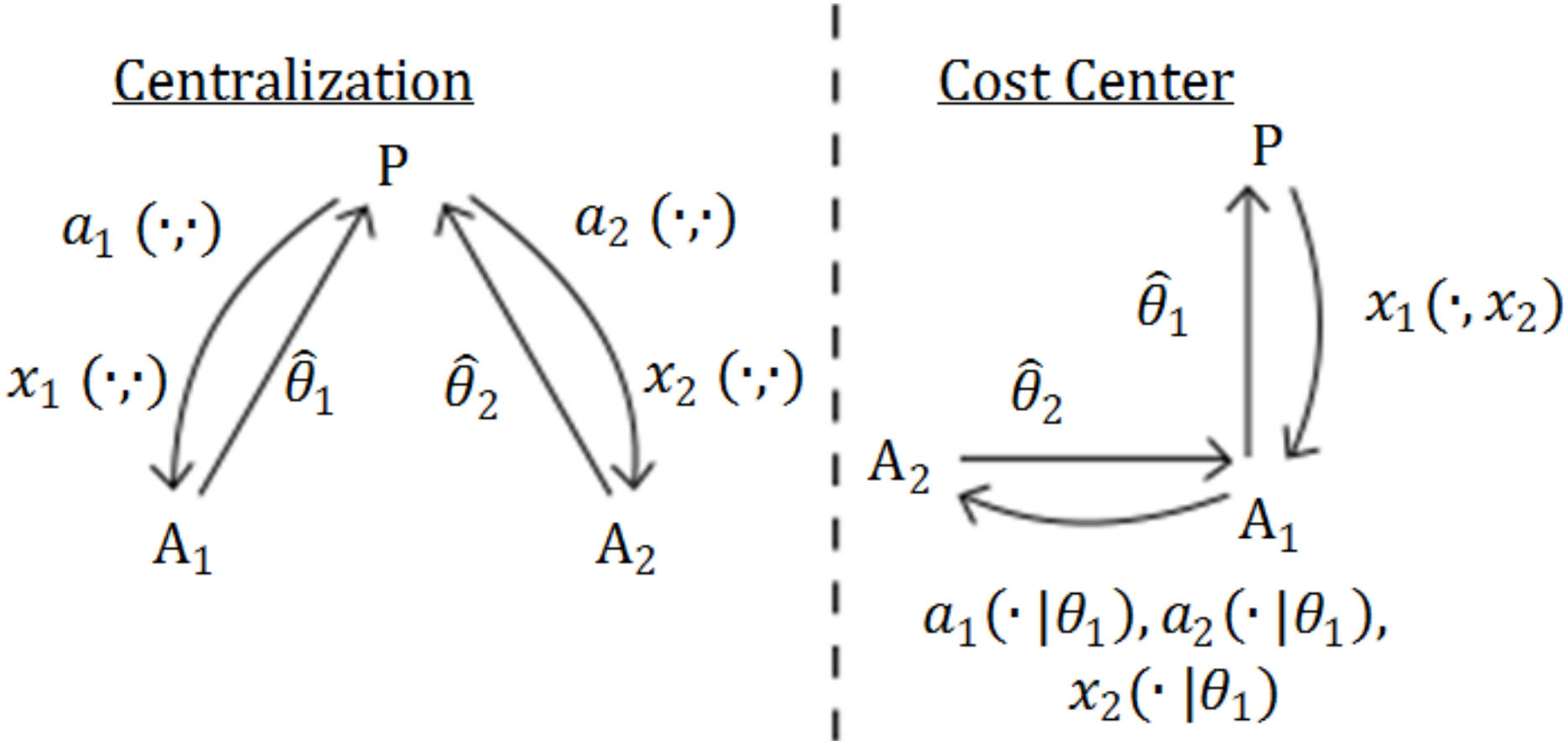
Centralization:

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 $\widehat{\theta}_1, \widehat{\theta}_2$ to P about their costs θ_1, θ_2.

