Pay to Quit

Robust Team Incentives with Hidden Need

Hyungmin Park Youngji Sohn

University of Warwick

[for 45 min]

Software release

A company is shipping a major release built from many interdependent modules

Each engineer privately inspects her module:

- Some modules have *latent bugs*
- Others are *already stable*

If an engineer intervenes, she privately chooses her diligence:

Thorough debugging (costly and effective) vs. a quick patch (cheap and ineffective)

What management observes:

- ✓ Who pushed commits (*participation*)
- × Whether the module actually had a bug (*type*)
- × How carefully the engineer tested (*effort*)
- × Individual module quality
- ✓ Whether the release succeeds (*coarse outcome*)

Pay only participants?

Common practice: organizations reward workers who participate in projects

- Participation is observable; actual value-added is not
- Paying non-participants feels wasteful

When workers differ in productive capacity, this creates problems

- Participation-based pay creates pooling: all types participate
- Team size bloats beyond what is necessary, inflating externalities
- Principal must size bonuses for the worst case as if every module needs fixing

Our question: How to optimally and robustly implement project success
when uncertainty spans both types and actions?

This paper

Setting. Team production with complementarities, hidden types, and hidden effort

Objective. Ensuring the maximum probability of project success,
robust to players' arbitrary beliefs about others' types and actions

Solution concept. Belief-free rationalizability (Bergemann & Morris, 2017)

Our findings. Optimal contract that achieves unique implementation of team project success

Today. Focus on fault-repair technology

Related literature

Contracting with externalities

- Segal (2003): divide-and-conquer under multilateral externalities
- Bernstein & Winter (2012): heterogeneous participation externalities

Team moral hazard with complementarities — unique implementation under action uncertainty

- Winter (2004): hierarchical contracts
- Halac, Lipnowski & Rappoport (2021): principal-designed uncertainty

Robust implementation — robust to beliefs about types

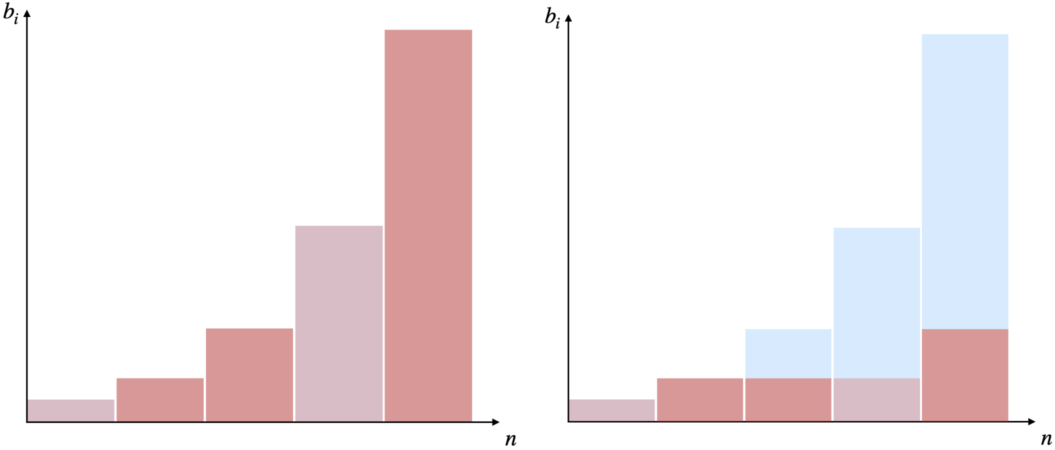
- Bergemann & Morris (2017); Dekel, Fudenberg & Morris (2007)

Mechanism design with evidence — exogenous evidence structure

- Green & Laffont (1986); Bull & Watson (2007)

This paper: unique implementation robust to *both* type and action uncertainty, with *endogenous* evidence created through contract design.

Preview: When only 3/5 needs fixing



Example: Symmetric $n = 2$

A team of 2 experts

A project is managed by two experts, $i \in \{1, 2\}$

Types. Each privately observes $\theta_i \in \{L, H\}$ with $\Pr(\theta_i = L) = p \in (0, 1)$

Effort. Each then privately chooses $e_i \in \{W, S\}$, where W costs $c > 0$

Fault-Repair Technology.

