# Pay to Quit

Hyungmin Park   Youngji Sohn

University of Warwick

PhD Micro Theory Workshop, May 2026

# A 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* — careful debugging can fix them
- Others are *already stable* — no new work is needed

**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

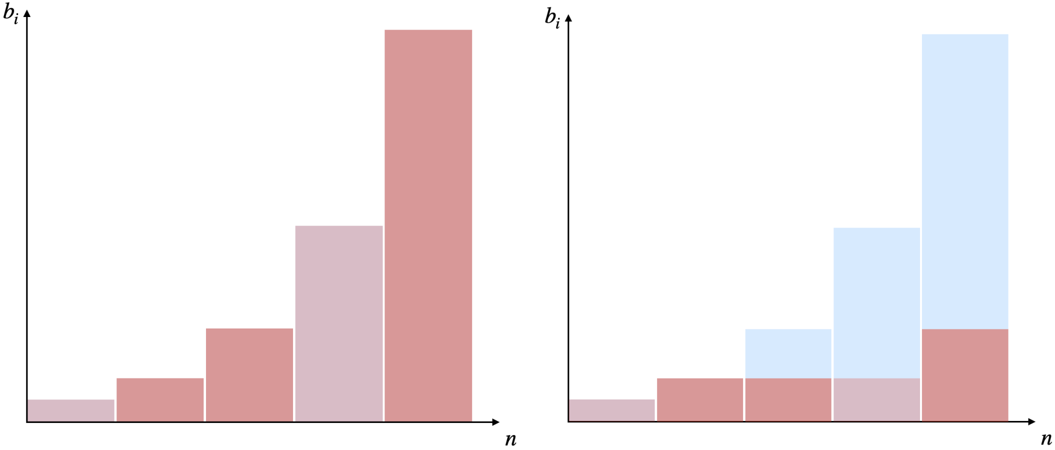
**Setting.** Team production with complementarities, hidden types, and hidden effort

**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)

**Today:** $n = 2$ examples, one technology (fault-repair)

# Preview: When only 3/5 needs fixing



# Setup

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

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

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

**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

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

# Principal's Problem

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

**What the principal observes:**

- × Types $\boldsymbol{\theta}$, effort $\mathbf{e}$, individual quality $q_i$
- ✓ Project outcome (success / failure)

**What can the principal pay for?**

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

# Setting: 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 — participation serves as partial evidence for effort

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

# Solution Concept?

**Challenge:** Complementarities ⇒ 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 NBR 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

We analyze two situations (1) PO for all experts vs (2) PtQ for all experts

# Participation-Only: Winter Hierarchy

Under PO ($b_i(0, a_{-i}) = 0$), all types participate. Participation is uninformative

Winter hierarchy for unique implementation:

**Rank 1:** Work must dominate shirk at worst belief $H_{-1} = \varnothing$:

$$b_1 \cdot \left[P(\{1\}) - P(\varnothing)\right] = b_1 \cdot \tfrac{1}{4} \geq c \qquad \Longrightarrow \qquad b_1 = 4c$$

**Rank 2:** Given expert 1 delivers $q_H$, worst belief $H_{-2} = \{1\}$:

$$b_2 \cdot \left[P(\{1,2\}) - P(\{1\})\right] = b_2 \cdot \tfrac{1}{2} \geq c \qquad \Longrightarrow \qquad b_2 = 2c$$

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

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

# Pay-to-Quit — Round 0

Set $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$. 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

