

CONTRACTING WITH ENDOGENOUSLY INCOMPLETE COMMITMENT: ESCAPE CLAUSES—APPENDIX*

Thomas P. Tangerås[†] and Wolfgang H. Gick[‡]

October 12, 2026

Abstract

This appendix contains the complementary proofs of formal statements in Tangerås and Gick (2026).

JEL classification: *D82, D84, D86.*

Keywords: *Constrained contracting, endogenously incomplete commitment, escape clauses, ratchet effect, stochastic revelation principle.*

*Financial support from Jan Wallanders och Tom Hedelius stiftelse (P15-0305) is gratefully acknowledged.

[†]Department of Business and Mathematics, Mälardalen University, P.O. Box 883, 72123 Västerås, Sweden, e-mail: thomas.tangeras@mdu.se. Affiliated with the Research Institute of Industrial Economics (IFN), Stockholm.

[‡]Faculty of Economics and Management, Free University of Bozen/Bolzano, Universitätsplatz 1, 39100 Bozen, Italy, e-mail: wolfgang.gick@unibz.it. Affiliated with IFN, Stockholm.

1 Introduction

This appendix contains the complementary proofs of formal statements in Tangerås and Gick (2026). We restate the model formulation in Section 2 to keep our appendix self-contained. Incomplete commitment causes the standard revelation principle to break down. We verify the applicability of the stochastic revelation principle developed by Bester and Strausz (2001) to our setting in Section 2. In that section, we also establish five claims about incentive feasible and incentive efficient mechanisms. The results in Section 3, Section 4 and the discussion Section 5 appear in the same order as in Tangerås and Gick (2026) to facilitate comparison with the main article. There are some extensions. In Section 4.1, we provide a complete characterization of incentive efficient mechanisms with complete commitment. Lemma 3 in Tangerås and Gick (2026) derives fundamental properties of incentive optimal reporting strategies under simplifying assumptions. In Section 4.2, we provide a characterization for the general case.

2 The contracting problem

The agent can be one of a finite number $I \geq 2$ of types. An agent of type $i \in \{1, 2, \dots, I\} = \mathcal{I}$ has constant marginal production cost $0 < \theta_i < \infty$. Types are ranked in order of increasing marginal production cost: $\theta_{i+1} > \theta_i$ for all $i \in \{1, \dots, I-1\}$. Let $\boldsymbol{\nu} = (\nu_1, \dots, \nu_i, \dots, \nu_I)$ be the probability distribution over the set of possible types $\boldsymbol{\theta} = (\theta_1, \dots, \theta_i, \dots, \theta_I)$, with $\nu_i > 0$ for all $i \in \mathcal{I}$, and $\sum_{i=1}^I \nu_i = 1$. To simplify indexation, we define a null type $\theta_0 \in [0, \theta_1)$ that occurs with probability $\nu_0 = 0$. Also, we let $G_i = \sum_{j=0}^i \nu_j$ be the probability that the agent has marginal production cost less than or equal to θ_i .

A contract $x = (q, t)$ is a pair specifying a quantity $q \geq 0$ the agent has to produce and an associated transfer of $t \geq 0$ from the principal to the agent. An agent with marginal cost θ_i operating under contract x obtains the rent

$$U_i(x) = t - \theta_i q.$$

The principal achieves the corresponding surplus of

$$W_i(x) = S(q) - t + \alpha U_i(x) = S(q) - \theta_i q - (1 - \alpha)U_i(x)$$

under contract x , where $S(q)$ is the principal's utility function of output q . This function is twice continuously differentiable and strictly concave, with $S(0) = 0$. The parameter $\alpha \in (0, 1)$ in the principal's objective function reflects the weight the principal attaches to the rent of the agent. We assume that the outside no-contract option has a value of

zero both to the principal and the agent and that agent participation is voluntary.

We employ the standard regularity assumption

$$\theta_i + \frac{G_{i-1}}{\nu_i}(1 - \alpha)(\theta_i - \theta_{i-1}) < \theta_{i+1} + \frac{G_i}{\nu_{i+1}}(1 - \alpha)(\theta_{i+1} - \theta_i) \quad \forall i \in \{1, \dots, I - 1\} \quad (1)$$

of increasing virtual marginal cost.