Part quality

$$q(\theta_i, e_i) = \begin{cases} q_H & \text{if } \theta_i = H \\ q_H & \text{if } \theta_i = L \text{ and } e_i = W \\ q_L & \text{if } \theta_i = L \text{ and } e_i = S \end{cases}$$

Project success

$$P(\{1, 2\}) = 1, \qquad P(\{1\}) = P(\{2\}) = \tfrac{1}{2}, \qquad P(\varnothing) = \tfrac{1}{4}$$

Individual effort decisions are strategic complements

Principal's problem

The principal aims to implement all-high-quality $(q_1 = q_2 = q_H)$ as a unique outcome and that at minimum cost

What can the principal condition for?

- Not effort — unobservable
- Not types — private
- But she can require a participation choice — observable, constrains subsequent actions

Participation as an observable margin

Each expert, after observing type, chooses whether to participate ($a_i = 1$) or opt out ($a_i = 0$)

- $a_i = 1$: Participants then privately choose $e_i \in \{W, S\}$
- $a_i = 0$: Opting out is effectively $e_i = S$

$\implies$ Actions: $(a_i, e_i) \in \{(0, S), (1, W), (1, S)\}$

We analyze a simultaneous-move game, where participation profile $\mathbf{a} = (a_1, a_2)$ is observed ex-post

Key features: opting out reveals shirking, so participation serves as partial evidence for effort

Contract. Bonus $b_i(\mathbf{a}) \geq 0$, conditioned on participation profile and project outcome

Payoffs and cost

Let $Q = \{j : q_j = q_H\}$

Expert i's payoff given a realization $(\boldsymbol{\theta}, \mathbf{a}, \mathbf{e})$:

$$u_i = P(Q)\, b_i(\mathbf{a}) \;-\; c\,\mathbf{1}\{e_i = W\}$$

Principal's expected cost under $\mathbf{p}$:

$$C_p(\mathbf{b}) = \mathbb{E}_{\mathbf{p}}\left[P(Q) \cdot \sum_{i \in N} b_i(\mathbf{a})\right]$$

Solution concept?

Challenge: Complementarities $\Rightarrow$ multiple equilibria
Uncertainty about others' **types** and **actions** interacts

Belief-Free Rationalizability (Bergemann & Morris, 2017):

For each type of each expert, iteratively eliminate actions that are never a best response for any **arbitrary belief** over others' types and surviving actions

- "Belief-free": beliefs over others' types and actions are unrestricted
  no prior, no consistency with a type space
- Very permissive solution concept, weaker than BNE and ICR

Unique implementation under BFR:

All-high-quality outcome uniquely survives, regardless of what experts believe about each other

The prediction does not rely on a specific belief model: **very robust**

Two contract forms

There can be two forms of contract offered to each expert

- Participation-Only: $b_i(0, a_{-i}) = 0$. Non-participants are not paid
- Pay-to-Quit: $b_i(0, a_{-i}) > 0$. Non-participants are paid. $b_i(0, a_{-i}) > b_i(1, a_{-i})$ for non-trivial separation

Let's analyze two situations (1) PO for both experts vs (2) PtQ for both experts

Benchmark: Participation-Only

Participation-only contracts

Under PO, all types participate $\Longrightarrow$ Only team moral hazard remains (Winter, 2004)

Resolve experts in order (Divide-and-conquer)

Expert 1 with $\theta_1 = L$: W must dominate S, even at the worst case belief $Q_{-1} = \varnothing$

$$b_1 \cdot P(\{1\}) - c > b_1 \cdot P(\varnothing) \quad \Longleftrightarrow \quad b_1 > \frac{c}{P(\{1\}) - P(\varnothing)} \quad \Longrightarrow \quad b_1^* = 4c$$

Expert 2 with $\theta_2 = L$: Once rank 1's part is assured, 2's worst case belief $Q_{-2} = \{1\}$

$$b_2 \cdot P(\{1,2\}) - c > b_2 \cdot P(\{1\}) \quad \Longleftrightarrow \quad b_2 > \frac{c}{P(\{1,2\}) - P(\{1\})} \quad \Longrightarrow \quad b_2^* = 2c$$

