

Pay to Quit*

Robust Team Incentives with Hidden Need

Hyungmin Park

Youngji Sohn

University of Warwick

University of Warwick

March 2026

Preliminary and Incomplete

Abstract

We study team incentives under complementarities, moral hazard, and adverse selection. Experts privately know whether their assigned components require costly repair. The principal observes only participation, not effort, and seeks to uniquely implement high quality under belief-free rationalizability (Bergemann and Morris, 2017). Standard contracts rewarding only participation fail to elicit private information, leaving the principal exposed to high coordination rents. We show that a “pay-to-quit” contract—which offers a higher success bonus for opting out than for participating—strictly reduces implementation costs. As the contract screens out experts who intend to shirk or whose components need no repair, participation becomes an on-path signal of active effort. Under complementarities, this endogenous information raises each expert’s marginal contribution, significantly lowering the bonuses required for unique implementation. We also show that, paradoxically, environments in which components are more likely to be already satisfactory can be more expensive to implement robustly, because intervention then becomes less informative.

JEL Code: D82, D86, L23

Keyword: Team production, Robust unique implementation, Complementarities, Moral hazard and adverse selection, Pay-to-quit contracts

*For helpful comments, we thank Costas Cavounidis, Daniele Condorelli, Alkis Georgiadis-Harris, Jacob Glazer, Sinem Hidir, Ilan Kremer, Motty Perry, Ferdinand Pieroth, Herakles Polemarchakis, Daniel Rappoport, Francesco Squintani, and participants in MWIP seminar at the University of Warwick and 15th Lisbon Meetings in Game Theory and Applications. Nevertheless, we absolve them from responsibility for any errors in the paper and the opinions expressed herein.

1 Introduction

Many team production problems arise not at the stage of creating new output, but at the stage of repairing, certifying, or stabilizing an existing project before release.¹ At this stage, each component is assigned to a specialist, and components differ in their state of readiness. Some contain genuine defects that require costly effort to fix, while others are already satisfactory and do not benefit from further intervention. Yet managers often observe only who joins the remediation effort, not which components actually needed attention or how much effort each specialist exerted.

A natural response is to reward participation and pay nothing to those who stand aside. This seems sensible, but it creates a problem: when participation is rewarded, everyone has an incentive to join, and the participation decision ceases to convey any information about the underlying need. What makes these environments particularly difficult is complementarity: the value of repairing one component is higher when more of the other components are already in good shape. In highly complementary projects, this loss of information is costly. Each specialist who truly needs to work must be compensated for the risk that others are merely present rather than actually resolving defects.

This paper studies the design of bonus contracts in team environments with both adverse selection and moral hazard. A principal assigns a project with n parts to n experts. Each expert privately learns whether her part is already of high quality or instead has a defect that can be repaired only through costly effort. The expert then chooses whether to participate, and a participant privately chooses whether to work or shirk. The participation profile is publicly observable, and hence contractible. However, neither the initial quality of a part nor the expert's effort is contractible. Project success depends on the set of parts that end up at high quality, and this success probability is strictly monotone and supermodular. The principal seeks to implement high quality for every part at minimum expected cost.

A useful benchmark clarifies where the difficulty comes from. If every part always required repair, rewarding intervention would create no screening conflict: anyone who intervenes would, by construction, be someone whose part should be repaired. The participation margin would then serve simply as the observable intervention decision through which the principal motivates effort. Our problem begins once experts privately know whether intervention is needed. Under a standard success bonus for intervention, both experts who should repair and experts whose parts are already satisfactory are drawn into the intervention margin, so observed intervention no longer distinguishes active re-

¹Prominent real-world examples include pre-release software quality assurance, where each module is assigned to a specialist who must privately identify and resolve defects; pre-certification reviews in aerospace manufacturing, where subsystems undergo independent expert inspection; and financial audit remediation, where each division's discrepancies must be resolved before a consolidated report can be filed.

pair from unnecessary presence.

This creates a basic tension. From a technological standpoint, a higher probability that a part is already of high quality is favorable to the principal, because fewer repairs are socially needed. From an incentive standpoint, however, the same possibility makes observable intervention less informative: an expert who intervenes may simply be collecting a participation reward rather than signaling that repair is actually needed. Under robust implementation, this informational effect can dominate. Easier fundamentals need not lower implementation costs, and may even raise them. A central role of pay-to-quit is to manage precisely this tension by restoring the informational content of intervention.

Our starting point is a simple observation: when the principal only needs high quality to be sustained as one possible outcome, private information about initial quality is of little use. In the partial-implementation benchmark, inducing type separation through participation is wasteful, and the optimal contract reduces to a standard success bonus for participants. The resulting cost is independent of the type distribution. The difficulty, however, is that this benchmark leaves the principal exposed to coordination failure. Under complementarities, pessimistic beliefs about others' types and behavior can make shirking rationalizable even when working is also rationalizable. Once the principal instead requires unique implementation that is robust to these beliefs, the problem changes fundamentally: she must design a contract under which working is the only rationalizable action for every low-quality expert, regardless of arbitrary beliefs about others' types and actions.²

In this regard, we adopt *belief-free rationalizability* (Bergemann and Morris, 2017) in our implementation. This allows us to study implementation that is robust not only to strategic uncertainty about what others will do, but also to informational uncertainty about what others believe. The principal's contract does not rely on common priors, consistent beliefs, or equilibrium selection. In this sense, the paper delivers implementation without beliefs and optimization with beliefs: robustness of behavior is obtained without imposing a belief structure on the agents, while the principal still uses the type distribution to minimize the expected implementation cost.

We first show that if the principal restricts attention to *participation-only* contracts, opting out is weakly dominated, so all types participate and the participation profile carries no information about which parts actually require repair. Because unique implementation is evaluated under belief-free rationalizability, the contract must then induce each low-quality expert to work under arbitrary beliefs about the status of others' parts. In particular, it must do so in the worst case in which every unresolved part may belong to a low-quality type whose effort is not yet pinned down. The key consequence

²This concern reflects the famous "Wilson doctrine" (Wilson, 1987): game-theoretic solutions should be robust to the fine details of players' beliefs, so adopting a belief-free solution concept ensures that the principal's contract performs well regardless of what experts believe about one another.

is not that our environment literally collapses to pure moral-hazard model by [Winter \(2004\)](#), but that private information destroys the simple use of participation as a contractible intervention margin. Once participation has become uninformative, [Winter's \(2004\)](#) divide-and-conquer logic becomes the relevant tool for economizing on the large bonuses needed for robust implementation.

Our main contribution is to show that the principal can do strictly better by reversing the usual logic of remediation pay. We study a *pay-to-quit* contract that offers a higher success-contingent bonus for opting out than for participating. In our setting, this is not a device for labor turnover; it is a screening device. High-quality experts, whose parts need no improvement, strictly prefer to opt out. Moreover, for low-quality experts, participating and shirking is also dominated by opting out: both actions leave the part unresolved, but opting out pays more. After one round of elimination, any expert who still participates must therefore be a low-quality type who works. Participation becomes an on-path signal that a defect is being actively repaired.

This change in the informational role of participation has a first-order implication for implementation. Once participation identifies repaired components, each expert can evaluate her own incentive to work against a larger publicly assured set of high-quality parts. Under supermodularity, this raises the marginal value of her effort and lowers the bonus needed to induce it. The key insight is that adverse selection, rather than being a friction to be neutralized, can be turned into a tool for reducing the cost of moral-hazard implementation. We show that, within our belief-free robust implementation problem, *pay-to-quit* achieves unique implementation at a strictly lower expected cost than participation-only contracts. We then characterize the principal's full design problem: the optimal uniquely implementing contract assigns *pay-to-quit* incentives at every expert-contingency pair, reducing the remaining problem to the choice of an optimal ranking. Under an O-ring technology, this ranking admits a nice closed-form priority index.

Taken together, our results highlight a broader lesson. A technologically more forgiving environment need not be easier to manage. When intervention is observable but the need for intervention is private, a higher probability that components are already satisfactory weakens the meaning of observed intervention and can raise the cost of robust implementation. *Pay-to-quit* is valuable precisely because it converts this hidden slack into usable information.

Our paper contributes to three strands of literature. First, it introduces an adverse-selection dimension to the literature on unique implementation and team incentives under complementarity. Relative to [Winter \(2004\)](#), the key novelty in our model is that experts privately know whether their parts actually require work, and the principal can exploit this through endogenous type separation. Second, our paper differs from recent approaches that mitigate strategic uncertainty through exogenous information channels,

such as monitoring, information design, or sequential contract choice. Instead, our mechanism makes the participation decision itself endogenously informative on path. Third, we highlight a novel economic role for pay-to-quit contracts: rather than managing turnover, they serve as a static screening instrument that resolves coordination failure by extracting private information about the need for effort.

The rest of the paper proceeds as follows. Section 2 presents the model. Section 3 studies partial implementation as a benchmark. Section 4 analyzes participation-only contracts and shows how the problem reduces to pure moral hazard in teams. Section 5 introduces pay-to-quit contracts and establishes their strict cost advantage. Section 6 characterizes the optimal uniquely implementing contract and derives a closed-form ranking rule under an O-ring technology. Section 7 discusses how the analysis changes under a common prior, with sequential effort choice and interim ranking, and under an alternative intervention technology. All proofs are relegated to the appendix.

Related Literature. This paper contributes to the literature on unique implementation in teams with moral hazard and complementarities. Our point of departure differs from Winter (2004) in an important respect. In our environment, if every component always required repair, the observable participation margin would already provide a straightforward route to robust implementation. The distinctive friction we study is that experts privately know whether repair is needed, which under standard reward-for-participation contracts makes participation uninformative. Once this happens, Winter’s divide-and-conquer logic becomes the relevant tool for economizing on robust incentives. Halac et al. (2021) show that, with privately offered contracts, rank uncertainty can substitute for explicit discrimination. The paper is also related to the broader divide-and-conquer tradition in contracting with externalities, going back to Segal (1999, 2003) and extended by Bernstein and Winter (2012) and Halac et al. (2020). Relative to this literature, our contribution is not merely to add adverse selection to a team moral-hazard problem, but to show that private information about the need for effort determines whether a coarse contractible action can be used to secure robust implementation at all. More broadly, the paper contributes to an emerging agenda on informationally robust organizational design, in which incentives must be structured not only to elicit effort but also to extract information about where effort is needed.

Several extensions of the baseline framework incorporate richer organizational assumptions. Winter (2006) and Winter (2010) study sequential production where workers observe predecessors’ effort; Gershkov and Winter (2015) analyze the interaction between formal and informal peer monitoring; and Lu and Song (2025) study network-based peer monitoring design. Moriya and Yamashita (2020) examine optimal information allocation in team production and show that asymmetrically informing even ex ante symmetric agents can be optimal for unique implementation. Camboni and Porcellacchia (2025)

study moral hazard in teams where the principal can use both a joint team outcome and imperfect individual performance signals to incentivize effort. More closely related to our work, [Park and Song \(2025\)](#) shows that costly and verifiable peer monitoring, though used only off-path, can eliminate all strategic uncertainty in teams, and [Cavounidis and Park \(2025\)](#) show that a subcontractor’s contract choice can reveal her intention to work through forward induction, reducing the cost of inducing effort. Our paper differs from these approaches in how information is generated: we use an opt-out decision that endogenously generates on-path type separation.

Our mechanism is also related in form to pay-to-quit and voluntary-separation contracts. [Lazear \(1979, 1983\)](#) studies deferred compensation and severance-like arrangements that discipline workers or induce efficient exit, [Au \(2020\)](#) analyzes pay-to-quit in repeated team settings where it supports dynamic cooperation through peer monitoring, and [Foarta and Sugaya \(2021\)](#) study how a menu of contracts can screen employees by effort cost, showing that inducing high-cost types to self-select out of demanding projects can be optimal. In contrast, our pay-to-quit contract operates in a team setting with complementarities, where screening serves not only to align participation with productive need but also to reduce the coordination rents required for robust implementation.

Finally, our paper connects to the literature on informationally robust mechanism design. We adopt belief-free rationalizability in the sense of [Bergemann and Morris \(2017\)](#), building on [Bergemann and Morris \(2005, 2017\)](#), to study unique implementation without relying on a common prior or on consistent higher-order beliefs. This solution concept is especially natural in our environment, where experts may know little about the quality of other parts and may hold heterogeneous beliefs about others’ beliefs.

2 Model

A principal assigns a project consisting of n distinct parts to a team of experts $N = \{1, \dots, n\}$, with expert i responsible for part i . Each part is initially in one of two quality states, $\theta_i \in \{H, L\}$: high quality (H) or low quality (L). A high-quality part is already satisfactory and requires no intervention; a low-quality part has a defect or gap that can be remedied through intervention. The initial quality of each part is drawn independently across parts, with $\Pr(\theta_i = H) = p_i \in (0, 1)$, and we write $\mathbf{p} = (p_1, \dots, p_n)$. The distribution \mathbf{p} need not be commonly known among agents.³

Each expert i privately observes her type θ_i and decides whether to participate, $a_i \in \{0, 1\}$, where $a_i = 1$ means undertaking the task and $a_i = 0$ means leaving it untouched. When choosing to participate, expert i selects an effort level $e_i \in \{W, S\}$, where W denotes working (high effort) at personal cost $c_i > 0$ and S denotes shirking (low effort)

³Our notion of implementation does not rely on experts knowing \mathbf{p} or holding common beliefs about others’ types. The principal’s knowledge of \mathbf{p} is used only for evaluating the cost of implementation.

at zero cost. If $a_i = 0$, the effort choice is moot; we set $e_i = S$ by convention. The participation profile $\mathbf{a} = (a_1, \dots, a_n)$ is observable to all experts and the principal, while effort choices are private. Each expert's action set is thus $\mathcal{A}_i = \{(1, W), (1, S), (0, S)\}$.⁴

The final quality of each part depends on the expert's type and effort through a technology function $q : \{H, L\} \times \{W, S\} \rightarrow \{q_H, q_L\}$ such that

$$q(\theta_i, e_i) = \begin{cases} q_H & \text{if } \theta_i = H, \text{ or } \theta_i = L \text{ and } e_i = W, \\ q_L & \text{if } \theta_i = L \text{ and } e_i = S. \end{cases} \quad (1)$$

That is, an initially high-quality part remains high quality regardless of effort, while an initially low-quality part is improved to q_H if and only if the responsible expert works. We call this a *fault-repair* technology.⁵

The project's success depends on which parts achieve high quality. Let $Q \subseteq N$ denote the set of parts that attain high quality, i.e., $Q(\boldsymbol{\theta}, \mathbf{e}) := \{i \in N : q(\theta_i, e_i) = q_H\}$ for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and $\mathbf{e} = (e_1, \dots, e_n)$. We model the probability of project success as a function $P : 2^N \rightarrow [0, 1]$, where $P(Q)$ gives the success probability when exactly the parts in Q achieve high quality. The function P exhibits *strict monotonicity* and *strict supermodularity*, which we formalize as follows.

Assumption 1 (Complementarities in success). The function $P : 2^N \rightarrow [0, 1]$ satisfies:

1. **Strict monotonicity:** $Q' \subsetneq Q''$ implies $P(Q') < P(Q'')$.
2. **Strict supermodularity:** For any $Q' \not\subseteq Q''$ and $Q'' \not\subseteq Q'$,

$$P(Q' \cup Q'') - P(Q') > P(Q'') - P(Q' \cap Q''). \quad (2)$$

Monotonicity means that more high-quality parts always improve the chances of success. Supermodularity captures complementarities in quality: the marginal contribution of improving one additional part is greater when more other parts are already of high quality.

The principal offers each expert a contract specifying a bonus $b_i(\mathbf{a}) \geq 0$ contingent on the participation profile $\mathbf{a} = (a_1, \dots, a_n)$ and paid upon project success; no payment is made in the event of failure.⁶ Only the participation profile and the project outcome are contractible; the realized quality of each part and the effort choices are not.

⁴Participation serves as a partial verification of effort choice. This distinction is central to our analysis: the principal can condition bonuses on the observable margin — who participates — but not on how hard they work. Since the participation profile is observed only after all decisions are made, the game is equivalent to a simultaneous choice among these three actions. In Section 7, we discuss the sequential variant in which experts observe the participation profile before choosing effort.

⁵We also discuss an alternative *fragile intervention* technology in Section 7.3, under which participating without exerting effort degrades an initially high-quality part. This captures tasks requiring high precision, such as medical surgery or sensitive hardware calibration, where merely tampering with a functioning system without costly care causes new damage.