First-best contracting For any given output q , the principal wants to minimize the agent's rent by setting the transfer t as small as possible. Under complete information about the agent's marginal production cost, the principal sets $U_i(x) = 0$ and maximizes

$$W_i^{fb}(q) = S(q) - \theta_i q$$

over $q \geq 0$. We assume that $\lim_{q \rightarrow 0} S'(q) > \theta_1$ and $S'(q) < \theta_I$ for some $q > 0$, so that the first-best contract $x_i^{fb} = (q_i^{fb}, t_i^{fb})$ entails strictly positive output and transfer payment:

$$q_i^{fb} = S'^{-1}(\theta_i) > 0, \quad t_i^{fb} = \theta_i q_i^{fb} > 0 \quad \forall i \in \mathcal{I}.$$

Let the first-best contract be strictly better from the principal's point of view than the outside option, $w_i^{fb} = W_i^{fb}(q_i^{fb}) > 0 \quad \forall i \in \mathcal{I}$. The menu $\mathbf{x}^{fb} = (x_1^{fb}, \dots, x_i^{fb}, \dots, x_I^{fb})$ of first-best contracts thus involves *full participation* in the sense that the agent produces a strictly positive output under complete information, regardless of its marginal cost.

Contracting under incomplete information Everything is common knowledge except the agent has private information about its marginal cost θ_i prior to contracting. The principal only knows the distribution characteristics $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$. We limit the total number K of different contracts the principal can offer the agent upfront. We refer to the polar extreme case $K = 1$ as one of *maximally constrained* contracting. We deviate from the standard paradigm also by assuming that the principal can enter into the mechanism offered to the agent a possibility to contract ex post under predefined circumstances. Formally, we analyze the following game between the principal and the agent:

Stage 0: The principal constructs two disjoint subsets $\mathcal{A} \subset \mathcal{I} \cup \emptyset$ and $\mathcal{B} \subset \mathcal{I} \cup \emptyset$ and a subset $\mathcal{C}$ which contains the types not in $\mathcal{A}$ or $\mathcal{B}$. The set $\mathcal{C}$ is empty if $\mathcal{A} \cup \mathcal{B}$ contains all types $\mathcal{I}$.

Stage 1: The principal commits to a menu $\mathbf{x}_{\mathcal{A}} = \{x_j\}_{j \in \mathcal{A}}$ of ex-ante contracts, $x_j = (q_j, t_j) > (0, 0)$ for all $j \in \mathcal{A}$ if $\mathcal{A} \neq \emptyset$, and to $x_j = x_0 = (0, 0)$ for all $j \in \mathcal{C}$ if $\mathcal{C} \neq \emptyset$. The menu $\mathbf{x}_{\mathcal{A}}$ consists of at most K different contracts: $|\mathbf{x}_{\mathcal{A}}| \leq K$.

Stage 2: The agent accepts or rejects the Stage 1 offer.

- Rejection: The principal and the agent receive their reservation utilities. Game over.
- Acceptance: The game continues to the next stage.

Stage 3: The agent reports marginal cost θ_j , $j \in \mathcal{I}$.

- If $\mathcal{A} \neq \emptyset$ and $j \in \mathcal{A}$, then the agent produces q_j in exchange for t_j . Game over.
- If $\mathcal{C} \neq \emptyset$ and $j \in \mathcal{C}$, then the agent receives the *null contract* x_0 . Game over.
- If $\mathcal{B} \neq \emptyset$ and $j \in \mathcal{B}$, then the game continues to the next stage.

Stage 4: The principal offers an *ex-post* contract $x_j = (q_j, t_j)$.

Stage 5: The agent accepts or rejects x_j .

- Rejection: The principal and the agent receive their reservation utilities. Game over.
- Acceptance: The agent produces q_j in exchange for t_j . Game over.

Let $\mathbf{x} = (\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{B}})$, $\mathbf{x}_{\mathcal{B}} = \{x_j\}_{j \in \mathcal{B}}$ be the menu of contracts. The mechanism features *pure ex-ante contracting* if $\mathcal{B} = \emptyset$. The mechanism has *incomplete commitment* if $\mathcal{B} \neq \emptyset$. The mechanism features *pure ex-post contracting* if $\mathcal{A} = \emptyset$ so that the principal does not offer any contract upfront before communicating with the agent.

The reporting strategy of an agent of type $i \in \mathcal{I}$ is a probability distribution $\boldsymbol{\sigma}_i = (\sigma_{1i}, \dots, \sigma_{ji}, \dots, \sigma_{Ii})^T \in \Delta^{I-1}$, where Δ^{I-1} is the $I - 1$ standard simplex. Specifically, $\sigma_{ji} \in [0, 1]$ is the probability that an agent with marginal cost θ_i claims to have marginal cost θ_j . We let $\sigma_i = \sigma_{ii}$ denote the probability that an agent with marginal cost θ_i truthfully reports its marginal cost. Let $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_i, \dots, \boldsymbol{\sigma}_I) \in \Delta^{2(I-1)}$ be the $I \times I$ matrix of reporting probabilities. We call μ_{ji} the posterior probability attached by the principal to the event that the agent has marginal cost θ_i when the agent has reported marginal cost θ_j , and let $\mu_j = \mu_{jj}$ be the posterior belief that the cost report θ_j has been truthful.