$C^{PO} = P(\{1,2\}) \cdot (4+2)c = 6c, \quad$ independent of p

Both experts paid regardless of type. Hierarchy sized for worst case $\boldsymbol{\theta} = (L, L)$

PO: Remarks

Connection to Winter (2004):

- The wages in PO contracts are as if complete information with $\boldsymbol{\theta} = \mathbf{L}$
- As if everyone has potential to contribute

Why pooling cannot improve on $\boldsymbol{\theta} \neq \mathbf{L}$:

- **Local inefficiency:** $\theta_i = H$ are pooled with $\theta_i = L$ despite adding no value
- **Global inefficiency:** having $\theta_i = H$ in the effort inducing problem inflates **all** higher-ranked workers' strategic uncertainty
- Even if $\boldsymbol{\theta} = \mathbf{H}$, everyone participates and creates strategic uncertainty of size n

Separation: PtQ contract

Pay to Quit: Round 0

Set $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$. Experts are paid for opting out.
Start from $BFR_i^0 = \{(0, S), (1, S), (1, W)\}$

Round 0. Type separation

For $\theta_i = H$:

- $(1, W)$ dominated by $(1, S)$: same quality, saves c
- $(1, S)$ dominated by $(0, S)$: same quality, but $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$

For $\theta_i = L$:

- $(1, S)$ dominated by $(0, S)$: same quality q_L, but higher bonus

$$BFR_i^1(\theta_i = H) = \{(0, S), \cancel{(1, S)}, \cancel{(1, W)}\}, \qquad BFR_i^1(\theta_i = L) = \{(0, S), \xcancel{(1, S)}, (1, W)\} \qquad \forall\, i$$

Participation assures quality

$$BFR_i^1(\theta_i = H) = \{(0, S)\}, \qquad BFR_i^1(\theta_i = L) = \{(0, S), (1, W)\} \qquad \forall\, i$$

Key observation: From expert 1's perspective

- If $a_2 = 1$: only $(1, W)$ survives $\Rightarrow q_2 = q_H$
- If $a_2 = 0$: type unknown $\Rightarrow q_2$ could be q_H or q_L

Therefore, $a_2 = 1$ implies $q_2 = q_H$ under any arbitrary belief of expert 1

$\Longrightarrow$ Expert 2's participation assures $q_2 = q_H$ to Expert 1

PtQ: Round 1

Round 1. Expert 1 with $\theta_1 = L$ compares $(1, W)$ vs. $(0, S)$

Critical bonus depending on a_2:

- If $a_2 = 1$, expert 2 participates $\implies$ $q_2 = q_H$ in any belief of expert 1

$$b_1 \cdot \big[P(\{1,2\}) - P(\{2\})\big] \geq c \implies b_1^* = 2c$$

- If $a_2 = 0$, expert 2 opts out, type unknown $\implies$ $q_2 = q_L$ in worst-case belief of 1

$$b_1 \cdot \big[P(\{1\}) - P(\varnothing)\big] \geq c \implies b_1^* = 4c$$

We can find perturbation $1 < \frac{\varepsilon_1(a_1 = 0)}{\varepsilon_1(a_1 = 1)} < \frac{P(\{1\})}{P(\varnothing)}, \frac{P(\{1,2\})}{P(\{2\})}$

making sure $\varepsilon_1(a_1 = 0)$ is not too large reversing $\theta_1 = L$'s incentive making her to opt out.

$$BFR_1^2(\theta_1 = H) = \{(0, S)\}, \qquad BFR_1^2(\theta_1 = L) = \{\cancel{(0, S)}, (1, W)\}$$
$$BFR_2^2(\theta_2 = H) = \{(0, S)\}, \qquad BFR_2^2(\theta_2 = L) = \{(0, S), (1, W)\}$$

PtQ: Round 2

Round 2. Expert 2 with $\theta_2 = L$. Since expert 1 is resolved, 2 believes that $q_1 = q_H$

$$b_2^* = \frac{c}{P(\{1,2\}) - P(\{1\})} = 2c$$

Similarly, we can find $1 < \frac{\varepsilon_2(a_2 = 0)}{\varepsilon_2(a_2 = 1)} < \frac{P(\{1,2\})}{P(\{2\})}$ so that

$$BFR_1^3(\theta_1 = H) = \{(0, S)\}, \qquad BFR_1^3(\theta_1 = L) = \{(1, W)\}$$

$$BFR_2^3(\theta_2 = H) = \{(0, S)\}, \qquad BFR_2^3(\theta_2 = L) = \{\cancel{(0, S)}, (1, W)\}$$