2. A_1 and A_2 simultaneously choose reports $\widehat{\theta}_1$ and $\widehat{\theta}_2$ (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
P
$a_1(\cdot,\cdot)$
$a_2(\cdot,\cdot)$
$x_1(\cdot,\cdot)$
$\hat{\theta}_1$
$\hat{\theta}_2$
$x_2(\cdot,\cdot)$
A_1
A_2
Cost Center
P
$\hat{\theta}_1$
$x_1(\cdot, x_2)$
$\hat{\theta}_2$
A_2
A_1
$a_1(\cdot|\theta_1), a_2(\cdot|\theta_1),$
$x_2(\cdot|\theta_1)$

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 θ_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 $\widehat{\theta}_1$ about θ_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 θ_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 $\widehat{\theta}_2$ about his cost θ_2.
4. A_2 chooses report $\widehat{\theta}_2$.
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.*

- P's ability, under a cost center, to make the payment x_1 to A_1 contingent on A_1's payment x_2 to A_2 is crucial for this result.

Benchmark with no limitations on communication

Example: $B(a_1, a_2) = a_1 + a_2$; θ_1, θ_2 i.i.d.

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

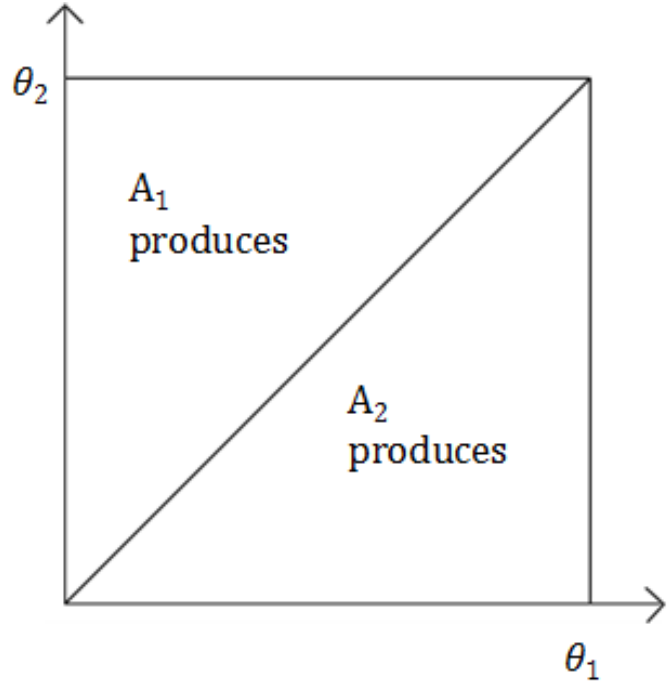
- P pays information rents to both agents to ensure truthful reporting is a BNE.

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 θ_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 θ_i is a continuous variable, each agent can make only a finite number of possible reports. Here, assume only 2 possible reports.

- **How** private information is coarsened can be determined optimally under each structure.

- 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 θ_1, not on a coarse summary of it.

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

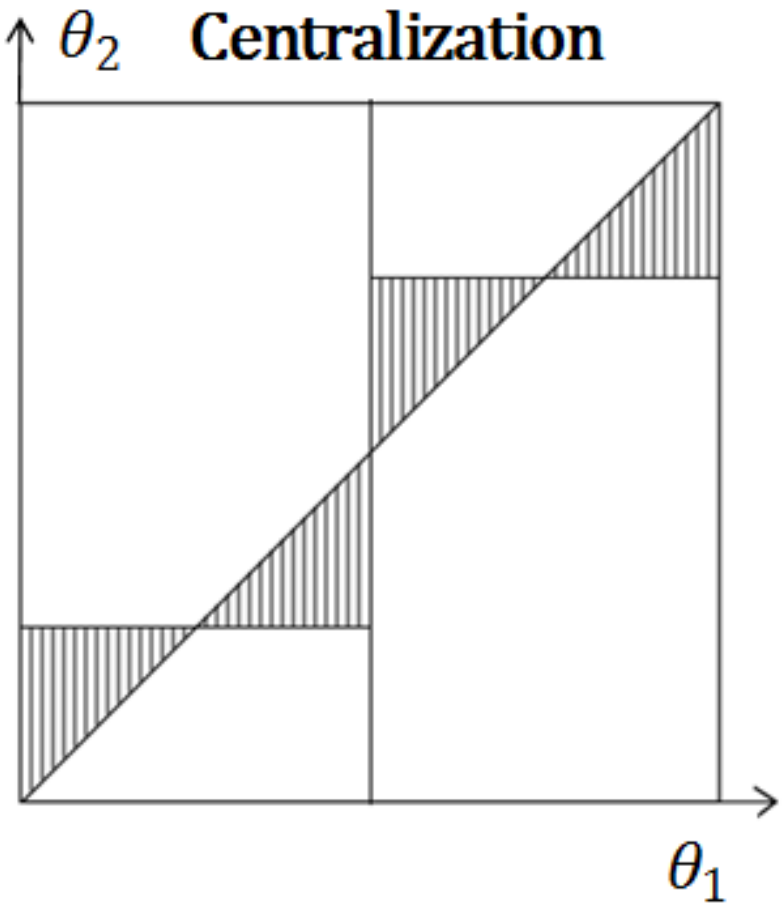
Centralization vs. delegation given limitations on communication