⁶It is without loss to consider only success-contingent bonuses: since $P(Q) > 0$ for every Q , shifting any

Let $\mathbf{b} = \{b_i(\mathbf{a})\}_{i \in N, \mathbf{a} \in \{0,1\}^n}$ denote the complete bonus schedule. Given \mathbf{b} , each expert i observes her type θ_i and chooses an action $(a_i, e_i) \in \mathcal{A}_i$. Given a type profile $\boldsymbol{\theta}$ and action profile (\mathbf{a}, \mathbf{e}) , expert i 's payoff is

$$u_i(\boldsymbol{\theta}, \mathbf{a}, \mathbf{e}) = P(Q(\boldsymbol{\theta}, \mathbf{e})) \cdot b_i(\mathbf{a}) - \mathbb{1}\{e_i = W\} \cdot c_i. \quad (3)$$

The principal's expected cost of the bonus schedule \mathbf{b} when every part achieves high quality is

$$C(\mathbf{b}) = P(N) \cdot \mathbb{E}_{\mathbf{p}} \left[\sum_{i \in N} b_i(\mathbf{a}(\boldsymbol{\theta})) \right], \quad (4)$$

where $\mathbf{a}(\boldsymbol{\theta})$ denotes the participation profile at the desired outcome.⁷ The principal seeks to maximize the probability of success, that is, every part attains q_H regardless of the realized type profile, at minimum expected cost.

Due to complementarities in quality, bonus schedules that make the desired outcome incentive compatible also sustain equilibria in which some parts attain q_L . Beyond this strategic uncertainty, in practice each expert's knowledge about other parts is likely limited—experts may hold heterogeneous beliefs about \mathbf{p} .

To address both concerns, we adopt *belief-free rationalizability* (BFR) (Bergemann and Morris, 2017) as our solution concept. BFR is an iterative elimination procedure in which an action survives for a given type if and only if it is a best response to *some* belief over opponents' surviving type-action pairs. Since these beliefs need not be consistent across experts or types, the solution concept is well-defined even when experts do not share a (common) prior over types.⁸

We say that a bonus schedule \mathbf{b} *uniquely implements* high quality (UI) if there exists a sequence of perturbations $\{\boldsymbol{\varepsilon}^k\}_{k \geq 1}$, where $\boldsymbol{\varepsilon}^k = (\varepsilon_i^k(\mathbf{a}))_{i \in N, \mathbf{a} \in \{0,1\}^n}$ converge component-wise to zero, such that under each perturbed bonus schedule $\mathbf{b} + \boldsymbol{\varepsilon}^k$, every belief-free rationalizable action for each expert $i \in N$ with any type $\theta_i \in \{H, L\}$ yields $q(\theta_i, e_i) = q_H$.⁹ The principal then solves

$$\min_{\mathbf{b} \text{ UI}} C(\mathbf{b}). \quad (5)$$

failure payment to success weakly relaxes incentives and never increases cost. The constraint $b_i(\mathbf{a}) \geq 0$ reflects *limited liability*: experts cannot be required to make positive transfers to the principal. Note also that the bonus to each expert depends on the entire participation profile, not merely on her own action. The contract is thus multilateral in nature.

⁷The game induced by \mathbf{b} may admit multiple outcomes with $Q(\boldsymbol{\theta}, \mathbf{e}) = N$ only if type H 's action is not pinned down. However, as shown in the proof of Proposition 4, this multiplicity is absent at any optimal contract.

⁸The formal definition of BFR is provided in Appendix A.

⁹The perturbation serves as a tie-breaking rule in favor of the desired action. Without the perturbation sequence, the set of contracts under which all high-quality is the *unique* belief-free rationalizable outcome would be open: at the boundary, exact indifferences allow an additional action that induces q_L to survive. Including the limiting argument in the definition ensures that such boundary contracts are also classified as uniquely implementing.

We call any minimizer an *optimal UI bonus schedule*.

The key issue is therefore not hidden effort alone, but the interaction between hidden effort and hidden need for intervention. When every part requires repair, rewarding intervention does not distort who enters the task: intervention and productive need are aligned. Once experts privately know whether intervention is needed, the same reward must perform two conflicting roles—it must motivate repair and at the same time avoid attracting experts whose parts are already satisfactory. As we shall see, under standard intervention-contingent pay, it fails at the second task, and the intervention margin loses its value as evidence of productive repair.

Remark 1 (Robustness in Implementation). A contract that uniquely implements high quality under BFR also does so under more restrictive solution concepts. The set of BFR-surviving actions contains the set of actions surviving interim correlated rationalizability for any type space, which in turn contains the set of Bayesian Nash equilibrium actions for any common prior. The unique BFR prediction therefore remains the unique prediction under these solution concepts as well. The principal thus obtains an implementation that is immune to both strategic uncertainty and informational uncertainty without relying on common knowledge of beliefs or equilibrium coordination.

3 Benchmark: Partial Implementation

Before solving the principal’s problem in (5), we consider *partial implementation*, a relaxed benchmark that drops the uniqueness requirement. We say a bonus schedule \mathbf{b} *partially implements* high quality (PI) if for each player, there exists a belief under which achieving q_H for her part is a best response in the game induced by \mathbf{b} .¹⁰

In the fault-repair technology, partial implementation reduces to a condition on L -types alone, since H -types’ actions have no effect on quality. Specifically, participating and working must be belief-free rationalizable for every expert with $\theta_i = L$.

To verify that a given bonus schedule partially implements high quality, it suffices to show that, for each expert with type L , $(1, W)$ is a best response under the belief that all other experts participate and all other L -types work, and that this belief is indeed supported on surviving type–action profiles. Such a contract is characterized by the incentive constraints that $(1, W)$ is weakly preferred to both shirking and opting out for L -types, given that all other parts achieve high quality:

$$P(N) \cdot b_i(1, \mathbf{a}_{-i}) - c_i \geq \max \{ P(N \setminus \{i\}) \cdot b_i(1, \mathbf{a}_{-i}), P(N \setminus \{i\}) \cdot b_i(0, \mathbf{a}_{-i}) \}. \quad (6)$$

¹⁰Relative to the unique implementation problem in (5), partial implementation drops the requirement that the high-quality outcome be the *only* belief-free rationalizable outcome; it asks only that the high-quality outcome be *among* the belief-free rationalizable outcomes.

We next ask whether the principal benefits from inducing type separation strictly—setting $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$ so that H -types opt out while L -types participate. Given that (6) holds for all $i \in N$ and all L -types choose $(1, W)$, type separation only increases the total expected bonus payment. The principal therefore optimally sets $b_i(0, \mathbf{a}_{-i}) \leq b_i(1, \mathbf{a}_{-i})$. Under this condition, the value of $b_i(0, \mathbf{a}_{-i})$ does not affect the expected cost, and we set $b_i(0, \mathbf{a}_{-i}) = 0$ without loss of generality.

Proposition 1 (Optimal partial implementation). *An optimal PI bonus schedule sets*

$$b_i^{\text{PI}}(1, \mathbf{a}_{-i}) = \frac{c_i}{P(N) - P(N \setminus \{i\})}, \quad b_i^{\text{PI}}(0, \mathbf{a}_{-i}) = 0 \quad (7)$$

for all $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$. The expected cost is

$$C^{\text{PI}} = P(N) \cdot \sum_{i \in N} \frac{c_i}{P(N) - P(N \setminus \{i\})}, \quad (8)$$

which is independent of \mathbf{p} .

This optimal partial implementation means the lower bound of cost in implementing all parts with high quality as a BFR action profile. Each expert's bonus equals her effort cost divided by her marginal contribution to success, $P(N) - P(N \setminus \{i\})$, and does not depend on others' participation decisions. Since no type separation occurs, the cost is independent of \mathbf{p} . This implies that private information about types is entirely benign when uniqueness is not required, effectively reducing the principal's challenge to a standard moral hazard problem.

The optimal PI contract, however, leaves the principal exposed to coordination failure. Working is rationalizable for L -types under the belief that all other L -types also work, but shirking is also rationalizable under a different belief. If expert i believes that some other L -types shirk, complementarity implies that her marginal contribution to success falls below $P(N) - P(N \setminus \{i\})$, so the bonus no longer covers her effort cost and shirking becomes a best response. Eliminating such beliefs requires raising bonuses beyond the optimal PI bonus. This motivates the analysis of unique implementation, where contracts are designed so that $(1, W)$ is the *only* rationalizable action for every L -type expert.

4 Participation-Only Contracts

We now turn to the principal's unique implementation problem. The partial implementation analysis showed that type separation through participation is wasteful when uniqueness is not required. Motivated by this observation, we begin by restricting attention to *participation-only* contracts, which promise a bonus only to participants upon project success.

Definition 1 (Participation-only contracts). A contract \mathbf{b} is *participation-only* (PO) if $b_i(0, \mathbf{a}_{-i}) = 0$ for all $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$.

Under a PO contract, a positive bonus is paid only to participants, so opting out is dominated by participating and shirking; every expert can participate regardless of her type. Participation therefore reveals no information about whether intervention is actually needed. Once both H - and L -types intervene, observed intervention no longer separates productive repair from unnecessary presence. The principal must therefore treat intervention as an uninformative observable and use the bonus schedule only to eliminate shirking among those experts who actually need to work.

Given that participation is now uninformative, the principal faces a coordination problem: under complementarities, each expert's incentive to work depends on whether others also work, and pessimistic beliefs can sustain a self-fulfilling equilibrium in which everyone shirks. The key insight, following Winter (2004), is that the principal can break this circularity by structuring bonuses asymmetrically across experts, so that working becomes the only rationalizable action *iteratively*. The idea is to make one expert's bonus large enough that she works regardless of what others do. Once her action is pinned down, the next expert faces a less uncertain environment—one part is already guaranteed to be of high quality—so her required bonus is lower. Proceeding in sequence, the principal “divides” the coordination problem and “conquers” each expert's incentive constraint one at a time. The order in which experts are resolved—the *ranking*—determines how quickly the assured set grows and hence the total cost.

Fix a permutation $\sigma \in \text{Perm}(N)$ representing the order in which experts' incentive constraints are resolved, where $\sigma(i)$ denotes expert i 's ranking. Define

$$R_{\leq i}^\sigma := \{j \in N : \sigma(j) \leq \sigma(i)\}, \quad R_{< i}^\sigma := \{j \in N : \sigma(j) < \sigma(i)\},$$

with $R_{< \sigma^{-1}(1)}^\sigma = \emptyset$. Define the *critical bonus* for expert i as

$$b_i^{\text{PO}}(\sigma) := \frac{c_i}{P(R_{\leq i}^\sigma) - P(R_{< i}^\sigma)}. \quad (9)$$

This is the bonus at which expert i is exactly indifferent between working and shirking, given a belief that all experts ranked before her work and no expert ranked after her does. Setting each expert's bonus strictly above her critical bonus level along the ranking σ ensures that the contract achieves unique implementation, as the following proposition formalizes.

Proposition 2 (Unique implementation under PO). *Fix a ranking σ . Any PO contract with $b_i(1, \mathbf{a}_{-i}) > b_i^{\text{PO}}(\sigma)$ and $b_i(0, \mathbf{a}_{-i}) = 0$ for all $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$ uniquely implements high quality: $(1, W)$ is the only belief-free rationalizable action for every expert with $\theta_i = L$.*

The ranking σ determines the bonus schedule and hence the expected cost. Since all experts participate and the project succeeds with probability $P(N)$ when every part achieves q_H , the expected cost under ranking σ is

$$C^{\text{PO}}(\sigma) = P(N) \cdot \sum_{i \in N} \frac{c_i}{P(R_{\leq i}^\sigma) - P(R_{< i}^\sigma)}. \quad (10)$$

As under partial implementation, the cost is independent of \mathbf{p} , since the contract must induce working as the only rationalizable action under every belief about others' types.

Remark 2. The cost of unique implementation under any PO contract is strictly higher than the cost of partial implementation. Under PI, each expert's bonus is scaled by $P(N) - P(N \setminus \{i\})$, the marginal contribution evaluated when all other parts are of high quality. Under PO, expert i has her bonus scaled by $P(R_{\leq i}^\sigma) - P(R_{< i}^\sigma)$, the marginal contribution evaluated at a smaller set. Supermodularity implies that the latter is strictly smaller for every expert except the last-ranked, so $b_i^{\text{PO}}(\sigma) > c_i/[P(N) - P(N \setminus \{i\})]$ for all i with $\sigma(i) < n$.

5 Pay-to-Quit Contracts

Under participation-only contracts, all experts participate regardless of type, so the participation profile cannot serve as a screening device for underlying quality states. The principal must therefore provide bonuses large enough to make working the only rationalizable action even in the worst case in which every expert is of type L . We now ask whether allowing experts to quit for a positive bonus can reduce the cost of unique implementation by making participation decisions informative.

Definition 2 (Pay-to-quit contracts). A contract \mathbf{b} is *pay-to-quit* (PtQ) if $b_i(0, \mathbf{a}_{-i}) > 0$ for all $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$.

PtQ changes the implementation problem only if participation decisions become type-dependent. To see this, consider an H -type expert i choosing between participation and non-participation. Since $\theta_i = H$, part i attains q_H regardless of effort, so opting out and participating without working induce the same quality outcome. An H -type therefore chooses whichever action yields the higher bonus. To obtain separation, we thus restrict attention to contracts that induce H -types to opt out: for all i and \mathbf{a}_{-i} ,

$$b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i}). \quad (11)$$

Under (11), participating and shirking is also strictly dominated by opting out for an L -type, since both actions leave part i at q_L , while opting out yields the higher bonus.

After the first round of iterated elimination, the surviving actions are $\{(0, S)\}$ for H -types and $\{(0, S), (1, W)\}$ for L -types. For each expert i and participation profile of the others \mathbf{a}_{-i} , let

$$W_{-i}(\mathbf{a}_{-i}) := \{j \in N \setminus \{i\} : a_j = 1\}$$

denote the set of other experts who participate and hence, by the elimination above, work. The nesting of surviving action sets, $\{(0, S)\} \subset \{(0, S), (1, W)\}$, implies that participation partially verifies types as well as effort: if $a_j = 1$, then expert j must be an L -type choosing $(1, W)$, which guarantees q_H for part j . If $a_j = 0$, the observed action is consistent with both types, so the quality of part j remains uncertain. It follows that, for any expert i , every part in $W_{-i}(\mathbf{a}_{-i})$ achieves q_H regardless of her beliefs about others' types and actions.¹¹

After the first round of elimination, what remains for unique implementation is to choose bonus levels such that $(1, W)$ strictly dominates $(0, S)$ for every L -type expert. Expert i does not observe \mathbf{a}_{-i} when choosing her action, so this comparison depends on her belief about which other parts achieve q_H . If the bonus schedule makes $(1, W)$ strictly dominate $(0, S)$ for every realization of \mathbf{a}_{-i} , however, then the same conclusion follows under any belief.

Fix a realization \mathbf{a}_{-i} , and let $Q_{-i}(\mathbf{a}_{-i}) \subseteq N \setminus \{i\}$ denote a rationalizable set of other parts that expert i regards as achieving q_H . Since $a_j = 1$ is consistent only with $(1, W)$ after the first round, we must have $W_{-i}(\mathbf{a}_{-i}) \subseteq Q_{-i}(\mathbf{a}_{-i})$. Then $(1, W)$ strictly dominates $(0, S)$ for an L -type if

$$b_i(1, \mathbf{a}_{-i}) \cdot P(Q_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - c_i > b_i(0, \mathbf{a}_{-i}) \cdot P(Q_{-i}(\mathbf{a}_{-i})) \quad (12)$$

for every rationalizable $Q_{-i}(\mathbf{a}_{-i}) \supseteq W_{-i}(\mathbf{a}_{-i})$ and every realization $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$. Because type separation requires only that $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$, the two bonuses can be chosen arbitrarily close. The corresponding infimum threshold is therefore characterized by

$$b_i(1, \mathbf{a}_{-i}) [P(W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(W_{-i}(\mathbf{a}_{-i}))] > c_i. \quad (13)$$

By supermodularity, the tightest constraint arises at the smallest feasible assured set. As more experts opt out, this base shrinks, expert i 's marginal contribution falls, and the bonus must rise to compensate.

The principal could directly choose bonuses to satisfy (13) for all experts, taking $Q_{-i}(\mathbf{a}_{-i}) = W_{-i}(\mathbf{a}_{-i})$ for each i . But the implementation cost can be reduced further by exploiting the same recursive logic of iterated elimination as in Section 4. Once the contract guarantees that part i reaches q_H regardless of expert i 's type, every other expert

¹¹The nesting of surviving action sets across types is reminiscent of the nested range condition in [Green and Laffont \(1986\)](#). A key difference is that, in our setting, the nesting arises endogenously from contract design and iterated elimination rather than being a primitive feature of the evidence structure.

j can condition on this, so Q_{-j} expands to include i , lowering the bonus required for j . We therefore construct the PtQ contract inductively along a ranking $\sigma \in \text{Perm}(N)$.