Expected cost.

$$C^{PtQ}(p) = P(\{1,2\}) \, [\underbrace{p \cdot 2c + (1-p) \cdot 4c}_{\mathbb{E}[b_1^*]} + 2c] = (6 - 2p)\, c \; < \; 6c = C^{\mathrm{PO}}$$

Participation as endogenous evidence

Under PO, participation is **uninformative**: everyone participates regardless of type.

Under PtQ, the contract **endogenously creates** an evidence structure:

- Effort remains a hidden action the principal never observes
- Yet after one round of elimination, $a_i = 1$ becomes conclusive evidence for $e_i = W$
- Unlike the *exogenous* evidence structures in the literature,
  here the participation margin is evidence the principal *builds* through contract design

PtQ: Remarks

1. Participation becomes conclusive evidence for effort

   $a_i = 1$ verifies $e_i = W$, $a_i = 0$ verifies $e_i = S$

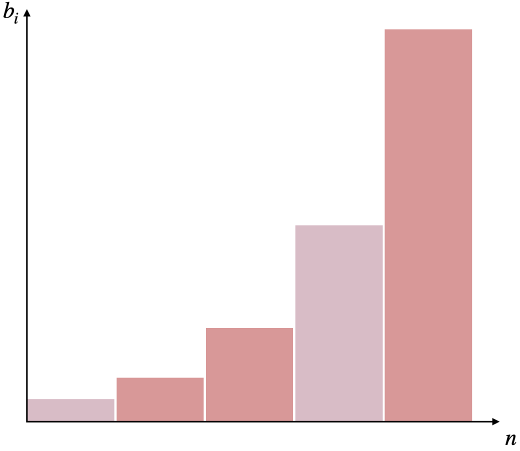
2. Hierarchy depends on number of participants

   i's critical bonus $b_i^*(\mathbf{a})$ is hierarchical bonus only counting lower-ranked *non-participants*

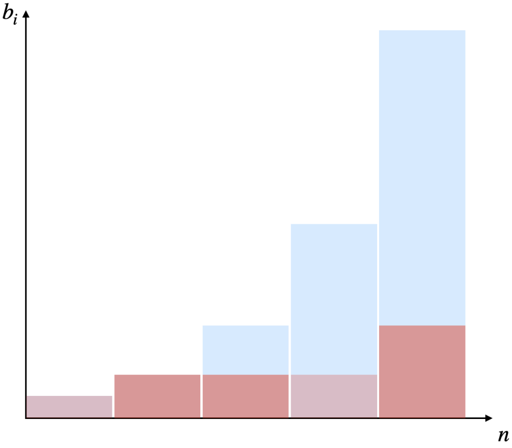
3. **Expected cost** depends on the prior $\mathbf{p}$