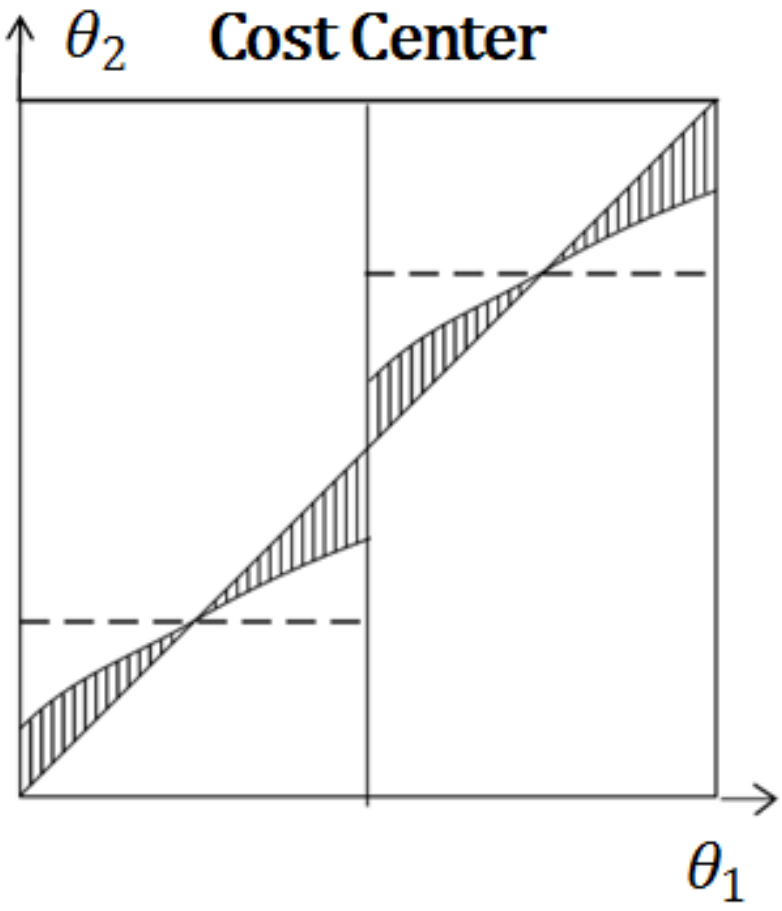
Proposition: *With limited communication, delegation to a cost center dominates centralization.*

Centralization: Coarseness of communication makes it preferable for P to communicate with agents **sequentially**, rather than simultaneously.

- The optimal way to ask A_2 to partition the values of θ_2 depends on which of the two reports A_1 made about θ_1.

Delegation to a cost center: The inefficiency in production decisions resulting from coarseness of communication can be made 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 θ_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 $\Longrightarrow$ 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.