At round $\sigma(i) + 1$ of the elimination process, the set of parts that expert i regards as assured to be high quality, $Q_{-i}(\mathbf{a}_{-i})$, includes

$$R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}). \quad (14)$$

Parts in $R_{<i}^\sigma$ are assured unconditionally: earlier rounds have pinned down q_H for each such part regardless of its expert's type. Parts in $W_{-i}(\mathbf{a}_{-i})$ are assured because participation is consistent only with $(1, W)$ after the first round of elimination. Substituting (14) into (13), the critical PtQ bonus for expert i at rank $\sigma(i)$ is

$$b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) := \frac{c_i}{P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}))}. \quad (15)$$

Now fix any $\varepsilon_0 > \varepsilon_1 > 0$, and define a perturbed PtQ contract by

$$b_i^\varepsilon(1, \mathbf{a}_{-i}) = b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_1, \quad b_i^\varepsilon(0, \mathbf{a}_{-i}) = b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_0.$$

Under this perturbed contract, the elimination proceeds along the ranking: at round $\sigma(i) + 1$, $(1, W)$ strictly dominates $(0, S)$ for L -type expert i , so $(0, S)$ is eliminated. Expert i 's part then enters Q_{-j} for every j ranked after her, and the process continues. After $n + 1$ rounds, $(1, W)$ is the only surviving action for every L -type expert.¹²

Proposition 3 (Unique implementation under PtQ). *For any ranking $\sigma \in \text{Perm}(N)$, the PtQ contract defined by $b_i(\mathbf{a}) = b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i})$ for all $i \in N$ and $\mathbf{a} \in \{0, 1\}^n$ uniquely implements high quality.*

For a fixed ranking σ , the expected implementation cost of the PtQ contract is

$$C^{\text{PtQ}}(\sigma, \mathbf{p}) = P(N) \cdot \mathbb{E}_{\mathbf{p}} \left[\sum_{i \in N} \frac{c_i}{P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}))} \right], \quad (16)$$

where the expectation is over type profiles $\boldsymbol{\theta}$ drawn from \mathbf{p} , which under the PtQ construction induce the participation profile through $a_j = \mathbb{1}\{\theta_j = L\}$. Unlike the PO cost, the PtQ cost depends on \mathbf{p} : when L -types are prevalent, more experts participate, enlarging $W_{-i}(\mathbf{a}_{-i})$ and thereby increasing the relevant marginal contributions by supermodularity, so the required bonuses fall. Conversely, when most parts are already of high quality, fewer experts participate, the assured set shrinks, and the required bonuses rise.

¹²The construction requires $n + 1$ rounds of iterated elimination—one for type separation and n for the inductive elimination along the ranking. Equivalently, the argument relies on $(n + 1)$ -th order rationality among the experts.

This dependence reveals a basic tension between technology and incentives. A higher p_j is technologically favorable, because part j is then less likely to require repair. But it is also informationally harmful: expert j is less likely to participate, so her participation decision is less informative on path about whether part j attains high quality. The resulting contraction of assured sets lowers other experts' marginal contributions and raises the bonuses needed to induce them. In this sense, components that are more likely to be already satisfactory can be more expensive to manage under robust implementation.

Corollary 1 (Technologically easier, informationally harder). *Fix a ranking σ . Under the pay-to-quit contract, $C^{\text{PtQ}}(\sigma, \mathbf{p})$ is weakly increasing in each p_j , and strictly increasing for every expert j with $\sigma(j) > 1$.*

We now compare the costs under the PtQ and the PO contracts. Under PO, participation reveals no information about types, so the assured set for expert i at round $\sigma(i) + 1$ of elimination consists only of $R_{<i}^\sigma$. Under PtQ, by contrast, participation is consistent only with $(1, W)$ after the first round, so observed participation provides an additional source of assurance. In particular, every participating opponent enters expert i 's assured set, and hence $Q_{-i}(\mathbf{a}_{-i}) = R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i})$. Relative to PO, PtQ therefore weakly enlarges the assured set in every contingency and strictly enlarges it whenever some opponent participates.

The cost reduction is driven by an *externality effect*. When expert j is type L and participates under PtQ, her participation enlarges the assured sets of other experts and thereby strictly lowers their required bonuses by supermodularity. Since $p_j \in (0, 1)$ for all j , such participation occurs with positive probability, so this externality reduces the expected implementation cost.

Corollary 2 (PtQ is strictly cheaper than PO). *Fix a ranking σ . Then*

$$C^{\text{PtQ}}(\sigma, \mathbf{p}) < C^{\text{PO}}(\sigma)$$

for every $\mathbf{p} \in (0, 1)^n$.

6 Optimal Contracts

The previous sections compared two polar contract forms under a fixed elimination ranking. While participation-only (PO) contracts leave the act of participation uninformative, pay-to-quit (PtQ) contracts transform it into an informative, on-path signal of the expert's type. We now turn to the principal's full design problem and ask what an *optimal* uniquely implementing contract looks like when the principal is free to tailor the incentive structure arbitrarily.

The principal's design problem is to choose, for each expert i and each observable participation profile of others \mathbf{a}_{-i} , whether to locally reward participation or non-participation more. Thus, at any expert-contingency pair (i, \mathbf{a}_{-i}) , the principal can structure the local bonuses to mimic either a PO format, $b_i(0, \mathbf{a}_{-i}) = 0 < b_i(1, \mathbf{a}_{-i})$, or a PtQ format, $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$.

The crucial insight for the optimal design is that the choice of the local contract form generates asymmetric information spillovers. Holding the elimination ranking and all other local schedules fixed, shifting an expert-contingency pair from a PO to a PtQ format preserves unique implementation for expert i at an identical limiting cost. However, it fundamentally alters the informational content of i 's participation. Expert i 's participation under a PtQ schedule guarantees to all other experts that she is an L -type who will exert effort.

This on-path informational refinement immediately enlarges the assured set of high-quality components for every predecessor in the elimination ranking. By supermodularity, evaluating effort against this larger assured set strictly increases these predecessors' marginal contributions to success, thereby strictly lowering the expected bonuses required to keep them working. Thus, while PO and PtQ can uniquely implement expert i 's own action at the same cost, PtQ additionally generates a positive informational externality on earlier-ranked experts. Consequently, the principal weakly prefers the PtQ format at every expert-contingency pair, and strictly prefers it wherever expert i 's participation is payoff-relevant for a predecessor. This yields the following characterization of the optimal contract.

Proposition 4 (Optimal UI contract). *An optimal uniquely implementing contract is a PtQ contract for some $\sigma \in \text{Perm}(N)$.*

Proposition 4 characterizes the optimal *local form* of a uniquely implementing contract, conditional on a ranking. Once the elimination order σ is fixed, the principal strictly prefers a PtQ schedule to any PO schedule at every expert-contingency pair, because PtQ implements the same local incentive constraint while additionally making participation informative and thereby lowering the bonuses required for predecessors. The full optimal UI problem therefore reduces to the choice of ranking:

$$\sigma^* \in \underset{\sigma \in \text{Perm}(N)}{\text{argmin}} C^{\text{PtQ}}(\sigma, \mathbf{p}).$$

We analyze this ordering problem below, and obtain a closed-form characterization under O-ring production.

The optimal ranking depends on three objects: effort costs \mathbf{c} , type distribution \mathbf{p} , and the success technology P . Placing expert i earlier or later changes not only the size of her own critical bonus, but also the way in which her presence helps others. If i is

higher-ranked, resolving her incentive problem is expensive because relatively few other parts are already assured; however, once resolved, her part enters the assured set of every lower-ranked expert unconditionally through the elimination order. If instead i is lower-ranked, her own bonus is cheaper because more parts have already been resolved, but the benefit she creates for others is then only contingent on participation: she enters the assured set of higher-ranked experts only in those realizations in which she participates. The ranking problem therefore trades off a costly but unconditional effect from being early against a cheaper but state-contingent effect from being late.

In general, this trade-off does not reduce to a one-dimensional comparison. The difficulty is that an expert's marginal contribution need not be globally ordered across assured sets: expert i may contribute more than expert j when the assured set is small, but less when it is large or differently composed. As a result, the preferred ordering of two experts can depend on where they are placed in the hierarchy and on which participation profiles are likely to arise.

A useful heuristic to interpret this difficulty is the distinction between *generalists* and *specialists*. For a generalist expert, participation has a broad informational value: once her incentive problem is resolved, many other experts benefit because her contribution scales in a roughly uniform way with the overall quality of the project. For a specialist expert, the value of resolution is much more targeted: her contribution depends sharply on which particular parts are already assured. The more specialist the team, the more the optimal ranking is about arranging complex complementarities rather than simply ordering “important” experts early, making a closed-form index elusive.

To obtain a clear analytical characterization, additional structure on the primitives is required. In the next subsection, we show that the ranking problem becomes tractable under a multiplicative O-ring production technology. Under this structure, every expert behaves like a generalist: marginal contributions are strictly proportional to the current success probability, making pairwise comparisons invariant across assured sets. This unique property allows us to deliver a closed-form priority index for the optimal hierarchy.

6.1 O-Ring Production

We now specialize to a multiplicative success technology under which the optimal ranking admits a closed-form characterization.

Definition 3 (O-ring production). The production technology P is an *O-ring technology* if

$$P(Q) = \prod_{i \in N \setminus Q} \lambda_i,$$

where $\lambda_i \in (0, 1)$ for each $i \in N$.

Under O-ring production, each part $i \notin Q$ that fails to achieve high quality reduces the project's success probability by a factor $\lambda_i < 1$, and $P(N) = 1$. The function satisfies Assumption 1. The marginal contribution of expert i given assured set $S \subseteq N \setminus \{i\}$ is

$$P(S \cup \{i\}) - P(S) = P(S) \cdot \frac{1 - \lambda_i}{\lambda_i},$$

which is proportional to $P(S)$ with a factor depending only on i . The ratio of any two experts' marginal contributions is therefore constant across assured sets, so pairwise ranking comparisons are invariant. This is what makes the O-ring case tractable.

Under O-ring production, the critical PtQ bonus in (15) simplifies to

$$b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) = \frac{c_i}{1 - \lambda_i} \cdot \frac{1}{\prod_{\substack{j: \sigma(j) > \sigma(i), \\ a_j = 0}} \lambda_j}.$$

The first term, $\rho_i := c_i/(1 - \lambda_i)$, is expert i 's *base bonus*: the payment required to incentivize her when her part is the only unresolved source of failure. The second term inflates the bonus by $1/\lambda_j$ for each later-ranked expert j who opts out.

Taking expectations over type profiles yields a multiplicative expression for the expected cost. For expert j ranked after i : with probability $1 - p_j$ she is type L , participates, and enters the assured set; with probability p_j she is type H , opts out, and inflates expert i 's bonus by $1/\lambda_j$. Since types are independent, these effects multiply across all later-ranked experts. Define expert j 's *uncertainty loading* as

$$\kappa_j := (1 - p_j) + \frac{p_j}{\lambda_j} > 1.$$

The uncertainty loading is increasing in p_j , since

$$\frac{\partial \kappa_j}{\partial p_j} = \frac{1}{\lambda_j} - 1 > 0.$$

Thus, an expert who is more likely to be already of high quality imposes a larger expected loading on all predecessors when placed late in the hierarchy. The O-ring case therefore makes the general tension especially transparent: components that are technologically less likely to need repair can be informationally more expensive in the incentive hierarchy.

The expected cost under ranking σ is

$$C_{\mathbf{p}}^{\text{PtQ}}(\sigma) = \sum_{i \in N} \frac{c_i}{1 - \lambda_i} \prod_{\sigma(j) > \sigma(i)} \kappa_j. \quad (17)$$

Equation (17) makes the ranking problem transparent. Each expert's expected bonus

is her base bonus multiplied by the uncertainty loadings of all later-ranked experts. Ranking expert i later reduces the number of loadings that inflate her own bonus, but leaves her loading κ_i inflating every predecessor's bonus. The optimal ranking trades off these two forces. Because the ratio of any two experts' marginal contributions is constant across assured sets, the trade-off can be resolved by pairwise adjacent swaps.

Proposition 5 (Optimal ranking under O-ring). *Under O-ring production, the optimal PtQ contract ranks experts by increasing priority index*

$$I_i := \frac{c_i \lambda_i}{p_i (1 - \lambda_i)^2}. \quad (18)$$

That is, $\sigma(i) < \sigma(j)$ whenever $I_i < I_j$.

The priority index has a natural interpretation. It decomposes as $I_i = \rho_i / \phi_i$, where

$$\rho_i := \frac{c_i}{1 - \lambda_i}, \quad \phi_i := \frac{p_i(1 - \lambda_i)}{\lambda_i} = \kappa_i - 1.$$

The numerator ρ_i captures the cost of placing expert i early: a high effort cost and a weak impact on success both make early placement less attractive. The denominator ϕ_i captures the benefit of resolving expert i early: the proportional reduction in each of i 's predecessor's expected bonus when expert i is type L and participates. This benefit is larger when expert i is likely to be type H (high p_i), since a high type placed late opts out and fails to generate the participation-based spillover for predecessors, and when failure of her part is especially damaging (low λ_i), since each such opt-out inflates predecessors' bonuses by a larger factor $1/\lambda_i$. An expert should therefore be ranked early when she is cheap to incentivize, when failure of her part is especially damaging, and when she is unlikely to participate if left late.

Two special cases clarify the index. Under symmetric technology ($\lambda_i = \lambda$ for all i), the index reduces to

$$I_i \propto c_i / p_i.$$

Experts with low cost-to-high-type-probability ratios are ranked first. Under symmetric costs and technology, the ranking simplifies to decreasing p_i : experts most likely to be type H are placed earliest, as leaving them late creates the largest dead-weight inflation for predecessors.

7 Discussions

7.1 Common Prior

Throughout the paper, the principal's contract achieves unique implementation without assuming that experts share common beliefs about \mathbf{p} . We now briefly explore how the analysis changes when experts do share a common prior and the principal seeks unique implementation under Bayesian Nash equilibrium.

Because both the PtQ and PO contracts achieve unique implementation under belief-free rationalizability, they remain valid under any common prior. The dominance and iterated-elimination arguments that sustain unique implementation do not depend on what beliefs each player holds about others' types or actions; they hold at every step of the elimination procedure. Consequently, the original contracts achieve unique implementation under BNE at the same expected cost as under BFR. A common prior can only help, not hurt. The question is whether the principal can design strictly cheaper contracts by exploiting the weaker requirement of BNE, where each expert's incentive constraint need only hold in expectation under the prior rather than at every realization of others' actions.

Under PO, all types participate. With a common prior, an L -type expert knows that each opponent j is type H with probability p_j , in which case part j achieves q_H automatically. This raises expert i 's expected marginal contribution relative to the BFR worst case, where every unresolved part may belong to a non-working L -type. The critical bonus therefore falls strictly for every expert except for the last in the ranking, and the expected cost of unique implementation under PO strictly decreases.

For PtQ contract, the screening mechanism is unchanged: H -types opt out and L -types choose between working and opting out through dominance. What changes is the cost. In the candidate equilibrium, only L -types participate, so the L -type's incentive constraint becomes Bayesian: the expected benefit of working must exceed the cost, but it need not do so realization by realization. Under BNE, this benefits from two forces. First, the same automatic-quality effect as in PO: each non-participant j is an H -type whose part achieves q_H for certain, and this occurs with probability p_j . This raises expert i 's expected marginal contribution. Second, because the participation profile varies with type realizations, the principal gains an additional instrument: she can *tilt* the participation bonus $b_i(1, \mathbf{a}_{-i})$, raising it when fewer others participate and lowering it when more do, while keeping the expected on-path payment constant.

To illustrate, consider two experts with ranking $\sigma(1) = 1, \sigma(2) = 2$. On the desired path both L -types work, so the on-path expected bonus for expert 1 is $p_2 \cdot b_1(1, 0) + (1 - p_2) \cdot b_1(1, 1) = c_1 / (P_{12} - P_2)$, where $b_1(1, 0)$ applies when expert 2 does not participate and $b_1(1, 1)$ when she does. In the undesired equilibrium where both L -types opt out, expert 1's

deviation payoff involves the weighted marginal $p_2 \cdot \Delta_1(\{2\}) + (1 - p_2) \cdot \Delta_1(\emptyset)$, which is strictly below $\Delta_1(\{2\})$ by supermodularity—so a flat bonus at the PI level cannot break this equilibrium. But raising $b_1(1, 0)$ and lowering $b_1(1, 1)$ preserves the on-path expected payment while making deviation profitable. Feasibility under limited liability holds with slack, and symmetric arguments rule out asymmetric undesired equilibria. The result is unique implementation at exactly C^{PI} . The argument extends to n experts, where the participation profile provides enough variation to tilt against all undesired equilibria at the PI cost.

The cost ranking under BNE is therefore $C^{\text{PI}} = C^{\text{PtQ,BNE}} < C^{\text{PO,BNE}} < C^{\text{PO,BFR}}$: the PtQ uniqueness premium vanishes entirely, while the PO premium shrinks but remains positive because all types participate and there is no variation to exploit.

The PI-cost result for PtQ under BNE relies on two features the BFR analysis does not require. First, each expert must hold exactly the common prior belief; even a small perturbation in beliefs restores a bad equilibrium where not all parts achieve q_H . Second, each expert conditions only on her own type when acting. If experts observe any additional information—such as the realized participation profile—the tilting unravels. This reinforces the main paper’s approach: the BFR contracts deliver unique implementation without relying on a common prior or the absence of auxiliary information.