4. **Implementation** does not

PO vs. PtQ: When 3/5 needs fixing



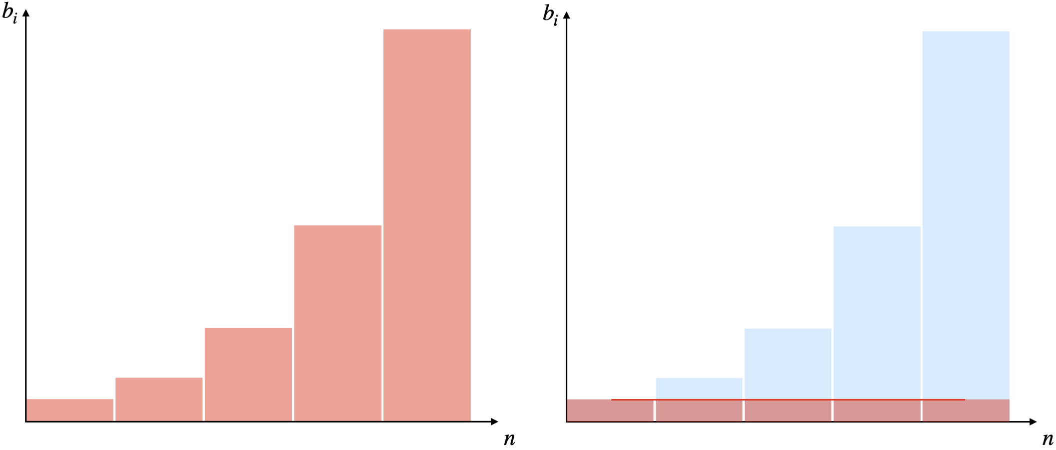
(a) PO



(b) PtQ

In PO, 5 participate and receive Winter payments with full length 5 hierarchy
In PtQ, 3 participate and the hierarchy reduces to 3 = (5 - 3 + 1)

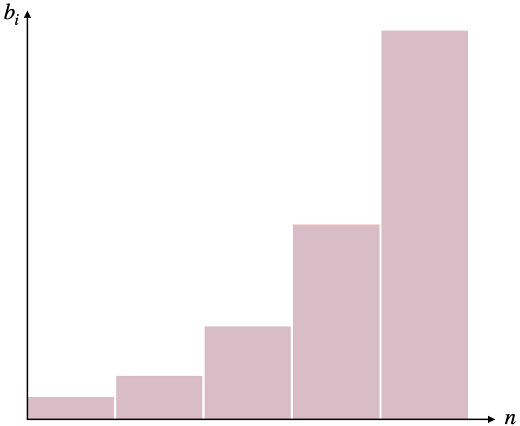
PO vs. PtQ: When 5/5 needs fixing



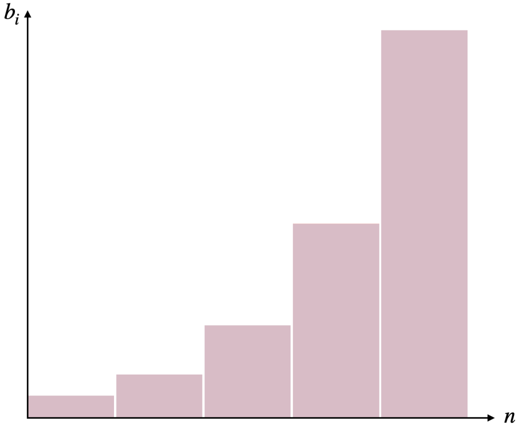
(a) PO (b) PtQ

In PO, 5 participate and receive Winter payments with full length 5 hierarchy
In PtQ, 5 participate and the hierarchy reduces to $1 = (5 - 5 + 1)$

PO vs. PtQ: When 0/5 needs fixing



(a) PO



(b) PtQ

In PO, 5 participate and receive Winter payments with full length 5 hierarchy
In PtQ, 5 opt out and the hierarchy is full 5

General Results

General setting

- n experts with heterogeneous type dist. $p_i = \Pr(\theta_i = L) \in (0,1)$ and effort costs $c_i > 0$
- Team production success probability $P(Q)$ is (1) *strictly increasing* and (2) *strictly supermodular* in high-quality set Q:

$$P(Q \cup \{i\}) - P(Q) \;<\; P(Q' \cup \{i\}) - P(Q') \qquad \forall\, i,\; Q \subsetneq Q' \subseteq N \setminus \{i\}$$

Bonus schedule b uniquely implements high quality (UI)
if there is a vanishing sequence of perturbations $\varepsilon^k > 0$ such that every BFR outcome under $\mathbf{b} + \varepsilon^k$ yields all-high-quality. (as the set of UI contracts is typically open)

Optimal UI contract:

$$\min_{\mathbf{b} \,\in\, \mathrm{UI}} \; P(N)\,\mathbf{E}_{\mathbf{p}}\left[\sum_{i \in N} b_i\big(\mathbf{a}(\theta)\big)\right]$$

$n = 2$ easily extends

- **PO**: divide-and-conquer, a hierarchy sized for n
- **PtQ**: quit bonus* $=$ participation bonus *, with a hierarchy sized for $\{i : a_i = 0\}$
  (1) every H opts out, then
  (2) each L is induced to participate and work, one by one along σ

Optimal contract?

- UI contract is not necessarily PO or PtQ uniform across $\mathbf{a}_{-i}$ for each expert.
- But at each $(i, \mathbf{a}_{-i})$, $\mathbf{b}$ must either *pool* or *separate* i's type, so the design is a choice between PO and PtQ among these local options.