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

# PtQ: Safe Harbor

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

**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

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

With some perturbations $\varepsilon_1(\mathbf{a}) \geq 0$,

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

# PtQ: Round 2

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

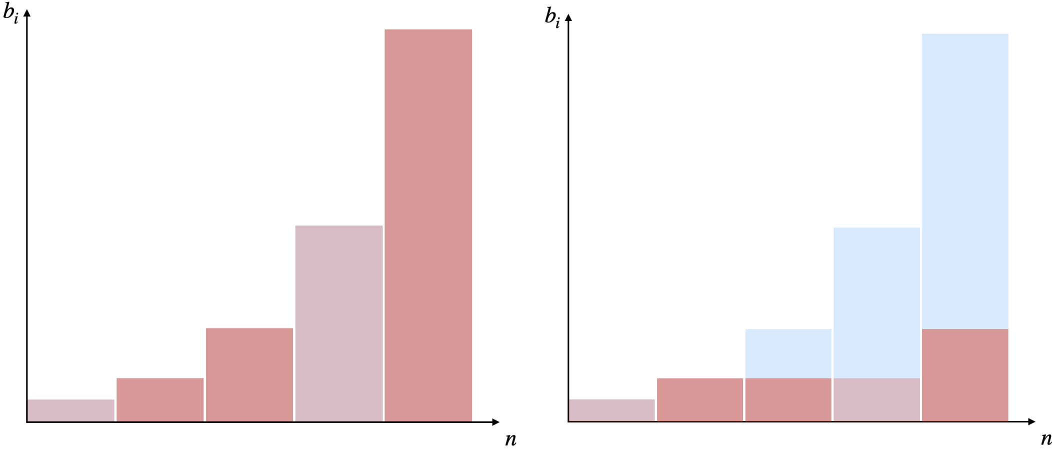
$$b_2^* = \frac{c}{P(\{1,2\}) - P(\{1\})} = 2c$$