7.2 Sequential Effort Choice and Interim Ranking

Our baseline model restricts information about others’ actions by assuming that experts choose participation and effort simultaneously. This setup captures the limited observability inherent in complex real-world projects, where members often commit to tasks without knowing exactly who else has already opted in. By achieving unique implementation in this opaque environment, our approach provides a particularly strong form of robustness: the mechanism succeeds regardless of the strategic uncertainty experts face regarding others’ participation decisions.

A natural question, then, is what happens if the environment is more transparent—specifically, if experts can observe the realized participation profile \mathbf{a} before choosing their effort. In this sequential setting, the strategy space enlarges, as an expert can now specify a continuation plan contingent on the observed \mathbf{a}_{-i} .

Our robust implementation works naturally in this sequential environment. Because the PtQ incentive constraints in the baseline model are satisfied pointwise for *every* conceivable participation profile \mathbf{a}_{-i} , any continuation plan that prescribes shirking at some realized profile is strictly dominated by an otherwise identical plan that prescribes working. Therefore, observing \mathbf{a} before choosing effort does not alter the fundamental implementation logic, and the exact same PtQ contract continues to uniquely implement high quality.

This conclusion aligns perfectly with the general informational robustness of belief-free rationalizability. In the sequential extension, the realized participation profile acts as an additional signal available to experts before they choose effort. Because the fixed-ranking PtQ contract already makes working strictly optimal contingency by contingency, this extra signal can only refine experts’ beliefs—it cannot generate new rationalizable deviations. Thus, the observability of participation does not weaken the robustness property established in the main model; rather, it provides a broader epistemic interpretation of why the original implementation result survives any release of interim information.

While our implementation works as before, the observability of participation allows the principal to use another tool in the optimization problem. Once the participation profile is observed, the principal is no longer restricted to a single, rigid ex ante hierarchy. Instead, the principal can tailor the continuation hierarchy to the realized profile by using an interim ranking rule

$$\sigma : \{0, 1\}^n \rightarrow \text{Perm}(N).$$

This creates a new design margin that is absent in the simultaneous model. Under PtQ, the realized profile is informative: participants are already partly resolved by the screening logic, while non-participants remain the main source of residual uncertainty. A profile-dependent ranking can therefore economize on incentive payments by adapting the divide-and-conquer order to the realized contingency, rather than relying on a single ex ante hierarchy that must work uniformly across all type realizations.

The sequential extension is therefore best understood as preserving robustness while enriching optimization. The fixed-ranking PtQ contract from the main text remains a valid robust benchmark, but observability of participation may allow the principal to further reduce expected cost by conditioning the continuation ranking on the realized participation profile. In this sense, the extension is related to the broader idea, emphasized by [Doval and Ely \(2020\)](#), that the design of the extensive form itself can create additional value beyond what is achievable under a fixed timing structure.

7.3 Alternative Technology

Our baseline model features a *fault-repair* technology in which unskilled intervention is harmless: participating without working leaves the part at its initial quality. However, many real-world team production problems involve environments where systems are highly sensitive. In settings such as medical surgeries or the calibration of delicate machinery, unnecessary or careless tampering can actively destroy the value of an otherwise sound component.

To capture this important class of problems, we consider a *fragile intervention* technology, under which a high-quality part degrades if an expert intervenes but manages it

poorly. Specifically, the final quality is given by

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

Under this technology, an H -type who intervenes without exerting effort destroys the quality of the part. This fundamentally changes the baseline setup: unskilled participation is no longer merely wasteful, but actively harmful, meaning that preserving high quality now requires costly effort once an expert decides to participate.

The key difference from the baseline fault-repair technology is that H -types can no longer free-ride on participation: preserving the high quality of their part now requires costly effort. This fundamental shift directly alters the cost of partial implementation. Because an H -type who participates must now incur cost c_i to avoid degrading the part, the principal can separate H -types out at a non-participation payment $b_i(0, \mathbf{a}_{-i}) = b_i(1, \mathbf{a}_{-i}) - c_i/P(N)$, strictly below the participation bonus. This saves the principal $p_i \cdot c_i$ per expert in expectation, yielding an optimal PI cost of $C_{\text{fragile}}^{\text{PI}} = C^{\text{PI}} - \sum_i p_i \cdot c_i$. Unlike the baseline model, this cost is strictly lower and now depends on the type distribution \mathbf{p} .

In terms of unique implementation, both the PO and PtQ contracts derived in the previous sections continue to work exactly as before. Beyond preserving these existing mechanisms, however, the fact that participating H -types now face a cost creates an entirely new screening channel. Under the baseline fault-repair technology, a contract with $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$ induces all types to participate and produces no separation. Under fragile technology, this same bonus ordering can successfully separate types because an H -type who participates must now either work (which is costly) or shirk (which destroys quality). We now characterize this alternative separating contract, where $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$, and refer to it as *pay-to-participate* (PtP).

While the PtP contract achieves the exact same type separation as PtQ— H -types opt out and L -types participate—the direction of *strategic assurance* is fundamentally reversed. Under BFR, an observable action provides strategic assurance to the rest of the team only if every rationalizable strategy consistent with that action guarantees q_H .

Under PtQ, this assurance comes from the *participants*. The surviving rationalizable strategies are $\{(0, S)\}$ for H -types and $\{(1, W), (0, S)\}$ for L -types. Observing participation ($a_j = 1$) can therefore only be rationalized by an L -type playing $(1, W)$, safely assuring the team that the part achieves q_H .

Under PtP, however, the assurance comes from the *non-participants*. The surviving strategies become $\{(0, S)\}$ for H -types and $\{(1, W), (1, S)\}$ for L -types. Because these sets dictate disjoint participation decisions across types, observing non-participation ($a_j = 0$)

can only arise from an H -type preserving their high initial quality. Thus, every non-participant's part securely achieves q_H . Participants, on the other hand, are revealed to be L -types, but their choice between working and shirking remains strategically unresolved at this stage.

To use this in achieving unique implementation, the contract must ensure that working strictly dominates shirking for every participating L -type. Under PtP, the assured set consists of predecessors in the hierarchy plus all non-participating H -types. As formalized in Appendix B, this yields a critical bonus structurally identical to that of PtQ, but with non-participants replacing working participants in the assured set.

Ultimately, under fragile technology, the principal can overcome the one-sided screening limitation of the baseline model. Under fault-repair, a high probability of H -types (p_j) was strictly a source of cost inflation, as these experts opted out and shrank the assured set for subsequent team members. Under fragile technology, when p_j is large, the principal can instead deploy the PtP contract. This channels the abundance of H -types into the assured set through non-participation, turning a former liability into a powerful source of cost reduction. By tailoring the contract form to each expert's type distribution, the principal maximally exploits adverse selection from both directions: L -types reveal themselves by participating under PtQ, and H -types reveal themselves by opting out under PtP.

More broadly, the analysis shows that technological slack and informational slack need not move in the same direction. A project may be easier in the sense that fewer parts require costly repair, yet harder to implement robustly because observed intervention becomes less informative about productive need. The role of pay-to-quit is to reconcile these two forces by turning hidden slack into an informative participation margin.

References

- Au, P. H. (2020). Pay to quit and team incentives. *Journal of the European Economic Association*, 18(2):964–1008.
- Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, pages 1771–1813.
- Bergemann, D. and Morris, S. (2017). Belief-free rationalizability and informational robustness. *Games and Economic Behavior*, 104:744–759.
- Bernstein, S. and Winter, E. (2012). Contracting with heterogeneous externalities. *American Economic Journal: Microeconomics*, 4(2):50–76.
- Camboni, M. and Porcellacchia, M. (2025). Monitoring team members: Information waste

- and the transparency trap. *American Economic Journal: Microeconomics*, 17(3):324–368.
- Cavounidis, C. and Park, H. (2025). Contracts from the horse’s mouth. Working Paper.
- Doval, L. and Ely, J. C. (2020). Sequential information design. *Econometrica*, 88(6):2575–2608.
- Foarta, D. and Sugaya, T. (2021). The management of talent: Optimal contracting for selection and incentives. *The RAND Journal of Economics*, 52(1):49–77.
- Gershkov, A. and Winter, E. (2015). Formal versus informal monitoring in teams. *American Economic Journal: Microeconomics*, 7(2):27–44.
- Green, J. R. and Laffont, J.-J. (1986). Partially verifiable information and mechanism design. *The Review of Economic Studies*, 53(3):447–456.
- Halac, M., Kremer, I., and Winter, E. (2020). Raising capital from heterogeneous investors. *American Economic Review*, 110(3):889–921.
- Halac, M., Lipnowski, E., and Rappoport, D. (2021). Rank uncertainty in organizations. *American Economic Review*, 111(3):757–786.
- Lazear, E. P. (1979). Why is there mandatory retirement? *Journal of political economy*, 87(6):1261–1284.
- Lazear, E. P. (1983). Pensions as severance pay. In *Financial aspects of the United States pension system*, pages 57–90. University of Chicago Press.
- Lu, Z. and Song, Y. (2025). Network-based peer monitoring design. *Journal of Economic Theory*, 224:105969.
- Moriya, F. and Yamashita, T. (2020). Asymmetric-information allocation to avoid coordination failure. *Journal of Economics & Management Strategy*, 29(1):173–186.
- Park, H. and Song, C.-O. (2025). Non-monitoring monitoring in teams. *Available at SSRN 5417192*.
- Segal, I. (1999). Contracting with externalities. *The Quarterly Journal of Economics*, 114(2):337–388.
- Segal, I. (2003). Coordination and discrimination in contracting with externalities: Divide and conquer? *Journal of Economic Theory*, 113(2):147–181.
- Wilson, R. (1987). *Game-theoretic analyses of trading processes*, page 33–70. Econometric Society Monographs. Cambridge University Press.

- Winter, E. (2004). Incentives and discrimination. *American Economic Review*, 94(3):764–773.
- Winter, E. (2006). Optimal incentives for sequential production processes. *The RAND Journal of Economics*, 37(2):376–390.
- Winter, E. (2010). Transparency and incentives among peers. *The RAND Journal of Economics*, 41(3):504–523.

A Belief-Free Rationalizability

We formally define belief-free rationalizability (BFR) for the game induced by a bonus schedule \mathbf{b} , following [Bergemann and Morris \(2017\)](#). Our setting is a payoff-type environment: each expert i has payoff-type $\theta_i \in \{H, L\}$, the payoff-relevant state is the type profile $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, and every expert considers every profile of others' types possible. Under these assumptions, a payoff-type is a sufficient statistic for a player's higher-order beliefs, and we identify types with payoff-types throughout. The definition below corresponds to the simplified formulation in [Bergemann and Morris \(2017, Section 3.1\)](#) for payoff-type environments with full support.

Each expert i simultaneously chooses an action $(a_i, e_i) \in \mathcal{A}_i = \{(1, W), (1, S), (0, S)\}$ after privately observing her type θ_i . The participation profile \mathbf{a} is realized only after all choices are made.

Definition 4 (Belief-free rationalizability). Fix a bonus schedule \mathbf{b} . For each expert i and type $\theta_i \in \{H, L\}$, define the sequence of surviving action sets $\{\text{BFR}_i^k(\theta_i)\}_{k \geq 0}$ as follows:

1. **Initialize:** $\text{BFR}_{i,0}(\theta_i) = \mathcal{A}_i$ for all $i \in N$ and $\theta_i \in \{H, L\}$.
2. **Iterate:** For $k \geq 0$, an action $(a_i, e_i) \in \text{BFR}_i^k(\theta_i)$ belongs to $\text{BFR}_i^{k+1}(\theta_i)$ if and only if there exists a belief $\nu \in \Delta(\prod_{j \neq i} (\{H, L\} \times \mathcal{A}_j))$ such that:

- (i) *Support consistency:* for every $(\boldsymbol{\theta}_{-i}, \mathbf{a}_{-i}, \mathbf{e}_{-i})$ in the support of ν and every $j \neq i$,

$$(a_j, e_j) \in \text{BFR}_j^k(\theta_j);$$

- (ii) *Best response:*

$$(a_i, e_i) \in \operatorname{argmax}_{(a'_i, e'_i) \in \mathcal{A}_i} \mathbb{E}_\nu [u_i(\boldsymbol{\theta}, (a'_i, \mathbf{a}_{-i}), (e'_i, \mathbf{e}_{-i}))].$$

3. **Limit:** $\text{BFR}_i(\theta_i) = \bigcap_{k \geq 0} \text{BFR}_i^k(\theta_i)$.

The set of belief-free rationalizable action profiles is $\text{BFR}(\mathbf{b}) = \prod_{i \in N} \prod_{\theta_i \in \{H, L\}} \text{BFR}_i(\theta_i)$.

The sequence $\{\text{BFR}_{i,k}^{\mathbf{b}}(\theta_i)\}_{k \geq 0}$ is weakly decreasing. Since \mathcal{A}_i is finite for each i , the sequence stabilizes in finitely many steps.

B Fragile intervention and pay-to-participate contracts

This appendix derives the alternative screening contract available under the fragile intervention technology discussed in Section 7.3. Throughout, fix a ranking $\sigma \in \text{Perm}(N)$. For each expert i , define

$$\bar{\rho}_i := \max_{Q \subseteq N \setminus \{i\}} \frac{P(Q)}{P(Q \cup \{i\})}.$$

Strict monotonicity implies $\bar{\rho}_i \in (0, 1)$.

Lemma B.1 (First-round separation under fragile intervention). *Fix expert i and a contingency \mathbf{a}_{-i} . Suppose*

$$\max \left\{ b_i(1, \mathbf{a}_{-i}) - \frac{c_i}{P(N)}, \bar{\rho}_i b_i(1, \mathbf{a}_{-i}) \right\} < b_i(0, \mathbf{a}_{-i}) < b_i(1, \mathbf{a}_{-i}). \quad (19)$$

Then, after the first round of belief-free elimination,

$$\text{BFR}_i^1(H) = \{(0, S)\}, \quad \text{BFR}_i^1(L) = \{(1, W), (1, S)\}.$$

Proof of Lemma B.1. Consider first an H -type expert i . To make $(0, S)$ survive and eliminate both participation actions, $(0, S)$ must strictly dominate $(1, W)$ and $(1, S)$ for every belief about the other experts' surviving type-action pairs.

Compare $(0, S)$ and $(1, W)$. For any set $Q \subseteq N \setminus \{i\}$ of other parts that expert i believes attain high quality, both actions preserve part i at q_H , so both generate the same success probability $P(Q \cup \{i\})$. The only difference is that $(1, W)$ incurs cost c_i . Hence $(0, S)$ strictly dominates $(1, W)$ whenever

$$P(Q \cup \{i\}) b_i(0, \mathbf{a}_{-i}) > P(Q \cup \{i\}) b_i(1, \mathbf{a}_{-i}) - c_i$$

for every $Q \subseteq N \setminus \{i\}$. This is equivalent to

$$b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i}) - \frac{c_i}{P(Q \cup \{i\})}$$

for every Q . The right-hand side is largest when $Q = N \setminus \{i\}$, yielding the first term in (19).

Next compare $(0, S)$ and $(1, S)$. If expert i chooses $(0, S)$, part i remains at q_H and the success probability is $P(Q \cup \{i\})$. If she chooses $(1, S)$, part i falls to q_L and the success probability is $P(Q)$. Thus $(0, S)$ strictly dominates $(1, S)$ whenever

$$P(Q \cup \{i\}) b_i(0, \mathbf{a}_{-i}) > P(Q) b_i(1, \mathbf{a}_{-i})$$

for every $Q \subseteq N \setminus \{i\}$, or equivalently,

$$b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i}) \cdot \frac{P(Q)}{P(Q \cup \{i\})}$$

for every Q . By the definition of $\bar{\rho}_i$, this is guaranteed by the second term in (19). Since both lower bounds in (19) are strictly below $b_i(1, \mathbf{a}_{-i})$, the interval is nonempty.

Now consider an L -type expert i . Both $(0, S)$ and $(1, S)$ leave part i at q_L , so they generate the same success probability for every belief. Because $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$, action $(0, S)$ is strictly dominated by $(1, S)$. No condition in (19) yet ranks $(1, W)$ and $(1, S)$ for the L -type, so both remain. Hence

$$\text{BFR}_i^1(H) = \{(0, S)\}, \quad \text{BFR}_i^1(L) = \{(1, W), (1, S)\}.$$

□

The lemma shows that fragile intervention creates a new separating margin with the opposite bonus ordering from pay-to-quit. The contract pays more for participation, yet high types still sort into non-participation because participation is no longer costless for them.

After the first round, a realized non-participation decision identifies a high type. For each expert i and realization \mathbf{a}_{-i} , let

$$O_{-i}(\mathbf{a}_{-i}) := \{j \in N \setminus \{i\} : a_j = 0\}$$

denote the set of non-participants. By Lemma B.1, every part in $O_{-i}(\mathbf{a}_{-i})$ is guaranteed to attain q_H .