Optimal contract

Key spillover: switching $(i, \mathbf{a}_{-i})$ from PO to PtQ

- implements i's own action at the *same* critical value
- but makes i's participation informative $\Rightarrow$ enlarges predecessors' assured sets $\Rightarrow$ strictly lowers their bonuses

The design problem collapses to finding an optimal ranking: $\sigma^* \in \underset{\sigma \in \mathrm{Perm}(N)}{\operatorname{argmin}} C^{\mathrm{PtQ}}(\sigma, \mathbf{p})$

Proposition (Optimal UI contract)

An optimal UI contract is a PtQ contract for all $(i, \mathbf{a}_{-i})$ for some $\sigma \in \mathrm{Perm}(N)$.

In general, not possible to characterize the optimal ranking

Optimal UI contract: Proof sketch

Start from any UI contract $\mathbf{b}$:

1. Characterize: Let $\tau(i)$=elimination round for expert i and $S_i(\mathbf{a}_{-i})$=assured sets
2. Fill in: Choose a ranking σ refining (breaking ties in) τ
   Extend bonuses to all off-path profiles (unreachable) to critical values $\Rightarrow$ same cost
3. Reduce: Replace bonuses with critical values along σ $\Rightarrow$ weakly lower cost
4. Switch to PtQ: Expert by expert, PO-form $\rightarrow$ PtQ-form
   Each switch preserves UI and enlarges predecessors' assured sets $\Rightarrow$ weakly lower cost
5. Optimize: $C(\mathbf{b}) \geq C^{\mathrm{PtQ}}(\sigma, \mathbf{p}) \geq \min_{\sigma'} C^{\mathrm{PtQ}}(\sigma', \mathbf{p})$

Special case: O-ring production

O-ring production function: $P(Q) = \prod_{i \notin Q} \lambda_i$ with $0 < \lambda_i < 1$

- Every expert i is a "generalist", with i's marginal contribution $\propto P(Q_{-i})$
- Each later-resolved j multiplies predecessors' bonuses by $\kappa_j := p_j + \dfrac{1 - p_j}{\lambda_j} > 1$:

$$b^*_{\sigma(i)} = \frac{c_i}{1 - \lambda_i} \prod_{\sigma(j) > \sigma(i)} \kappa_j$$

Proposition (Optimal ranking under O-ring)

The optimal contract in O-ring production ranks experts by increasing priority index $I_i = \dfrac{c_i \, \lambda_i}{(1 - p_i)(1 - \lambda_i)^2}$ *and offers PtQ for each* $(i, \mathbf{a}_{-i})$.

Rank early: *cheap* (low c_i), *critical* (low λ_i), and *likely-already-good* (low p_i) experts.
Under symmetric tech λ, $I_i \propto c_i/(1 - p_i)$.
Under symmetric tech & cost, rank by increasing p_i (most-likely-good first).

Conclusion

Paying non-participants is not paying for idleness

It purchases **endogenous evidence** that makes participation informative about effort

Why it works:

- PtQ converts adverse selection into a coordination device
- Participation separates types $\rightarrow$ shrinks strategic uncertainty $\rightarrow$ lowers bonuses
- Optimal among all UI contracts, not just within a restricted class

Takeaway:
Informational content of observable actions is not a primitive but shaped by incentive structure

Thank you!

Appendix

Belief-Free Rationalizability

Definition (Belief-free rationalizable actions)

Initialize: $B^{\mathbf{b}}_{i,0}(\theta_i) = \mathcal{A}_i = \{(0, S), (1, S), (1, W)\}$ for all $i \in N$, $\theta_i \in \Theta$

Iterate: $(a_i, e_i) \in B^{\mathbf{b}}_{i,k}(\theta_i)$ survives in $B^{\mathbf{b}}_{i,k+1}(\theta_i)$ iff there exists an arbitrary conjecture $\nu \in \Delta(\Theta_{-i} \times \mathcal{A}_{-i})$ such that:

(i) $\nu(\boldsymbol{\theta}_{-i}, \mathbf{a}_{-i}, \mathbf{e}_{-i}) > 0 \Rightarrow (a_j, e_j) \in B^{\mathbf{b}}_{j,k}(\theta_j)$ for all $j \neq i$