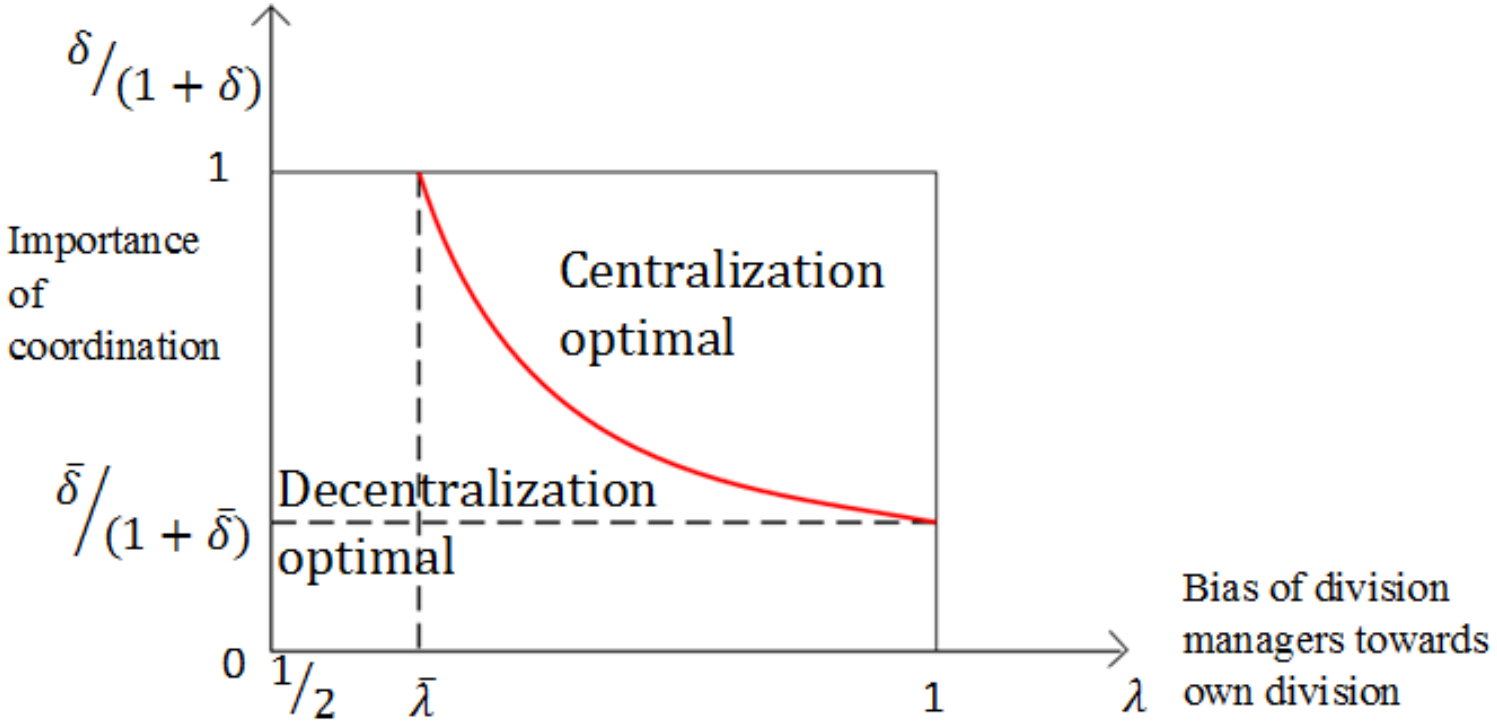
Many recent papers assume that although top management cannot commit to rules specifying actions and payments as functions of reports, it **can commit to delegate decision rights**. Ask: When is such "formal delegation" of decision rights advantageous?

- Aghion + Tirole (*JPE*, 97): Formal delegation beneficial when crucial to give subordinates incentives to generate ideas and/or acquire information .

- Dessein (*REStud*, 02): Formal delegation beneficial when subordinate has a small bias or a big info advantage (top mgt's prior very dispersed).

- Alonso et al (*AER*, 08): Study an environment like Dessein's but with two division mgrs, each privately informed about shocks to own division and each biased towards own division. HQ cares about both adapting decisions to shocks and coordinating them.

Centralization vs. delegation when only decision rights are contractible (Alonso, Dessein, and Matouschek, *AER*, 08):



For headquarters to prefer centralization, coordination must be sufficiently important **and** *division managers must be sufficiently biased towards their own divisions.*

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 θ_i is privately observed by M_i, and (θ_1, θ_2) are i.i.d., $\sim U[-s, s]$
  - Parameter $\delta \in (0, \infty)$ is common knowledge; measures, for each division, **importance of coordination** (of d_1 and d_2) relative to adaptation (of d_i to θ_i)

- M_i's payoff is $\lambda\pi_i + (1 - \lambda)\pi_j$
  - Parameter $\lambda \in (\frac{1}{2}, 1]$ is common knowledge; 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 θ_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 θ_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.

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

- **ADM show Decentralization peforms better w.r.t. adaptation:** yields lower expected value of $(d_1 - \theta_1)^2 + (d_2 - \theta_2)^2$.

Centralization (C): Each M_i simultaneously makes a cheap-talk report about θ_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 θ_i to M_j, and then each M_i chooses d_i.