Define the critical participation bonus for expert i under ranking σ by

$$b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) := \frac{c_i}{P(R_{<i}^\sigma \cup O_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup O_{-i}(\mathbf{a}_{-i}))}. \quad (20)$$

Proposition B.1 (Unique implementation under pay-to-participate). *Fix a ranking σ . Suppose that, for every expert i and every realization \mathbf{a}_{-i} ,*

$$b_i(1, \mathbf{a}_{-i}) = b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_i^1(\mathbf{a}_{-i}), \quad (21)$$

$$\max \left\{ b_i(1, \mathbf{a}_{-i}) - \frac{c_i}{P(N)}, \bar{\rho}_i b_i(1, \mathbf{a}_{-i}) \right\} < b_i(0, \mathbf{a}_{-i}) < b_i(1, \mathbf{a}_{-i}) \quad (22)$$

for some $\varepsilon_i^1(\mathbf{a}_{-i}) > 0$. Then the contract uniquely implements high quality under belief-free rationalizability: $(0, S)$ is the only surviving action for every H -type expert and $(1, W)$ is the only surviving action for every L -type expert.

Proof of Proposition ??. Step 1: First-round elimination. By Proposition B.1, after the

first round we have

$$\text{BFR}_i^1(H) = \{(0, S)\}, \quad \text{BFR}_i^1(L) = \{(1, W), (1, S)\}$$

for every expert i .

Step 2: Inductive elimination along the ranking. Let i_m denote the expert with rank m , that is, $\sigma(i_m) = m$. We show by induction that for each $m = 1, \dots, n$,

$$\text{BFR}_{i_m}^{m+1}(L) = \{(1, W)\}.$$

Base case. Consider the first-ranked expert i_1 . After the first round, every non-participant in \mathbf{a}_{-i_1} is known to be an H -type and therefore achieves q_H . The worst-case belief for comparing $(1, W)$ and $(1, S)$ for an L -type i_1 is that every participant in $N \setminus (O_{-i_1}(\mathbf{a}_{-i_1}) \cup \{i_1\})$ shirks. Under that belief, working yields success probability

$$P(O_{-i_1}(\mathbf{a}_{-i_1}) \cup \{i_1\}),$$

while shirking yields

$$P(O_{-i_1}(\mathbf{a}_{-i_1})).$$

Since both actions receive the same bonus $b_{i_1}(1, \mathbf{a}_{-i_1})$, working strictly dominates shirking whenever

$$b_{i_1}(1, \mathbf{a}_{-i_1}) \left[P(O_{-i_1}(\mathbf{a}_{-i_1}) \cup \{i_1\}) - P(O_{-i_1}(\mathbf{a}_{-i_1})) \right] > c_{i_1},$$

which holds by (21) and (20). Hence

$$\text{BFR}_{i_1}^2(L) = \{(1, W)\}.$$

Induction step. Fix $m \geq 2$ and suppose that for every $k < m$,

$$\text{BFR}_{i_k}^{k+1}(L) = \{(1, W)\}.$$

Then every predecessor in $R_{<i_m}^\sigma$ is already pinned down to produce q_H , regardless of type. In addition, every non-participant in $O_{-i_m}(\mathbf{a}_{-i_m})$ is a revealed H -type and therefore also produces q_H . The worst-case belief for an L -type i_m is again that every remaining participant outside

$$R_{<i_m}^\sigma \cup O_{-i_m}(\mathbf{a}_{-i_m}) \cup \{i_m\}$$

shirks. Under that belief, working yields success probability

$$P(R_{<i_m}^\sigma \cup O_{-i_m}(\mathbf{a}_{-i_m}) \cup \{i_m\}),$$

while shirking yields

$$P(R_{<i_m}^\sigma \cup O_{-i_m}(\mathbf{a}_{-i_m})).$$

Therefore $(1, W)$ strictly dominates $(1, S)$ whenever

$$b_{i_m}(1, \mathbf{a}_{-i_m}) \left[P(R_{<i_m}^\sigma \cup O_{-i_m}(\mathbf{a}_{-i_m}) \cup \{i_m\}) - P(R_{<i_m}^\sigma \cup O_{-i_m}(\mathbf{a}_{-i_m})) \right] > c_{i_m},$$

which again holds by (21) and (20). Hence

$$\text{BFR}_{i_m}^{m+1}(L) = \{(1, W)\}.$$

This completes the induction.

Step 3: Conclusion. After $n + 1$ rounds, every H -type expert has the unique surviving action $(0, S)$ and every L -type expert has the unique surviving action $(1, W)$. Under fragile intervention, both actions produce q_H . Hence every belief-free rationalizable action for every expert and every type yields high quality, so the contract uniquely implements high quality. \square

To state the cost of the pay-to-participate contract, define the lowest non-participation bonus consistent with first-round separation by

$$\underline{b}_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) := \max \left\{ b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) - \frac{c_i}{P(N)}, \bar{\rho}_i b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) \right\}. \quad (23)$$

At the desired outcome, an L -type expert receives the participation bonus $b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i})$ while an H -type expert receives the non-participation bonus $\underline{b}_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i})$. The infimum expected cost under ranking σ is therefore

$$C^{\text{PtP}}(\sigma, \mathbf{p}) = P(N) \cdot \mathbb{E}_{\mathbf{p}} \left[\sum_{i:\theta_i=L} b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) + \sum_{i:\theta_i=H} \underline{b}_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) \right], \quad (24)$$

where the participation profile is given by $a_j = \mathbb{1}\{\theta_j = L\}$.

Corollary 3 (O-ring specialization). *Suppose the production technology is O-ring, so that*

$$P(Q) = \prod_{j \in N \setminus Q} \lambda_j, \quad \lambda_j \in (0, 1).$$

Then the critical participation bonus becomes

$$b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) = \frac{c_i}{1 - \lambda_i} \prod_{\substack{j:\sigma(j) > \sigma(i) \\ a_j=1}} \frac{1}{\lambda_j}. \quad (25)$$

Moreover, $\bar{\rho}_i = \lambda_i$, the first term in (23) binds, and hence

$$\underline{b}_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) = b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) - c_i. \quad (26)$$

Therefore the expected cost simplifies to

$$C^{\text{PtP}}(\sigma, \mathbf{p}) = \sum_{i \in N} \frac{c_i}{1 - \lambda_i} \prod_{\sigma(j) > \sigma(i)} \left(p_j + \frac{1 - p_j}{\lambda_j} \right) - \sum_{i \in N} p_i c_i. \quad (27)$$

For a fixed contract form—that is, when every expert is assigned pay-to-participate—the optimal ranking is obtained by increasing priority index

$$I_i^{\text{PtP}} := \frac{c_i \lambda_i}{(1 - p_i)(1 - \lambda_i)^2}. \quad (28)$$

Proof of Corollary 3. Under O-ring production,

$$P(R_{<i}^\sigma \cup O_{-i}(\mathbf{a}_{-i})) = \lambda_i \prod_{\substack{j: \sigma(j) > \sigma(i) \\ a_j = 1}} \lambda_j.$$

Hence

$$P(R_{<i}^\sigma \cup O_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup O_{-i}(\mathbf{a}_{-i})) = (1 - \lambda_i) \prod_{\substack{j: \sigma(j) > \sigma(i) \\ a_j = 1}} \lambda_j,$$

which yields (25).

Next,

$$\frac{P(Q)}{P(Q \cup \{i\})} = \lambda_i$$

for every $Q \subseteq N \setminus \{i\}$, so $\bar{\rho}_i = \lambda_i$. Also, substituting (25) into the two terms in (23) gives

$$b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}) - c_i \geq \lambda_i b_i^{\text{PtP}}(\sigma, \mathbf{a}_{-i}),$$

because

$$\prod_{\substack{j: \sigma(j) > \sigma(i) \\ a_j = 1}} \lambda_j \leq 1.$$

Hence the first term binds, establishing (26).

Finally, a later-ranked expert j multiplies expert i 's participation bonus by $1/\lambda_j$ ex-

actly when $a_j = 1$, that is, exactly when $\theta_j = L$. Independence of types therefore implies

$$\mathbb{E} \left[\prod_{\substack{j:\sigma(j)>\sigma(i) \\ a_j=1}} \frac{1}{\lambda_j} \right] = \prod_{\sigma(j)>\sigma(i)} \left(p_j + \frac{1-p_j}{\lambda_j} \right).$$

Substituting this expression and (26) into (24) yields (27).

The ranking rule follows from the standard adjacent-swap argument. Write

$$\rho_i := \frac{c_i}{1-\lambda_i}, \quad \kappa_i^{\text{PtP}} := p_i + \frac{1-p_i}{\lambda_i}.$$

Ignoring the additive constant $\sum_i p_i c_i$, the expected cost is

$$\sum_{i \in N} \rho_i \prod_{\sigma(j)>\sigma(i)} \kappa_j^{\text{PtP}}.$$

Comparing two adjacent experts i and j , placing i before j is cheaper if and only if

$$\rho_i (\kappa_j^{\text{PtP}} - 1) \leq \rho_j (\kappa_i^{\text{PtP}} - 1),$$

which is equivalent to $I_i^{\text{PtP}} \leq I_j^{\text{PtP}}$ with

$$I_i^{\text{PtP}} = \frac{\rho_i}{\kappa_i^{\text{PtP}} - 1} = \frac{c_i \lambda_i}{(1-p_i)(1-\lambda_i)^2}.$$

Thus the optimal ranking is obtained by increasing (28). \square

The O-ring formulas make the comparison with pay-to-quit transparent. Under pay-to-quit, a later expert raises predecessors' expected bonuses when she is a high type and therefore fails to generate an informative participation decision; under pay-to-participate, the corresponding inflation arises when she is a low type and must participate. Thus the two contracts are mirror images on the informational margin. The additional term $-\sum_i p_i c_i$ in (27) captures a further difference: pay-to-participate also lowers the principal's direct payment to high types, because they sort into the lower non-participation bonus.

If the principal allows both contract forms simultaneously, the expected cost under O-ring production and a fixed ranking σ can be written as

$$C(\sigma, f) = \sum_{i \in N} \frac{c_i}{1-\lambda_i} \prod_{\sigma(j)>\sigma(i)} \kappa_j(f_j) - \sum_{i:f_i=\text{PtP}} p_i c_i, \quad (29)$$

where $f_i \in \{\text{PtQ}, \text{PtP}\}$ denotes the form assigned to expert i , and

$$\kappa_i(\text{PtQ}) = (1 - p_i) + \frac{p_i}{\lambda_i}, \quad \kappa_i(\text{PtP}) = p_i + \frac{1 - p_i}{\lambda_i}.$$

Conditional on the form profile f , the optimal ranking is still obtained by an adjacent-swap argument: expert i is ranked before expert j whenever

$$\frac{c_i}{(1 - \lambda_i)(\kappa_i(f_i) - 1)} \leq \frac{c_j}{(1 - \lambda_j)(\kappa_j(f_j) - 1)}.$$

The assignment of forms is no longer fully separable expert by expert, however, because choosing pay-to-participate for expert i both changes the loading borne by all predecessors and lowers expert i 's own expected payment by $p_i c_i$. What remains robust is that if $p_i \geq 1/2$, pay-to-participate weakly dominates pay-to-quit for expert i : it yields a weakly smaller loading and, in addition, a weakly lower direct payment. When $p_i < 1/2$, by contrast, the optimal form depends on the balance between these two effects.

C Proofs

C.1 Proof of Proposition 1

Proof of Proposition 1. We must show that the stated bonus schedule minimizes the principal's expected cost among all contracts under which, for every expert i with $\theta_i = L$, the action $(1, W)$ is belief-free rationalizable.

Step 1: PI reduces to an L-type condition. Under the fault-repair technology, $q(H, e_i) = q_H$ for any e_i , so an H -type's action has no effect on the quality of her part. The requirement that every part attain q_H therefore imposes no restriction on H -types. Partial implementation thus requires only that $(1, W)$ be belief-free rationalizable for every expert with $\theta_i = L$.

Step 2: A necessary incentive condition. Fix expert i with $\theta_i = L$. Suppose that the belief under which every other expert participates and every other L -type works is supported on surviving type–action pairs. Under this belief, all parts other than i 's attain high quality, so $(1, W)$ is a best response only if it weakly dominates both $(1, S)$ and $(0, S)$:

$$P(N) b_i(1, \mathbf{a}_{-i}) - c_i \geq P(N \setminus \{i\}) b_i(1, \mathbf{a}_{-i}), \quad (30)$$

$$P(N) b_i(1, \mathbf{a}_{-i}) - c_i \geq P(N \setminus \{i\}) b_i(0, \mathbf{a}_{-i}). \quad (31)$$

In particular, (30) implies

$$b_i(1, \mathbf{a}_{-i}) \geq \frac{c_i}{P(N) - P(N \setminus \{i\})}. \quad (32)$$

Step 3: Non-participation bonuses are wasteful. We claim that, without loss of optimality, the principal sets

$$b_i(0, \mathbf{a}_{-i}) = 0 \quad \text{for all } i \text{ and } \mathbf{a}_{-i}.$$

First, if for some i and \mathbf{a}_{-i} we had

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

then replacing $b_i(0, \mathbf{a}_{-i})$ by 0 leaves (30) unchanged and relaxes (31), while strictly lowering the expected payment whenever expert i is type H . Hence no optimal PI contract can satisfy $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$.

Second, once $b_i(0, \mathbf{a}_{-i}) \leq b_i(1, \mathbf{a}_{-i})$, an H -type weakly prefers $(1, S)$ to $(0, S)$, since both actions yield q_H and involve no effort cost. Thus the favorable belief used in Step 2 can be supported with H -types participating. But then $b_i(0, \mathbf{a}_{-i})$ serves no incentive purpose: it enters only on the right-hand side of (31), so lowering it weakly relaxes the

L -type incentive constraint and weakly lowers cost. Therefore we may set $b_i(0, \mathbf{a}_{-i}) = 0$ for all i and \mathbf{a}_{-i} .

Step 4: Optimal participation bonus. With $b_i(0, \mathbf{a}_{-i}) = 0$, the relevant lower bound is (32). Since the principal's expected cost is increasing in bonuses, the minimum feasible participation bonus is obtained at equality:

$$b_i^{\text{PI}}(1, \mathbf{a}_{-i}) = \frac{c_i}{P(N) - P(N \setminus \{i\})}, \quad b_i^{\text{PI}}(0, \mathbf{a}_{-i}) = 0$$

for all i and all $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$.

Step 5: Verification. Under the proposed contract, for any expert of type H , action $(1, S)$ strictly dominates both $(1, W)$ and $(0, S)$ for every belief over other experts' types and actions. Hence $(1, S)$ is the unique belief-free rationalizable action for type H .

Now consider the profile in which every expert participates, every L -type chooses W , and every H -type chooses S . Fix an expert i with $\theta_i = L$, and consider the belief that every other expert follows this profile. Since all other parts then attain q_H , we have

$$P(N) b_i^{\text{PI}}(1, \mathbf{a}_{-i}) - c_i = P(N \setminus \{i\}) b_i^{\text{PI}}(1, \mathbf{a}_{-i}) > 0 = P(N \setminus \{i\}) b_i^{\text{PI}}(0, \mathbf{a}_{-i}),$$

It follows that $(1, W)$ is indifferent to $(1, S)$ and $(1, W)$ is strictly preferred to $(0, S)$. Therefore $(1, W)$ is a belief-free rationalizable action for every expert of type L .

We have thus showed that the belief used in Step 2 is supported by the surviving actions in the iterated elimination. Therefore the all-high-quality profile described above is belief-free rationalizable, so the contract partially implements high quality.

Step 6: Cost and independence from \mathbf{p} . Under the rationalizable all-high-quality profile just constructed, every expert participates, so the participation profile is $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{1}$ for every type profile $\boldsymbol{\theta}$. Therefore the expected cost is

$$C^{\text{PI}} = P(N) \sum_{i \in N} b_i^{\text{PI}}(1, \mathbf{1}_{-i}) = P(N) \sum_{i \in N} \frac{c_i}{P(N) - P(N \setminus \{i\})}.$$

This expression depends only on (c_i, P) and not on \mathbf{p} . □

C.2 Proof of Proposition 2

Proof of Proposition 2. Fix a ranking σ . Suppose the contract is PO and satisfies

$$b_i(1, \mathbf{a}_{-i}) > b_i^{\text{PO}}(\sigma) = \frac{c_i}{P(R_{\leq i}^\sigma) - P(R_{< i}^\sigma)} \quad \text{and} \quad b_i(0, \mathbf{a}_{-i}) = 0$$

for all $i \in N$ and all $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$. We show that this contract uniquely implements high quality.

Step 1: First-round elimination. For an expert of type H , action $(1, S)$ strictly dominates both $(1, W)$ and $(0, S)$ for every belief over others' types and actions. Indeed, $(1, S)$ and $(1, W)$ induce the same quality q_H , but $(1, W)$ incurs cost $c_i > 0$, while $(1, S)$ and $(0, S)$ also induce the same quality q_H , and

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

Hence

$$\text{BFR}_i^1(H) = \{(1, S)\}.$$

For an expert of type L , action $(1, S)$ strictly dominates $(0, S)$ for every belief, since both actions leave the part at q_L , but $(1, S)$ yields the strictly higher bonus. Therefore

$$\text{BFR}_i^1(L) = \{(1, S), (1, W)\}.$$

Step 2: Inductive elimination along the ranking. We claim that for each $i = 1, \dots, n$,

$$\text{BFR}_i^{m+1}(L) = \{(1, W)\}.$$

Thus the L -types are pinned down iteratively along the ranking.