With some perturbations $\varepsilon_2(\mathbf{a}) \geq 0$,

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

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

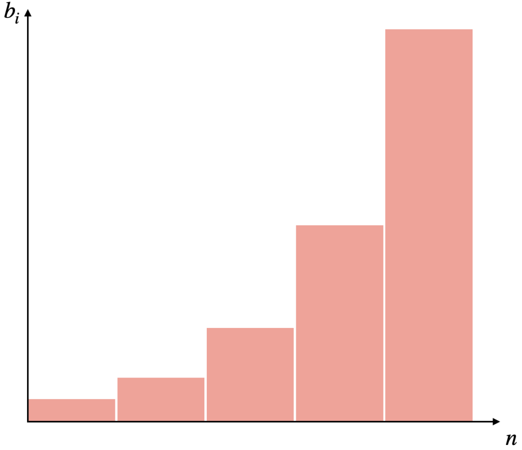
# PO vs. PtQ: When 3/5 needs fixing



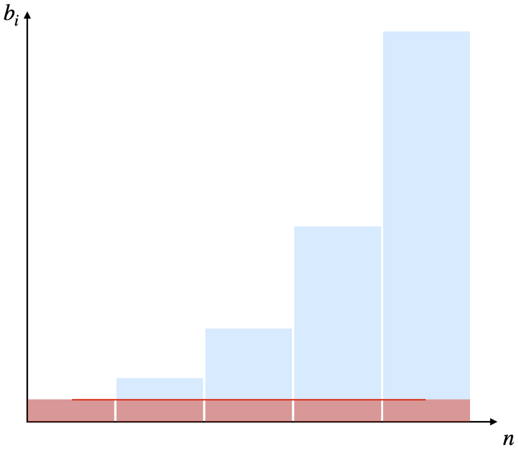

(a) PO (b) PtQ

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



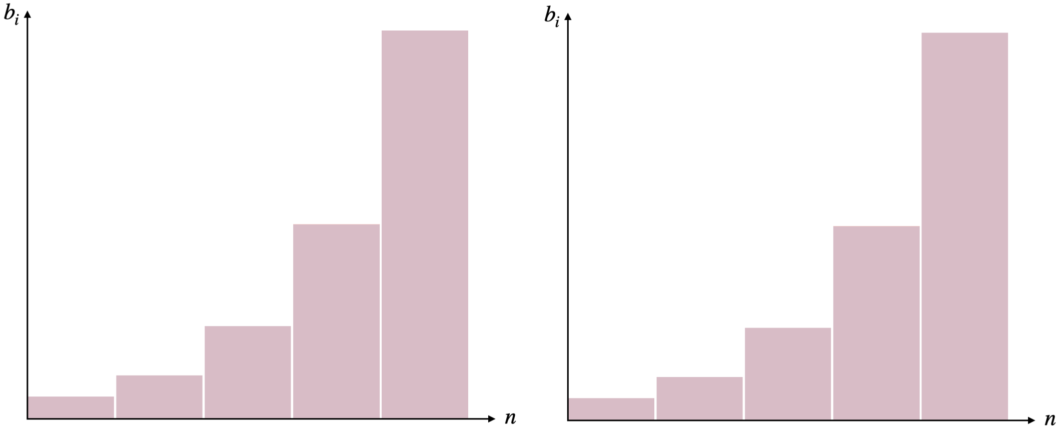

(a) PO




(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

# PtQ: Remarks

1. Participation becomes conclusive evidence for effort

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

2. Critical bonus depends on number of participants

   More participants $\rightarrow$ higher marginal productivity $\rightarrow$ lower bonus required
   Fewer participants $\rightarrow$ lower marginal productivity $\rightarrow$ higher bonus required

3. Technologically easier $\neq$ informationally easier

   Bonus increasing in $\Pr(\theta_j = H)$

4. **Expected cost** depends on the prior **p**

   Principal uses **p** to evaluate which rankings are cheap

5. **Implementation** does not

# General Results

Results generalize to $n$ experts, arbitrary strictly supermodular $P(H)$, heterogeneous $(c_i, p_i)$

1. **Optimal contract is PtQ** for every expert PtQ strictly dominates PO for every $\mathbf{p} \in (0,1)^n$
2. Under O-ring production $\left(P(H) = \prod_{i \notin H} \lambda_i\right)$, optimal contract is PtQ with ranking by a priority index:

$$I_i = \frac{c_i \, \lambda_i}{(1 - p_i)(1 - \lambda_i)^2}$$

   Rank by increasing $I_i$: cheap, likely-stable, and critical experts first

# 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

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

# 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

# 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!

# 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 Formula

This generalizes to any $n > 0$:

$$b_i^* = \frac{c_i}{P(A_1 \cup \{i\} \cup R^\sigma_{<i}) - P(A_1 \setminus \{i\} \cup R^\sigma_{<i})}$$

where $A_1(\mathbf{a}) = \{j \in N : a_j = 1\}$ are participants guaranteed $q_H$ 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 $b_i(0, \mathbf{a}_{-i}) = b_i^* + \varepsilon_{i,0}$.

$$1 < \frac{\varepsilon_{i,0}}{\varepsilon_{i,1}} < \frac{P(A_1 \cup \{i\} \cup R^\sigma_{<i})}{P(A_1 \setminus \{i\} \cup R^\sigma_{<i})} \qquad \forall\, \mathbf{a}_{-i}$$

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

By strict monotonicity, $\frac{P(A_1 \cup \{i\} \cup R^\sigma_{<i})}{P(A_1 \setminus \{i\} \cup R^\sigma_{<i})} > 1$ for all $\mathbf{a}_{-i}$. Such perturbations always exist

# Optimality of PtQ (Proof Sketch)

**Proposition:** An optimal UI contract is PtQ for some fixed $\sigma \in \mathrm{Perm}(N)$.

**Proof idea.** Start from any UI contract $\mathbf{b}$:

1. **Characterize:** Extract elimination times $\tau(i)$ and assured sets $S_i(\mathbf{a}_{-i})$.
2. **Fill in:** Choose ranking $\sigma$ refining $\tau$. Extend bonuses to all profiles $\Rightarrow$ same cost.
3. **Reduce:** Replace all bonuses with critical values along $\sigma \Rightarrow$ weakly lower cost.
4. **Switch to PtQ:** For each expert, switch from PO-form to PtQ-form. Each switch preserves UI and weakly reduces cost (enlarges predecessors' assured sets).
5. **Optimize:** $C(\mathbf{b}) \geq C^{PtQ}(\sigma, \mathbf{p}) \geq \min_{\sigma'} C^{PtQ}(\sigma', \mathbf{p})$.

# O-Ring: Priority Index

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

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

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

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

Under symmetric technology: $I_i \propto c_i/(1-p_i)$. Under symmetric costs and tech: rank by decreasing $p_i$.

# Common Prior and Bayesian NE

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

**PtQ under BNE achieves PI cost.**

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 at PI level while breaking all bad equilibria.

**PO under BNE:** Cost falls relative to BFR, but remains strictly above PI.

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

**Cost ranking:**

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

# 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.