- Benefits of **C** from HQ's perspective:
  1. decision-making is unbiased—under **D**, managers underweight coordination, since $\lambda > \frac{1}{2}$
  2. equilibrium quality of communication is better (less coarse)—$\lambda > \frac{1}{2}$ means preferences of M_i and HQ are more aligned than those of M_i and M_j
- Benefit of **D** from HQ's perspective: flexibility gain since each M_i bases choice of d_i on precise value of own θ_i
- Contrast Melumad et al model: There, in **D**, transfers to A_1 can eliminate A_1's bias. And coarseness of communication is exogenous and equal in **C** and **D**. So there, only flexibility gain matters for comparison btw **C** and **D**.

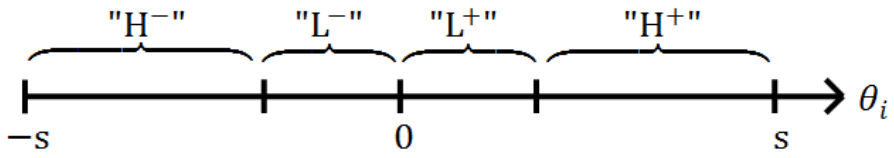
Equilibrium communication under C and D

Claim: *Under both* **C** *and* **D**, *each* M_i *has incentive to* **exaggerate** $|\theta_i|$.

- NB: θ_1, θ_2 independent $\Longrightarrow E(\theta_j|\theta_i) = 0 \ \forall \theta_i \Longrightarrow E(d_j|\theta_i) = 0 \ \forall \theta_i$.
- HQ's payoff: $-(d_1 - \theta_1)^2 - (d_2 - \theta_2)^2 - 2\delta(d_1 - d_2)^2$
- M1's payoff: $-\lambda(d_1 - \theta_1)^2 - (1-\lambda)(d_2 - \theta_2)^2 - \delta(d_1 - d_2)^2$
- M2's payoff: $-(1-\lambda)(d_1 - \theta_1)^2 - \lambda(d_2 - \theta_2)^2 - \delta(d_1 - d_2)^2$
- Under **C**, from M_i's perspective, HQ puts too much weight on coordinating d_i and d_j relative to adapting d_i to θ_i $(\frac{2\delta}{1} > \frac{\delta}{\lambda})$, so $|d_i|$ is not sensitive enough to $|\theta_i|$. Hence, M_i has incentive to exaggerate $|\theta_i|$ to HQ.
- Under **D**, from M_i's perspective, M_j puts too little weight on coordinating d_i and d_j relative to adapting d_j to θ_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 $|\theta_i|$ to M_j.

Equilibrium communication under C and D

Claim: *Under both* **C** *and* **D**, *all communication eqa are interval eqa, with size of intervals* $\uparrow$ *in* $|\theta_i|$.



- If and only if $\theta_i = 0$, there is no preference misalignment btw M_i (sender) and HQ/M_j (receiver).

- As $|\theta_i|$ increases, preference misalignment btw M_i (sender) and HQ/M_j (receiver) increases, hence incentive to exaggerate $|\theta_i|$ increases.

- And as in Crawford-Sobel model, where preference misalignment (bias) did not vary with the state, in eqm more tempting messages must be coarser.

- Hence here, communication is coarser the larger is $|\theta_i|$, for two reasons.

- Because, for $\theta_i = 0$, there is no preference misalignment, there is no upper limit on number, N, of messages that can arise in eqm. Coarseness gets very small as $|\theta_i| \to 0$.

- ADM focus on the most informative eqm ($N \to \infty$). But this is not a FRE. Informativeness depends on λ and δ.

Comparing equilibrium informativeness of communication under C vs D

Claim: *Eqm communication is coarser, hence less informative, under* **D** *than* **C**.

- Preference misalignment is greater btw M_1 and M_2 (who communicate under **D**) than btw M_i and HQ (who communicate under **C**)—because $\lambda > \frac{1}{2}$. Hence incentive to exaggerate $|\theta_i|$ is greater under **D** than **C**.
  - As $\lambda \to \frac{1}{2}$, preferences of M_i, M_j, and HQ become aligned, so both **D** and **C** yield fully informative communication in eqm.

- 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 $|\theta_i| \downarrow$.