We prove this by induction on the ranking $\sigma(i) = m$.

Base case: $m = 1$. Consider expert $i = \sigma^{-1}(1)$ with type L . By Step 1, for every $j \neq i$,

$$\text{BFR}_j^1(H) = \{(1, S)\}, \quad \text{BFR}_j^1(L) = \{(1, S), (1, W)\}.$$

Among all beliefs supported on these surviving actions, the worst case for the comparison between $(1, W)$ and $(1, S)$ is that every other expert is type L and chooses $(1, S)$. Under this belief, no other part attains q_H . Hence $(1, W)$ yields success probability $P(\{i\})$, while $(1, S)$ yields success probability $P(\emptyset)$. $(1, W)$ strictly dominates $(1, S)$ whenever

$$(P(\{i\}) - P(\emptyset)) b_i(1, \mathbf{a}_{-i}) > c_i,$$

which holds by the construction of $b_i^{PO}(\sigma)$. Hence

$$\text{BFR}_i^2(L) = \{(1, W)\}.$$

Induction step. Fix $m \geq 2$, and suppose that for every j such that $\sigma(j) < m$,

$$\text{BFR}_j^{\sigma(1)+1}(L) = \{(1, W)\}.$$

Then for every predecessor $j \in R_{<\sigma^{-1}(m)}^\sigma$, the part j is guaranteed to attain q_H : if $\theta_j = H$,

then $\text{BFR}_j^1(H) = \{(1, S)\}$; if $\theta_j = L$, then the induction hypothesis implies that the only surviving action from round $\sigma(j) + 1$ onward is $(1, W)$.

Now consider expert $k = \sigma^{-1}(m)$ with type L . At iteration m , every predecessor in $R_{<k}^\sigma$ is already pinned down to produce q_H , while every successor $j \notin R_{\leq k}^\sigma$ may still choose any action in $\text{BFR}_j^m(\theta_j)$. The worst-case belief for the comparison between $(1, W)$ and $(1, S)$ is therefore that every successor is type L and chooses $(1, S)$, so that no successor's part attains q_H . This belief is supported on the surviving type–action pairs at iteration m .

Under that belief, if expert k works, the set of high-quality parts is $R_{\leq k}^\sigma$; if she shirks, the set of high-quality parts is $R_{<k}^\sigma$. Since both actions involve participation, they yield the same bonus $b_k(1, \mathbf{a}_{-k})$. Hence $(1, W)$ strictly dominates $(1, S)$ whenever

$$(P(R_{\leq k}^\sigma) - P(R_{<k}^\sigma)) b_k(1, \mathbf{a}_{-k}) > c_k,$$

which holds by the assumed bonus bound

$$b_k(1, \mathbf{a}_{-k}) > \frac{c_k}{P(R_{\leq k}^\sigma) - P(R_{<k}^\sigma)} = b_k^{\text{PO}}(\sigma).$$

Therefore

$$\text{BFR}_k^{m+1}(L) = \{(1, W)\}.$$

By induction, this holds for every $m = 1, \dots, n$.

Step 3: Conclusion. After iteration $n + 1$, every expert of type H has the unique belief-free rationalizable action $(1, S)$, and every expert of type L has the unique belief-free rationalizable action $(1, W)$. Under both actions the part attains high quality:

$$q(H, S) = q_H, \quad q(L, W) = q_H.$$

Therefore every belief-free rationalizable action for every expert and every type yields q_H . Hence the contract uniquely implements high quality. \square

C.3 Proof of Proposition 3

Proof of Proposition 3. Fix a ranking σ . The proposed contract sets

$$b_i(1, \mathbf{a}_{-i}) = b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_1, \quad b_i(0, \mathbf{a}_{-i}) = b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_0,$$

with $\varepsilon_0 > \varepsilon_1 > 0$, where

$$b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) = \frac{c_i}{P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}))}.$$

Because the collection of pairs (i, Q) with $Q \subseteq N \setminus \{i\}$ is finite and $P(Q \cup \{i\}) > P(Q)$ for every such pair, the number

$$\bar{r} := \max_{i \in N} \max_{Q \subseteq N \setminus \{i\}} \frac{P(Q)}{P(Q \cup \{i\})}$$

is strictly less than one. Choose any $r \in (\bar{r}, 1)$, any $\varepsilon_0 > 0$, and set $\varepsilon_1 = r\varepsilon_0$. Then for every i and every $Q \subseteq N \setminus \{i\}$,

$$\varepsilon_1 P(Q \cup \{i\}) - \varepsilon_0 P(Q) = \varepsilon_0 (rP(Q \cup \{i\}) - P(Q)) > 0. \quad (33)$$

Step 1: First-round elimination. In the first-round elimination of belief-free rationalizability, we shrink the rationalizable sets for both types by inducing separation by type.

Fix expert i . If $\theta_i = H$, then $(0, S)$ and $(1, S)$ both leave part i at q_H , so they induce the same success probability for each realization of opponents' actions \mathbf{a}_{-i} . Since $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$, action $(0, S)$ strictly dominates $(1, S)$. Likewise, $(0, S)$ strictly dominates $(1, W)$ because both yield q_H while $(1, W)$ gives a lower bonus and also incurs cost $c_i > 0$. Hence

$$\text{BFR}_i^1(H) = \{(0, S)\}.$$

If $\theta_i = L$, then $(0, S)$ and $(1, S)$ both leave part i at q_L , hence again induce the same success probability for every opponents' realization. Because $b_i(0, \mathbf{a}_{-i}) > b_i(1, \mathbf{a}_{-i})$, action $(0, S)$ strictly dominates $(1, S)$. The comparison between $(0, S)$ and $(1, W)$, however, depends on bonus levels and others' actions, so neither dominates the other without further information. Hence

$$\text{BFR}_i^1(L) = \{(0, S), (1, W)\}.$$

Step 2: Inductive elimination along the ranking. Let $i = \sigma^{-1}(k)$. We claim that, for each $k = 1, \dots, n$, the action $(0, S)$ of an L -type i is eliminated at round $k + 1$, so that

$$\text{BFR}_i^{k+1}(L) = \{(1, W)\}.$$

We prove this by induction on k .

Base case: $k = 1$. Consider expert $i = \sigma^{-1}(1)$ with $\theta_i = L$. Fix a participation profile \mathbf{a}_{-i} and any belief supported on the surviving type-action pairs from Step 1. Since $R_{<i}^\sigma = \emptyset$, the set of other parts that are assured to be of high quality consists precisely of those experts who participate: by Step 1, any participant j with $a_j = 1$ must be a type- L expert playing $(1, W)$, so part j attains q_H . Let Q denote the set of other parts at high quality under this belief. Then $Q \supseteq W_{-i}(\mathbf{a}_{-i})$.

The payoff difference between $(1, W)$ and $(0, S)$ is

$$\begin{aligned} & (b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_1) P(Q \cup \{i\}) - (b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_0) P(Q) - c_i \\ &= b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) [P(Q \cup \{i\}) - P(Q)] - c_i + \varepsilon_1 P(Q \cup \{i\}) - \varepsilon_0 P(Q). \end{aligned}$$

Since $Q \supseteq W_{-i}(\mathbf{a}_{-i})$, supermodularity implies

$$P(Q \cup \{i\}) - P(Q) \geq P(W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(W_{-i}(\mathbf{a}_{-i})),$$

so the first two terms together are nonnegative by the definition of $b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i})$. The last two terms are strictly positive by (33). Hence $(1, W)$ strictly dominates $(0, S)$ for type- L expert i , and

$$\text{BFR}_i^2(L) = \{(1, W)\}.$$

Induction step. Fix $m \geq 2$, and suppose that for every j such that $\sigma(j) < m$,

$$\text{BFR}_j^{m+1}(L) = \{(1, W)\}.$$

Then for every predecessor $j \in R_{<\sigma^{-1}(m)}^\sigma$, part j is guaranteed to attain q_H .

Now consider expert i with $\theta_i = L$ where $\sigma(i) = k > m$. Fix a participation profile \mathbf{a}_{-i} and any belief supported on the surviving type-action pairs at round $k + 1$. The predecessors in $R_{<i}^\sigma$ are already resolved. Among the remaining experts, any participant j with $a_j = 1$ and $j \notin R_{<i}^\sigma$ must be a type- L expert whose surviving actions are $(0, S)$ or $(1, W)$; but since $a_j = 1$, only $(1, W)$ is consistent, so part j attains q_H . Let Q denote the set of other parts at high quality under this belief. Then

$$Q \supseteq R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}).$$

The payoff difference between $(1, W)$ and $(0, S)$ is

$$\begin{aligned} & (b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_1) P(Q \cup \{i\}) - (b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) + \varepsilon_0) P(Q) - c_i \\ &= b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i}) [P(Q \cup \{i\}) - P(Q)] - c_i + \varepsilon_1 P(Q \cup \{i\}) - \varepsilon_0 P(Q). \end{aligned}$$

Since $Q \supseteq R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i})$, supermodularity implies

$$P(Q \cup \{i\}) - P(Q) \geq P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i})),$$

so the first two terms together are nonnegative by the definition of $b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i})$. The last two terms are strictly positive by (33). Hence $(1, W)$ strictly dominates $(0, S)$ for type- L expert i , and

$$\text{BFR}_i^{k+1}(L) = \{(1, W)\}.$$

This completes the induction.

Step 3: Conclusion. After round $n + 1$, every expert of type H has the unique belief-free rationalizable action $(0, S)$, and every expert of type L has the unique belief-free rationalizable action $(1, W)$. Under actions $(0, S)$ for H -type and $(1, W)$ for L -type, the part attains high quality. Therefore every belief-free rationalizable action for every expert and every type yields q_H . Hence the contract uniquely implements high quality.

Finally, under the implemented outcome we have $a_j = \mathbb{1}\{\theta_j = L\}$ for each j . Project success then occurs with probability $P(N)$. Letting $\varepsilon_0, \varepsilon_1 \downarrow 0$ while maintaining $\varepsilon_0 > \varepsilon_1$ yields the infimum cost

$$C^{\text{PtQ}}(\sigma, \mathbf{p}) = P(N) \mathbb{E}_{\mathbf{p}} \left[\sum_{i \in N} \frac{c_i}{P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}) \cup \{i\}) - P(R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}))} \right].$$

□

C.4 Proof of Proposition 4

We say that a contract \mathbf{b} has *PO form* at expert–profile pair (i, \mathbf{a}_{-i}) , if $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$, and *PtQ form* if $b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i})$.

Lemma C.1 (Characterization of UI contracts). *Let \mathbf{b} be a UI contract. Then there exists a perturbation sequence $\{\boldsymbol{\varepsilon}^k\}_{k \geq 1}$ converging componentwise to zero such that, for every k , the perturbed contract $\mathbf{b}^k := \mathbf{b} + \boldsymbol{\varepsilon}^k$ uniquely implements all high quality.*

For each k and each $i \in N$, define the elimination time

$$\tau^k(i) := \min\{r \geq 1 : \text{BFR}_i^{k,r}(\theta_i) = \{(1, W)\}\},$$

where $\text{BFR}_i^{k,r}(\theta_i)$ denotes the surviving action set for type θ_i at round r of belief-free rationalizability in the game induced by \mathbf{b}^k .

For each k , each $i \in N$, and each $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, define the assured set

$$S_i^k(\mathbf{a}_{-i}) := \{j \in N \setminus \{i\} : q(\theta_j, e_j) = q_H \ \forall \theta_j \in \{H, L\} \ \forall (a_j, e_j) \in \text{BFR}_j^{k, \tau^k(i)-1}(\theta_j)\}.$$

Passing to a subsequence if necessary, there exist a map $\tau : N \rightarrow \mathbb{N}$ and sets $S_i(\mathbf{a}_{-i}) \subseteq N \setminus \{i\}$ such that, for every $i \in N$, every $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, and every k in the subsequence,

$$\tau^k(i) = \tau(i) \quad \text{and} \quad S_i^k(\mathbf{a}_{-i}) = S_i(\mathbf{a}_{-i}).$$

Moreover, for every $i \in N$ and every $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, the contract \mathbf{b} satisfies:

(a) If $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$ (PO form), then

$$b_i(1, \mathbf{a}_{-i}) \geq \frac{c_i}{\Delta_i(S_i(\mathbf{a}_{-i}))}. \quad (34)$$

(b) If $b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i})$ (PtQ form), then

$$P(S_i(\mathbf{a}_{-i}) \cup \{i\}) \cdot b_i(1, \mathbf{a}_{-i}) - P(S_i(\mathbf{a}_{-i})) \cdot b_i(0, \mathbf{a}_{-i}) \geq c_i. \quad (35)$$

Lemma C.2 (Fill-in at unreachable profiles). *Let \mathbf{b} be a UI contract with elimination times τ and assured sets $S_i(\mathbf{a}_{-i})$ from Lemma C.1. Let $\sigma \in \text{Perm}(N)$ be any ranking refining τ :*

$$\tau(i) < \tau(j) \implies \sigma(i) < \sigma(j).$$

For each $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, define the σ -enlarged assured set

$$A_i^\sigma(\mathbf{a}_{-i}) := S_i(\mathbf{a}_{-i}) \cup R_{< i}^\sigma,$$

where $R_{< i}^\sigma = \{j \in N : \sigma(j) < \sigma(i)\}$. Define the set of reachable profiles for expert i :

$$\mathcal{P}_i := \{\mathbf{a}_{-i} \in \{0, 1\}^{n-1} : \exists \boldsymbol{\theta}_{-i}, \mathbf{e}_{-i} \text{ s.t. } (a_j, e_j) \in \text{BFR}_j^{\tau(i)-1}(\theta_j) \forall j \neq i\},$$

where the BFR sets are those of the perturbed games \mathbf{b}^k (constant along the subsequence from Lemma C.1).

Define $\bar{\mathbf{b}}$ by: for each $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$,

$$\bar{b}_i(a_i, \mathbf{a}_{-i}) := \begin{cases} b_i(a_i, \mathbf{a}_{-i}) & \text{if } \mathbf{a}_{-i} \in \mathcal{P}_i, \\ \frac{c_i}{\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))} & \text{if } \mathbf{a}_{-i} \notin \mathcal{P}_i \text{ and } a_i = 1, \\ 0 & \text{if } \mathbf{a}_{-i} \notin \mathcal{P}_i, a_i = 0, \text{ and } b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i}), \\ \frac{c_i}{\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))} & \text{if } \mathbf{a}_{-i} \notin \mathcal{P}_i, a_i = 0, \text{ and } b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i}). \end{cases}$$

Then:

1. Contract form preservation: For all i and \mathbf{a}_{-i} ,

$$b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i}) \iff \bar{b}_i(1, \mathbf{a}_{-i}) > \bar{b}_i(0, \mathbf{a}_{-i}).$$

2. Round-1 invariance: The round-1 BFR surviving actions under $\bar{\mathbf{b}}$ are the same as under \mathbf{b} .

3. UI preserved: $\bar{\mathbf{b}}$ is UI.

4. Cost equality: $C(\bar{\mathbf{b}}) = C(\mathbf{b})$.

5. Lower bound: For all i and all $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$,

$$\bar{b}_i(1, \mathbf{a}_{-i}) \geq \frac{c_i}{\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))}. \quad (36)$$

Lemma C.3 (Reduction to sequential critical bonuses). *Let $\bar{\mathbf{b}}$ be the filled-in contract from Lemma C.2. Define $\tilde{\mathbf{b}}^\sigma$ for all $i \in N$ and all $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$ by*

$$\tilde{b}_i^\sigma(1, \mathbf{a}_{-i}) := \frac{c_i}{\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))},$$

and

$$\tilde{b}_i^\sigma(0, \mathbf{a}_{-i}) := \begin{cases} 0 & \text{if } b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i}) \quad (\text{PO form under } \mathbf{b}), \\ \tilde{b}_i^\sigma(1, \mathbf{a}_{-i}) & \text{if } b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i}) \quad (\text{PtQ form under } \mathbf{b}), \end{cases}$$

where the form is inherited from the original contract \mathbf{b} .

Then:

1. Bonus reduction: $\tilde{b}_i^\sigma(1, \mathbf{a}_{-i}) \leq \bar{b}_i(1, \mathbf{a}_{-i})$ for all i and \mathbf{a}_{-i} .
2. Round-1 invariance: The round-1 BFR surviving actions under $\tilde{\mathbf{b}}^\sigma$ are the same as under $\bar{\mathbf{b}}$ (and hence under \mathbf{b}).
3. UI: $\tilde{\mathbf{b}}^\sigma$ is UI.
4. Cost reduction: $C(\tilde{\mathbf{b}}^\sigma) \leq C(\bar{\mathbf{b}}) = C(\mathbf{b})$.