(ii) $(a_i, e_i) \in \arg\max_{(a'_i, e'_i) \in \mathcal{A}_i} \sum u^{\mathbf{b}}_i(\theta_i, a'_i, e'_i \mid \boldsymbol{\theta}_{-i}, \mathbf{a}_{-i}, \mathbf{e}_{-i})\, \nu(\cdot)$

Limit: $B_i(\theta_i \mid \mathbf{b}) = \bigcap_{k \geq 0} B^{\mathbf{b}}_{i,k}(\theta_i)$

PtQ: General characterization

$$b_i^* = \frac{c_i}{P(W_{-i} \cup \{i\} \cup R^{\sigma}_{<i}) - P(W_{-i} \setminus \{i\} \cup R^{\sigma}_{<i})}$$

where $W(\mathbf{a}) = \{j \in N : a_j = 1\}$ are participants who are guaranteed to work hence q_H and $R^{\sigma}_{<i}$ is the resolved experts before expert i.

Perturbations: Set $b_i(1, \mathbf{a}_{-i}) = b_i^* + \varepsilon_{i,1}$ and $b_i(0, \mathbf{a}_{-i}) = b_i^* + \varepsilon_{i,0}$.

$$1 < \frac{\varepsilon_{i,0}}{\varepsilon_{i,1}} < \frac{P(W_{-i} \cup \{i\} \cup R^{\sigma}_{<i})}{P(W_{-i} \setminus \{i\} \cup R^{\sigma}_{<i})} \qquad \forall \, \mathbf{a}_{-i}$$

must hold to break ties to induce the desired action for each type

By strict monotonicity, $\frac{P(W_{-i} \cup \{i\} \cup R^{\sigma}_{<i})}{P(W_{-i} \setminus \{i\} \cup R^{\sigma}_{<i})} > 1$ for all $\mathbf{a}_{-i}$. Such perturbations always exist

O-Ring characterization proof

Under $P(Q) = \prod_{i \notin Q} \lambda_i$, expected cost under ranking σ:

$$C_{\mathbf{p}}^{PtQ}(\sigma) = \sum_{k=1}^{n} \frac{c_{i_k}}{1 - \lambda_{i_k}} \prod_{\ell > k} \kappa_{i_\ell}, \qquad \kappa_i := p_i + \frac{1 - p_i}{\lambda_i}$$

Adjacent swap condition: rank i before j iff $\dfrac{c_i/(1-\lambda_i)}{\kappa_i - 1} \leq \dfrac{c_j/(1-\lambda_j)}{\kappa_j - 1}$

$\Longrightarrow$ Optimal ranking: increasing $I_i = \dfrac{c_i\,\lambda_i}{(1-p_i)(1-\lambda_i)^2}$.

Fragile technology

Intervention without care *degrades* quality:

$$q(\theta_i, a_i, e_i) = \begin{cases} q_H & \text{if } \theta_i = H, a_i = 0 \\ q_H & \text{if } a_i = 1 \text{ and } e_i = W \\ q_L & \text{otherwise} \end{cases}$$

Now two contracts can separate types:

Contract 1 (PtQ): $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$ (same as before)

- Participants = assured high quality. Better when defects are common

Contract 2 (PtP): $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i}) - c_i$

- Non-participants = assured high quality. Better when parts are mostly stable
- After type separation, hierarchical wage sizing for participants

Optimal: assign PtQ when $p_i \leq 1/2$, PtP when $p_i > 1/2$

Common prior and Bayesian NE

Under a common prior **p** and BNE:

PtQ under BNE achieves cost of partial implementation in BFR

The principal can *tilt* bonuses: raise $b_i(1, a_{-i})$ when few participate, lower it when many do.
This preserves on-path expected payment while breaking all bad equilibria.

PO under BNE, cost falls relative to PO-BFR, but stays strictly above the PtQ-BNE cost.
Principal needs to compensate i for pooling j with $\theta_j = L$.

Cost ranking:

$$C^{PtQ,BNE} \ (= C^{PI,BFR}) \ < \ C^{PO,BNE} \ < \ C^{PO,BFR}$$

In either contract form, implementation is sensitive to belief structure.