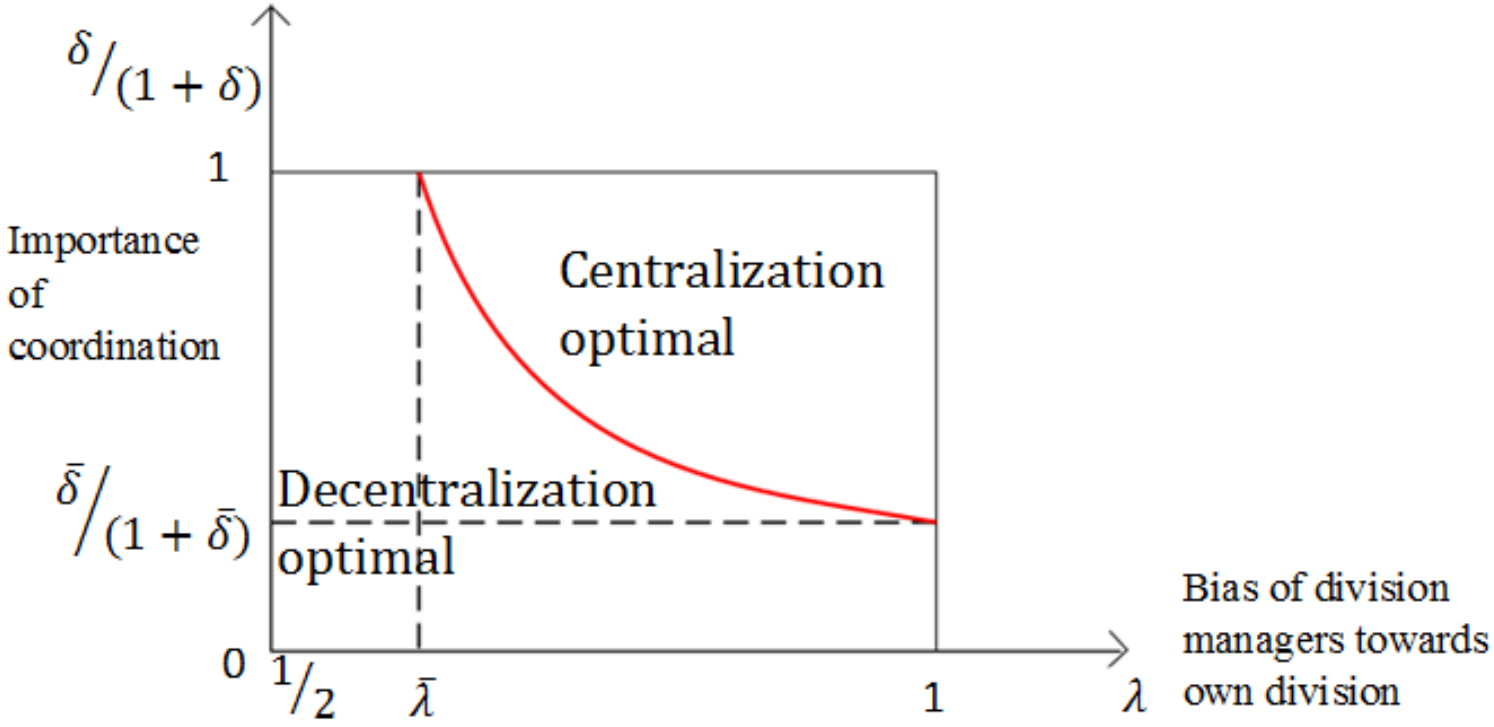
- As $\delta \to \infty$, eqm commun. becomes equally informative under **D** and **C**.
  - As $\delta \to \infty$, under both **D** and **C**, decisions are made in the same way given posterior means of θ_1 and θ_2: $d_1 = d_2 =$ average of these posterior means.
  - Hence communication incentives become the same as $\delta \to \infty$.

HQ's preference between centralization and decentralization

Claim: **C** *and* **D** *perform equally well as* $\lambda \to \frac{1}{2}$, $\delta \to 0$, *or* $\delta \to \infty$.

- As $\lambda \to \frac{1}{2}$ or $\delta \to 0$, preference misalignments among M_i, M_j, and HQ disappear, so HQ attains full-information payoff.