Lemma C.4 (Local PtQ improvement). *Fix a ranking $\sigma \in \text{Perm}(N)$ and the reduced contract $\tilde{\mathbf{b}}^\sigma$ from Lemma C.3. Fix an expert $k \in N$. Define the contract $\hat{\mathbf{b}}^{\sigma, k}$ that modifies $\tilde{\mathbf{b}}^\sigma$ by switching expert k to unconditional PtQ. For all $i \in N$ and $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, set*

$$\hat{b}_i^{\sigma, k}(1, \mathbf{a}_{-i}) := \frac{c_i}{\Delta_i(\hat{A}_i^{\sigma, k}(\mathbf{a}_{-i}))},$$

and

$$\hat{b}_i^{\sigma, k}(0, \mathbf{a}_{-i}) := \begin{cases} \hat{b}_i^{\sigma, k}(1, \mathbf{a}_{-i}) + \varepsilon_0 & \text{if } i = k, \\ 0 & \text{if } i \neq k \text{ and } b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i}), \\ \hat{b}_i^{\sigma, k}(1, \mathbf{a}_{-i}) & \text{if } i \neq k \text{ and } b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i}), \end{cases}$$

where the assured set is

$$\hat{A}_i^{\sigma, k}(\mathbf{a}_{-i}) := R_{< i}^\sigma \cup W_{-i}^k(\mathbf{a}_{-i}),$$

and $W_{-i}^k(\mathbf{a}_{-i})$ is the set of experts $j \neq i$ with $\sigma(j) > \sigma(i)$ and $a_j = 1$ who use PtQ form at all (j, \mathbf{a}_{-j}) under $\hat{\mathbf{b}}^{\sigma,k}$. Concretely, $j \in W_{-i}^k(\mathbf{a}_{-i})$ if and only if $\sigma(j) > \sigma(i)$, $a_j = 1$, and either $j = k$ or expert j had PtQ form at all (j, \mathbf{a}_{-j}) under \mathbf{b} .

Then:

1. UI: $\hat{\mathbf{b}}^{\sigma,k}$ is UI.
2. Bonus reduction: $\hat{b}_i^{\sigma,k}(1, \mathbf{a}_{-i}) \leq \tilde{b}_i^\sigma(1, \mathbf{a}_{-i})$ for all i and \mathbf{a}_{-i} .
3. Cost reduction: $C(\hat{\mathbf{b}}^{\sigma,k}) \leq C(\tilde{\mathbf{b}}^\sigma)$.
4. Strict improvement: If expert k did not already have PtQ form at all (k, \mathbf{a}_{-k}) under \mathbf{b} and $\sigma(k) > 1$, then $C(\hat{\mathbf{b}}^{\sigma,k}) < C(\tilde{\mathbf{b}}^\sigma)$.

Proof of Proposition 4. Let \mathbf{b} be an arbitrary UI contract.

Step 1: Characterize. By Lemma C.1, extract elimination times τ and assured sets $S_i(\mathbf{a}_{-i})$.

Step 2: Fill in unreachable profiles. Choose any ranking σ refining τ . By Lemma C.2, construct $\bar{\mathbf{b}}$ with $C(\bar{\mathbf{b}}) = C(\mathbf{b})$, the same round-1 elimination, and $\bar{b}_i(1, \mathbf{a}_{-i}) \geq c_i/\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))$ everywhere.

Step 3: Reduce to critical bonuses along σ . By Lemma C.3, construct $\tilde{\mathbf{b}}^\sigma$ with

$$C(\tilde{\mathbf{b}}^\sigma) \leq C(\bar{\mathbf{b}}) = C(\mathbf{b}).$$

The contract $\tilde{\mathbf{b}}^\sigma$ is UI, preserves the form from \mathbf{b} , and uses the critical bonus $c_i/\Delta_i(A_i^\sigma(\mathbf{a}_{-i}))$ at every (i, \mathbf{a}_{-i}) .

Step 4: Switch all experts to PtQ. Apply Lemma C.4 sequentially to each expert $k \in N$, switching k to unconditional PtQ. Each application preserves UI and weakly reduces cost.

After switching all n experts, the resulting contract assigns PtQ form at every (i, \mathbf{a}_{-i}) . Under this all-PtQ contract, every participant's action certifies q_H after round 1. The assured set for expert i at round $\sigma(i) + 1$ becomes

$$R_{<i}^\sigma \cup W_{-i}(\mathbf{a}_{-i}),$$

where $W_{-i}(\mathbf{a}_{-i}) = \{j \neq i : a_j = 1\}$ (since every expert now has PtQ form). The critical bonus is $b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i})$, and

$$C^{\text{PtQ}}(\sigma, \mathbf{p}) \leq C(\tilde{\mathbf{b}}^\sigma) \leq C(\bar{\mathbf{b}}) = C(\mathbf{b}).$$

Step 5: Optimize over rankings. Since \mathbf{b} was an arbitrary UI contract and σ was an arbitrary choice

$$\inf_{\mathbf{b} \text{ UI}} C(\mathbf{b}) \geq \min_{\sigma \in \text{Perm}(N)} C^{\text{PtQ}}(\sigma, \mathbf{p}).$$

By Proposition 3, for every ranking σ , the all-PtQ contract with critical bonus $b_i^{\text{PtQ}}(\sigma, \mathbf{a}_{-i})$ achieves UI at cost $C^{\text{PtQ}}(\sigma, \mathbf{p})$. Hence the minimum is attained, and an optimal UI contract is obtained by choosing

$$\sigma^* \in \underset{\sigma \in \text{Perm}(N)}{\text{argmin}} C^{\text{PtQ}}(\sigma, \mathbf{p})$$

and offering the all-PtQ contract with ranking σ^* and vanishing perturbations $\varepsilon_0 > \varepsilon_1 > 0$ that satisfies the construction in Proposition 3. \square

C.5 Proof of Proposition 5

Proof of Proposition 5. Write

$$\rho_i := \frac{c_i}{1 - \lambda_i}, \quad \kappa_i := (1 - p_i) + \frac{p_i}{\lambda_i}, \quad \phi_i := \kappa_i - 1 = \frac{p_i(1 - \lambda_i)}{\lambda_i}.$$

By the O-ring cost formula, the expected cost under ranking σ is

$$C_{\mathbf{p}}^{\text{PtQ}}(\sigma) = \sum_{i \in N} \rho_i \prod_{\sigma(j) > \sigma(i)} \kappa_j.$$

Consider two adjacent experts i and j in the ranking, with a common suffix factor K from all experts ranked after both. Placing i before j costs $K(\rho_i \kappa_j + \rho_j)$; placing j before i costs $K(\rho_j \kappa_i + \rho_i)$. Hence i before j is weakly better if and only if

$$\rho_i \kappa_j + \rho_j \leq \rho_j \kappa_i + \rho_i,$$

which simplifies to $\rho_i(\kappa_j - 1) \leq \rho_j(\kappa_i - 1)$, i.e.,

$$\rho_i \phi_j \leq \rho_j \phi_i \iff \frac{\rho_i}{\phi_i} \leq \frac{\rho_j}{\phi_j} \iff I_i \leq I_j.$$

Every adjacent inversion of the index order can be removed by a weakly cost-reducing swap. Repeating finitely many times, a ranking is optimal if and only if $\sigma(i) < \sigma(j)$ whenever $I_i < I_j$. \square

C.6 Proof of Lemmas C.1, C.2, C.3, and C.4

Proof of Lemma C.1. Since \mathbf{b} is UI, there exists a perturbation sequence $\{\varepsilon^k\}$ converging to zero such that \mathbf{b}^k uniquely implements high quality for every k . In particular, $\tau^k(i)$ is well-defined and finite for every i and k .

Step 1: Finding the subsequence. For each i , the sequence $\{\tau^k(i)\}_{k \geq 1}$ takes values in a finite set (BFR terminates in finitely many rounds). For each pair (i, \mathbf{a}_{-i}) , the set

$S_i^k(\mathbf{a}_{-i}) \subseteq N \setminus \{i\}$ also takes finitely many values. Since there are finitely many experts and profiles, a diagonal argument yields a subsequence (still indexed by k) along which $\tau^k(i) = \tau(i)$ and $S_i^k(\mathbf{a}_{-i}) = S_i(\mathbf{a}_{-i})$ for all i , \mathbf{a}_{-i} , and k .

Step 2: Worst-case belief. Fix $i \in N$, $\mathbf{a}_{-i} \in \{0, 1\}^{n-1}$, and any k in the subsequence. By definition of $S_i(\mathbf{a}_{-i})$:

- For $j \in S_i(\mathbf{a}_{-i})$: every surviving type–action pair at round $\tau(i) - 1$ consistent with a_j yields q_H .
- For $j \notin S_i(\mathbf{a}_{-i})$: there exists a surviving type–action pair at round $\tau(i) - 1$ consistent with a_j that yields q_L .

Choosing one such pair for each j gives a degenerate belief ν^k , supported on round- $(\tau(i) - 1)$ survivors and consistent with \mathbf{a}_{-i} , under which exactly the parts in $S_i(\mathbf{a}_{-i})$ attain q_H .

Step 3: Best-response inequality. Since $\text{BFR}_i^{k, \tau(i)}(L) = \{(1, W)\}$, the action $(1, W)$ must weakly dominate both $(1, S)$ and $(0, S)$ under the belief ν^k . Hence

$$P(S_i(\mathbf{a}_{-i}) \cup \{i\}) \cdot b_i^k(1, \mathbf{a}_{-i}) - c_i \geq \max\left\{P(S_i(\mathbf{a}_{-i})) \cdot b_i^k(1, \mathbf{a}_{-i}), P(S_i(\mathbf{a}_{-i})) \cdot b_i^k(0, \mathbf{a}_{-i})\right\}. \quad (37)$$

Step 4: Limit. Taking $k \rightarrow \infty$ in (37) gives

$$P(S_i(\mathbf{a}_{-i}) \cup \{i\}) \cdot b_i(1, \mathbf{a}_{-i}) - c_i \geq \max\left\{P(S_i(\mathbf{a}_{-i})) \cdot b_i(1, \mathbf{a}_{-i}), P(S_i(\mathbf{a}_{-i})) \cdot b_i(0, \mathbf{a}_{-i})\right\}. \quad (38)$$

Step 5: Two cases.

Case (a): $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$ (PO form). The maximum in (38) is the first term, giving $\Delta_i(S_i(\mathbf{a}_{-i})) \cdot b_i(1, \mathbf{a}_{-i}) \geq c_i$.

Case (b): $b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i})$ (PtQ form). The maximum is the second term, giving (35) directly. \square

Proof of Lemma C.2. (i) At reachable profiles, $\bar{\mathbf{b}} = \mathbf{b}$, so the form is trivially preserved. At unreachable profiles: if $b_i(1, \mathbf{a}_{-i}) > b_i(0, \mathbf{a}_{-i})$ (PO form), then $\bar{b}_i(1, \mathbf{a}_{-i}) = c_i / \Delta_i(A_i^\sigma) > 0 = \bar{b}_i(0, \mathbf{a}_{-i})$, preserving PO form. If $b_i(1, \mathbf{a}_{-i}) \leq b_i(0, \mathbf{a}_{-i})$ (PtQ form), then $\bar{b}_i(1, \mathbf{a}_{-i}) = \bar{b}_i(0, \mathbf{a}_{-i}) = c_i / \Delta_i(A_i^\sigma)$, so $\bar{b}_i(1, \mathbf{a}_{-i}) \leq \bar{b}_i(0, \mathbf{a}_{-i})$, preserving PtQ form.

(ii) The round-1 elimination depends on dominance comparisons among $\{(1, W), (1, S), (0, S)\}$. For L -types, the comparison between $(1, S)$ and $(0, S)$ depends on whether there exists *any* \mathbf{a}_{-i} at which $(1, S)$ is preferred to $(0, S)$ (i.e., $b_i(1, \cdot) > b_i(0, \cdot)$, PO form) or vice versa ($b_i(0, \cdot) > b_i(1, \cdot)$, strict PtQ form). By (i), the set of PO-form and PtQ-form profiles is the same under $\bar{\mathbf{b}}$ and \mathbf{b} . Hence:

- If \mathbf{b} has PO form at every \mathbf{a}_{-i} for expert i : $(1, S)$ dominates $(0, S)$ for L -types. Same under $\bar{\mathbf{b}}$.
- If \mathbf{b} has strict PtQ form at every \mathbf{a}_{-i} : $(0, S)$ dominates $(1, S)$. Same under $\bar{\mathbf{b}}$.
- If the form varies across \mathbf{a}_{-i} : neither dominates the other. Same under $\bar{\mathbf{b}}$.

The analogous argument applies for H -types. Hence the round-1 surviving actions are identical.

(iii) Since round-1 survivors are the same and $\bar{\mathbf{b}} = \mathbf{b}$ at all reachable profiles \mathcal{P}_i , the BFR elimination under the perturbed contract $\bar{\mathbf{b}}^k := \bar{\mathbf{b}} + \varepsilon^k$ proceeds identically to that under \mathbf{b}^k at all reachable profiles.

Formally: the best-response condition at round $\tau(i)$ requires $(1, W)$ to be a best response to beliefs supported on round- $(\tau(i) - 1)$ surviving profiles. These beliefs generate profiles in \mathcal{P}_i , at which $\bar{\mathbf{b}}^k = \mathbf{b}^k$. Hence the same best-response conditions hold, and $\text{BFR}_i^{k, \tau(i)}(L) = \{(1, W)\}$ under $\bar{\mathbf{b}}^k$.

(iv)

L-types. Under both \mathbf{b} and $\bar{\mathbf{b}}$, every L -type is pinned to $(1, W)$ by the UI property. Hence $a_i = 1$ and $e_i = W$ under both contracts.

H-types. Under \mathbf{b} , expert i 's H -type may have multiple surviving actions when the contract form varies across \mathbf{a}_{-i} (mixed form). In that case, the cost $C(\mathbf{b})$ depends on which H -type action is selected in the outcome. The same applies to $\bar{\mathbf{b}}$.

However, H -type actions matter for cost only through the realized payment $b_i(a_i, \mathbf{a}_{-i}(\boldsymbol{\theta}))$. At every realized profile $\mathbf{a}_{-i}(\boldsymbol{\theta}) \in \mathcal{P}_i$, we have $\bar{b}_i(a_i, \mathbf{a}_{-i}) = b_i(a_i, \mathbf{a}_{-i})$ for both values $a_i \in \{0, 1\}$, by construction.

Moreover, the set of H -type surviving actions is identical under $\bar{\mathbf{b}}$ and \mathbf{b} . This is because:

1. Round-1 survivors are the same by part (ii).
2. At every subsequent BFR round $r \leq \tau(i) - 1$, the best-response conditions for H -types depend only on payoffs at round- $(r - 1)$ surviving profiles. These profiles generate participation vectors in \mathcal{P}_i , at which $\bar{\mathbf{b}} = \mathbf{b}$. Hence the dominance comparisons are identical and the surviving sets evolve identically.

Therefore, for every type profile $\boldsymbol{\theta}$ and every selection of H -type actions from the surviving set, the resulting participation profile $\mathbf{a}(\boldsymbol{\theta})$ is the same under both contracts, and the realized payment $b_i(a_i, \mathbf{a}_{-i}(\boldsymbol{\theta})) = \bar{b}_i(a_i, \mathbf{a}_{-i}(\boldsymbol{\theta}))$ is the same for every expert i . Summing over experts and taking expectations over $\boldsymbol{\theta}$:

$$C(\bar{\mathbf{b}}) = C(\mathbf{b}).$$

This holds regardless of which selection among the H -type surviving actions defines the outcome: every selection yields the same cost under both contracts.

(v) At reachable profiles: $A_i^\sigma(\mathbf{a}_{-i}) \supseteq S_i(\mathbf{a}_{-i})$, so $\Delta_i(A_i^\sigma) \geq \Delta_i(S_i)$. By Remark ??, $b_i(1, \mathbf{a}_{-i}) \geq c_i/\Delta_i(S_i) \geq c_i/\Delta_i(A_i^\sigma)$. Hence $\bar{b}_i(1, \mathbf{a}_{-i}) = b_i(1, \mathbf{a}_{-i}) \geq c_i/\Delta_i(A_i^\sigma)$.

At unreachable profiles: $\bar{b}_i(1, \mathbf{a}_{-i}) = c_i/\Delta_i(A_i^\sigma)$ by construction. \square

Proof of Lemma C.3. (i) Immediate from Lemma C.2(v): $\bar{b}_i(1, \mathbf{a}_{-i}) \geq c_i/\Delta_i(A_i^\sigma(\mathbf{a}_{-i})) = \tilde{b}_i^\sigma(1, \mathbf{a}_{-i})$.

(ii) The form is preserved: at PO-form (i, \mathbf{a}_{-i}) , $\tilde{b}_i^\sigma(1, \cdot) = c_i/\Delta_i(A_i^\sigma) > 0 = \tilde{b}_i^\sigma(0, \cdot)$. At PtQ-form (i, \mathbf{a}_{-i}) , $\tilde{b}_i^\sigma(0, \cdot) = \tilde{b}_i^\sigma(1, \cdot)$ (with strict inequality supplied by perturbation). Since the set of PO-form and PtQ-form pairs is the same as under $\bar{\mathbf{b}}$ and \mathbf{b} , the round-1 surviving actions are identical by the same argument as in Lemma C.2(ii).

(iii) We verify UI by running BFR elimination on $\tilde{\mathbf{b}}^\sigma + \varepsilon$ directly along σ , where the perturbation is chosen as in Proposition 3: $\varepsilon_1 = r\varepsilon_0$ with $r \in (\bar{r}, 1)$ and $\bar{r} = \max_{i,Q} P(Q)/P(Q \cup \{i\}) < 1$.

Round 1. By (ii), the round-1 surviving actions are the same as under \mathbf{b} . For each expert i :

- H -types: every surviving action yields q_H (since $q(H, e) = q_H$ for all e).
- L -types: the surviving set is $\{(0, S), (1, W)\}$ (all PtQ), $\{(1, S), (1, W)\}$ (all PO), or $\{(0, S), (1, S), (1, W)\}$ (mixed form).

Certification claim.

Claim 1. For every $i \in N$ and every $j \in S_i(\mathbf{a}_{-i}) \setminus R_{<i}^\sigma$: at round $\sigma(i) + 1$ of the BFR elimination under $\tilde{\mathbf{b}}^\sigma + \varepsilon$, every surviving action for expert j consistent with a_j yields q_H .

Proof of Claim. Since $j \in S_i(\mathbf{a}_{-i}) \setminus R_{<i}^\sigma$, we have $\sigma(j) > \sigma(i)$, hence $\tau(j) \geq \tau(i)$ (since σ refines τ).

If $a_j = 0$: The action $(0, S)$ is consistent with $a_j = 0$ and yields q_L for L -types. For $j \in S_i(\mathbf{a}_{-i})$, we need $(0, S)$ to not survive for j 's L -type at round $\tau(i) - 1$ under \mathbf{b}^k . But $(0, S)$ for L -type j is eliminated only at round $\tau(j)$ (when $(1, W)$ strictly dominates it). Since $\tau(j) \geq \tau(i)$, the action $(0, S)$ survives at round $\tau(i) - 1$. Contradiction. Hence $a_j = 0$ cannot occur.

If $a_j = 1$: The action $(1, S)$ is consistent with $a_j = 1$ and yields q_L for L -types. For $j \in S_i(\mathbf{a}_{-i})$, the action $(1, S)$ must not survive for j 's L -type at round $\tau(i) - 1$ under \mathbf{b}^k . Since $\tau(j) \geq \tau(i)$, expert j 's L -type is not yet pinned to $(1, W)$ at that round. The only way $(1, S)$ can be absent from the surviving set is if it was eliminated in round 1. This occurs if and only if \mathbf{b} has PtQ form at every (j, \mathbf{a}_{-j}) (so that $(0, S)$ strictly dominates

$(1, S)$ globally). Since $\tilde{\mathbf{b}}^\sigma$ preserves the form, the round-1 elimination is the same by (ii). Hence $(1, S)$ is also eliminated in round 1 under $\tilde{\mathbf{b}}^\sigma$. Since surviving sets only shrink over subsequent rounds, $(1, S)$ remains absent at round $\sigma(i) + 1 \geq 2$. \square

Induction along σ . We prove that for each $m = 1, \dots, n$, the L -type of expert $i_m := \sigma^{-1}(m)$ is pinned to $(1, W)$ at round $m + 1$.

Suppose the claim holds for all predecessors of i_m in σ . Fix $\mathbf{a}_{-i_m} \in \{0, 1\}^{n-1}$ and any belief supported on the round- $(m+1)$ surviving type-action profiles consistent with \mathbf{a}_{-i_m} . Let $Q \subseteq N \setminus \{i_m\}$ be the set of other parts at q_H under this belief.

By the induction hypothesis, $R_{<i_m}^\sigma \subseteq Q$. By Claim 1, for each $j \in S_{i_m}(\mathbf{a}_{-i_m}) \setminus R_{<i_m}^\sigma$, every surviving action consistent with a_j yields q_H , so $j \in Q$. Hence $Q \supseteq A_{i_m}^\sigma(\mathbf{a}_{-i_m})$.

We consider the two cases for the form at (i_m, \mathbf{a}_{-i_m}) .

Case 1: PO form. The q_L -inducing action is $(1, S)$. The payoff difference between $(1, W)$ and $(1, S)$ is

$$\tilde{b}_{i_m}^\sigma(1, \mathbf{a}_{-i_m}) \cdot \Delta_{i_m}(Q) - c_{i_m} \geq \frac{c_{i_m}}{\Delta_{i_m}(A_{i_m}^\sigma)} \cdot \Delta_{i_m}(A_{i_m}^\sigma) - c_{i_m} = 0,$$

using $\Delta_{i_m}(Q) \geq \Delta_{i_m}(A_{i_m}^\sigma)$ (supermodularity and $Q \supseteq A_{i_m}^\sigma$). With the perturbation ($\varepsilon > 0$ added to the bonus), the inequality is strict.

Case 2: PtQ form. The q_L -inducing action is $(0, S)$. Write $m_i := \tilde{b}_{i_m}^\sigma(1, \mathbf{a}_{-i_m}) = c_{i_m}/\Delta_{i_m}(A_{i_m}^\sigma)$. The payoff difference between $(1, W)$ and $(0, S)$ is

$$\begin{aligned} & (m_i + \varepsilon_1) P(Q \cup \{i_m\}) - (m_i + \varepsilon_0) P(Q) - c_{i_m} \\ & = m_i \cdot \Delta_{i_m}(Q) - c_{i_m} + \varepsilon_1 P(Q \cup \{i_m\}) - \varepsilon_0 P(Q). \end{aligned}$$

The first two terms are nonnegative (as in Case 1). The last two are strictly positive since $\varepsilon_1 = r\varepsilon_0$ and $r > \bar{r} \geq P(Q)/P(Q \cup \{i_m\})$.

In both cases, the q_L -inducing action is eliminated. The induction is complete.

After $n + 1$ rounds, every L -type is pinned to $(1, W)$ and every H -type's surviving actions yield q_H . Hence $\tilde{\mathbf{b}}^\sigma$ is UI.

(iv) Under both $\bar{\mathbf{b}}$ and $\tilde{\mathbf{b}}^\sigma$, the form is preserved, so the participation mapping $\boldsymbol{\theta} \mapsto \mathbf{a}(\boldsymbol{\theta})$ is the same at the implemented outcome. For every type profile $\boldsymbol{\theta}$ and every expert i :

L-type ($\theta_i = L$): Expert i participates ($a_i = 1$) under both contracts and is paid $b_i(1, \mathbf{a}_{-i})$. By (i), $\tilde{b}_i^\sigma(1, \cdot) \leq \bar{b}_i(1, \cdot)$.

H-type ($\theta_i = H$) with PO form at all \mathbf{a}_{-i} : Expert i participates and is paid $b_i(1, \mathbf{a}_{-i})$. By (i), $\tilde{b}_i^\sigma(1, \cdot) \leq \bar{b}_i(1, \cdot)$.

H-type with PtQ form at all \mathbf{a}_{-i} : Expert i opts out ($a_i = 0$) and is paid $b_i(0, \mathbf{a}_{-i})$. We have $\tilde{b}_i^\sigma(0, \cdot) = \tilde{b}_i^\sigma(1, \cdot) \leq \bar{b}_i(1, \cdot) \leq \bar{b}_i(0, \cdot)$, where the last inequality holds because $\bar{\mathbf{b}}$ has PtQ form (so $\bar{b}_i(0, \cdot) \geq \bar{b}_i(1, \cdot)$).

H-type with mixed form: Expert i 's surviving actions all yield q_H and she is paid $b_i(a_i, \mathbf{a}_{-i})$ for some a_i . At PO-form (i, \mathbf{a}_{-i}) : $\tilde{b}_i^\sigma(0, \cdot) = 0 \leq \bar{b}_i(0, \cdot)$ and $\tilde{b}_i^\sigma(1, \cdot) \leq \bar{b}_i(1, \cdot)$ by (i). At PtQ-form (i, \mathbf{a}_{-i}) : $\tilde{b}_i^\sigma(0, \cdot) = \tilde{b}_i^\sigma(1, \cdot) \leq \bar{b}_i(1, \cdot) \leq \bar{b}_i(0, \cdot)$. In either case, $\tilde{b}_i^\sigma(a_i, \cdot) \leq \bar{b}_i(a_i, \cdot)$ regardless of a_i .

Summing over experts and taking expectations: $C(\tilde{\mathbf{b}}^\sigma) \leq C(\bar{\mathbf{b}}) = C(\mathbf{b})$. \square

Proof of Lemma C.4. (i) The UI verification proceeds by BFR elimination along σ , exactly as in Lemma C.3(iii).

Round 1. Expert k now has PtQ form at all (k, \mathbf{a}_{-k}) . For k 's L -type, $(1, S)$ is eliminated because $(0, S)$ weakly dominates it at every \mathbf{a}_{-k} (both yield q_L , but $\hat{b}_k(0, \cdot) > \hat{b}_k(1, \cdot)$, giving a strictly higher bonus). Hence $\text{BFR}_k^1(L) = \{(0, S), (1, W)\}$. For all other experts $i \neq k$, the form is unchanged from $\tilde{\mathbf{b}}^\sigma$, so their round-1 survivors are unchanged.

Certification. For any expert $j \in W_{-i}^k(\mathbf{a}_{-i})$: expert j uses PtQ form at all (j, \mathbf{a}_{-j}) under $\hat{\mathbf{b}}^{\sigma, k}$, so $(1, S)$ is eliminated in round 1 for j 's L -type. Hence $a_j = 1$ can only be consistent with $(1, W)$, yielding q_H .

Induction along σ . At round $\sigma(i) + 1$, the assured set contains $R_{<i}^\sigma$ (predecessors, by induction) and $W_{-i}^k(\mathbf{a}_{-i})$ (certified participants). Since the bonus is at the critical level $c_i/\Delta_i(\hat{A}_i^{\sigma, k})$, the perturbation makes the desired action strictly dominant, by the identical calculation as in Lemma C.3(iii).

(ii) We show $\hat{A}_i^{\sigma, k}(\mathbf{a}_{-i}) \supseteq A_i^\sigma(\mathbf{a}_{-i})$, i.e.,

$$S_i(\mathbf{a}_{-i}) \cup R_{<i}^\sigma \subseteq R_{<i}^\sigma \cup W_{-i}^k(\mathbf{a}_{-i}). \quad (39)$$

Take $j \in S_i(\mathbf{a}_{-i}) \setminus R_{<i}^\sigma$. By Claim 1 (from Lemma C.3), $\sigma(j) > \sigma(i)$, $a_j = 1$, and $(1, S)$ was eliminated for j 's L -type in round 1 under \mathbf{b} , which means j had PtQ form at all (j, \mathbf{a}_{-j}) under \mathbf{b} .

If $j = k$: expert k has PtQ form at all (k, \mathbf{a}_{-k}) under $\hat{\mathbf{b}}^{\sigma, k}$ by construction, so $k \in W_{-i}^k(\mathbf{a}_{-i})$.

If $j \neq k$: expert j already had PtQ form at all (j, \mathbf{a}_{-j}) under \mathbf{b} , so $j \in W_{-i}^k(\mathbf{a}_{-i})$ by definition.

Hence (39) holds, $\Delta_i(\hat{A}_i^{\sigma, k}) \geq \Delta_i(A_i^\sigma)$, and $\hat{b}_i^{\sigma, k}(1, \cdot) \leq \tilde{b}_i^\sigma(1, \cdot)$.

(iii) The cost comparison proceeds as in Lemma C.3(iv). For experts $i \neq k$: the form is unchanged, and the payment satisfies $\hat{b}_i^{\sigma, k}(a_i, \cdot) \leq \tilde{b}_i^\sigma(a_i, \cdot)$ for both values of a_i .

For $a_i = 1$: by (ii).

For $a_i = 0$ at PO-form (i, \mathbf{a}_{-i}) : $\hat{b}_i^{\sigma, k}(0, \cdot) = 0 = \tilde{b}_i^\sigma(0, \cdot)$.

For $a_i = 0$ at PtQ-form (i, \mathbf{a}_{-i}) : $\hat{b}_i^{\sigma, k}(0, \cdot) = \hat{b}_i^{\sigma, k}(1, \cdot) \leq \tilde{b}_i^\sigma(1, \cdot) = \tilde{b}_i^\sigma(0, \cdot)$.

Key observation for expert k : realized payments under $\tilde{\mathbf{b}}^\sigma$. Under $\tilde{\mathbf{b}}^\sigma$, expert k 's realized payment is $\tilde{b}_k^\sigma(1, \mathbf{a}_{-k})$ regardless of type: at PO-form pairs H -type participates and pays

$\tilde{b}_k^\sigma(1, \cdot)$; at PtQ-form pairs H -type opts out and pays $\tilde{b}_k^\sigma(0, \cdot) = \tilde{b}_k^\sigma(1, \cdot)$; at mixed-form pairs both surviving H -type actions yield $\tilde{b}_k^\sigma(1, \cdot)$. L -type participates and pays $\tilde{b}_k^\sigma(1, \cdot)$.

For expert k : L -type participates and is paid $\hat{b}_k^{\sigma,k}(1, \cdot) \leq \tilde{b}_k^\sigma(1, \cdot)$ by (ii). H -type opts out under $\hat{\mathbf{b}}^{\sigma,k}$ and is paid $\hat{b}_k^{\sigma,k}(0, \cdot) = \hat{b}_k^{\sigma,k}(1, \cdot) + \varepsilon_0$, which in the limit equals $\hat{b}_k^{\sigma,k}(1, \cdot) \leq \tilde{b}_k^\sigma(1, \cdot)$. The expected payment for expert k is:

$$\begin{aligned} \text{Under } \tilde{\mathbf{b}}^\sigma : \quad & \mathbb{E}[(1 - p_k) \tilde{b}_k^\sigma(1, \mathbf{a}_{-k}) + p_k \tilde{b}_k^\sigma(a_k^H, \mathbf{a}_{-k})] = \mathbb{E}[\tilde{b}_k^\sigma(1, \mathbf{a}_{-k})], \\ \text{Under } \hat{\mathbf{b}}^{\sigma,k} : \quad & \mathbb{E}[(1 - p_k) \hat{b}_k^{\sigma,k}(1, \mathbf{a}_{-k}) + p_k \hat{b}_k^{\sigma,k}(1, \mathbf{a}_{-k})] = \mathbb{E}[\hat{b}_k^{\sigma,k}(1, \mathbf{a}_{-k})], \end{aligned}$$

in the limit ($\varepsilon_0 \rightarrow 0$), where a_k^H is the H -type's participation choice under $\tilde{\mathbf{b}}^\sigma$. The first equality uses $\tilde{b}_k^\sigma(a_k^H, \cdot) = \tilde{b}_k^\sigma(1, \cdot)$ (the key observation above). Since $\hat{b}_k^{\sigma,k}(1, \cdot) \leq \tilde{b}_k^\sigma(1, \cdot)$ by (ii),

$$\mathbb{E}[\hat{b}_k^{\sigma,k}(1, \mathbf{a}_{-k})] \leq \mathbb{E}[\tilde{b}_k^\sigma(1, \mathbf{a}_{-k})].$$

Participation profiles of other experts are the same. Expert k was not globally PtQ under \mathbf{b} , so $k \notin S_i(\mathbf{a}_{-i})$ for any $i \neq k$ (by Claim 1). Hence $A_i^\sigma(\mathbf{a}_{-i})$ is invariant to a_k , and the bonuses $\tilde{b}_i^\sigma(1, \mathbf{a}_{-i})$ do not depend on a_k . The change in k 's H -type action (from possibly participating to always opting out) therefore does not alter the participation profiles or realized payments of other experts.

Thus $C(\hat{\mathbf{b}}^{\sigma,k}) \leq C(\tilde{\mathbf{b}}^\sigma)$.

(iv) Suppose k did not have PtQ form at all (k, \mathbf{a}_{-k}) under \mathbf{b} and $\sigma(k) > 1$. Then there exists a predecessor j with $\sigma(j) < \sigma(k)$.

Under $\tilde{\mathbf{b}}^\sigma$: since k did not have PtQ at all (k, \mathbf{a}_{-k}) , $(1, S)$ survived round 1 for k 's L -type. Hence $k \notin W_{-j}^k(\mathbf{a}_{-j})$ before the switch: k 's participation did not certify q_H .

After switching: $k \in W_{-j}^k(\mathbf{a}_{-j})$ whenever $a_k = 1$. Since $\theta_k = L$ (hence $a_k = 1$) with probability $1 - p_k > 0$, predecessor j 's assured set strictly expands on a positive-probability event. By strict supermodularity, Δ_j strictly increases, j 's critical bonus strictly decreases, and j 's expected payment strictly falls. Hence $C(\hat{\mathbf{b}}^{\sigma,k}) < C(\tilde{\mathbf{b}}^\sigma)$. \square