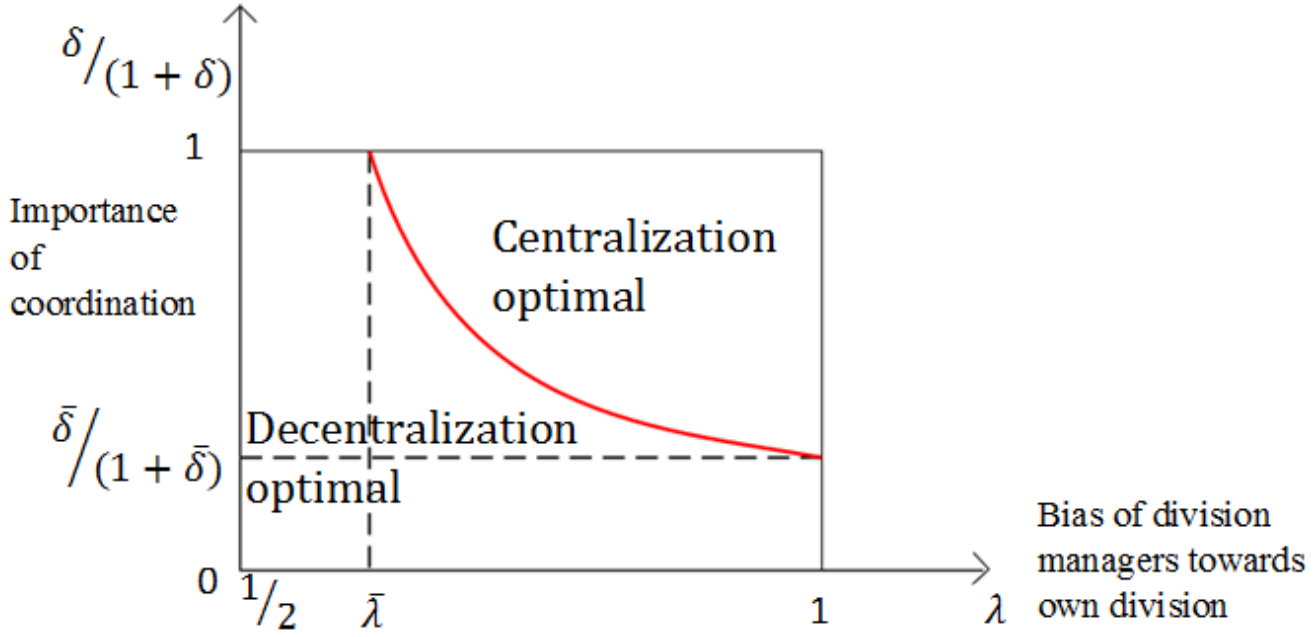
- As $\delta \to \infty$, decision-making and communication under **C** and **D** converge.



HQ's preference between centralization and decentralization

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

- Surprisingly, if $\lambda \leq \bar{\lambda}$, then even as δ gets very large, **C** never dominates **D**.
  - For λ small, bias in decision-making under **D**, from HQ's perspective, is small, and communication is not much less informative under **D** than **C**.
  - Flexibility gain under **D** outweighs both bias in decision-making and coarser communication.



Experimental test of Alonso et al model by Evdokimov and Garfagnini (*QE*, 2019)

1. Under both C and D, elicited receivers' beliefs about senders' shocks. Hence could **separately** analyze communication and decision-making.

2. Data largely consistent with predictions about eqm **communication** in most informative eqm.

3. Systematic deviations from predictions about eqm **use of communicated information** (decision-making):

   - Coordination was **overweighted** by managers under Decentralization.
   - Coordination was **underweighted** by HQs under Centralization.

4. Payoff losses (relative to eqm predictions) largely due to distortions in decision rules rather than miscommunication.

5. Deviations in decision rules from eqm predictions disappear in treatments where θ_1, θ_2 are **publicly observed**.

6. Hence results indicate that decision-makers overweight the objective where they face uncertainty (coordination under D; adaption under C).

Further questions about delegation

1) 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 exert effort 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, θ_p and θ_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 (θ_p, θ_a).
  - Optimal rule specifies that the larger is θ_a, the larger is the minimum acceptable level of θ_p.

2) 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 a relational contract support *informal delegation* of decision rights by a boss (P) to a subordinate (A)?

- In each period of ∞-horizon relationship, A privately chooses effort to find a good project. A's and P's preferences over projects are misaligned.

- 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" in eqm. 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.