By an extension of the terminology in Bester and Strausz (2001) to our environment, a mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ is *incentive feasible* given $(\mathcal{A}, \mathcal{B})$ if it meets the following conditions:

$$U_i(x_i) \geq 0 \quad \forall i \in \mathcal{I} \quad (2)$$

$$U_i(x_i) \geq U_i(x_j) = U_j(x_j) + (\theta_j - \theta_i)q_j \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I} \quad (3)$$

$$\sigma_i > 0, \sigma_{ji}(U_i(x_i) - U_i(x_j)) = 0 \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I} \quad (4)$$

$$x_j \in \arg \max_{x' \in R_+^2} \sum_{i=1}^I \mu_{ji} W_i(x') \quad \forall j \in \mathcal{B} \text{ if } \mathcal{B} \neq \emptyset \quad (5)$$

$$\mu_{ji} = \frac{\nu_i \sigma_{ji}}{\sum_{h=1}^I \nu_h \sigma_{jh}} \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I} \quad (6)$$

$$|\mathbf{x}_{\mathcal{A}}| \leq K \quad (7)$$

Constraints (2) and (3) are the standard individual rationality and incentive compatibility constraints. The ex-post menu $\mathbf{x}_{\mathcal{B}}$ and agent reporting strategy Σ must jointly form a Perfect-Bayesian Equilibrium (PBE) to be part of an incentive feasible contract if $\mathcal{B} \neq \emptyset$. Constraints (4)-(6) are the associated equilibrium conditions. First, (4) is a rationality constraint on Σ that keeps an agent of type i at least indifferent between truth-telling and lying given that the agent correctly expects to receive x_j if it invokes the escape clause by reporting $j \in \mathcal{B}$. Second, (5) is a sequential rationality constraint on $\mathbf{x}_{\mathcal{B}}$ requiring that x_j maximize the expected surplus of the principal subsequent to every cost report θ_j , $j \in \mathcal{B}$, and given the principal's Stage 4 distribution of beliefs about the agent's true marginal cost θ_i . Third, (6) is a consistency requirement that the principal's posterior beliefs satisfy Bayes' rule. The final constraint (7) appears because of constrained ex-ante contracting, and does not feature in Bester and Strausz (2001). We use $\Gamma(\mathcal{A}, \mathcal{B})$ to label the set of incentive feasible mechanisms given $(\mathcal{A}, \mathcal{B})$.

Following again Bester and Strausz (2001), a mechanism $(\hat{\mathbf{x}}, \hat{\Sigma}|\mathcal{A}, \mathcal{B})$ is *incentive efficient* given $(\mathcal{A}, \mathcal{B})$ if it maximizes the principal's expected surplus

$$W(\mathbf{x}, \Sigma|\mathcal{A}, \mathcal{B}) = \sum_{i=1}^I \sum_{j=1}^I \nu_i \sigma_{ji} W_i(x_j) \quad (8)$$

in the set $\Gamma(\mathcal{A}, \mathcal{B})$ of incentive feasible mechanisms. Observe that the principal optimizes both over the menu of contracts $\mathbf{x}$ and the reporting strategy Σ .

Bester and Strausz (2001) consider exogenously incomplete commitment, which amounts to exogenously given $(\mathcal{A}, \mathcal{B})$ in the present context. Our paper attempts to bridge the gap between the two paradigms by endogenizing commitment. Specifically, at Stage 0 of the game, the principal chooses $(\mathcal{A}, \mathcal{B})$ to maximize the expected surplus $W(\hat{\mathbf{x}}, \hat{\Sigma}|\mathcal{A}, \mathcal{B})$. An *incentive optimal* mechanism $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ solves this problem.

Instances can occur when an incomplete commitment mechanism can do as well as one with complete commitment, but no better. The principal reduces contract commitment at stage 0 if and only if doing so strictly increases expected surplus.

Definition 1 (Escape clauses are minimal) *A mechanism $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ containing an escape clause ($\mathcal{B}^* \neq \emptyset$) is incentive optimal if and only if it maximizes the principal's expected surplus in the set of incentive efficient mechanisms,*

$$W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) \geq W(\hat{\mathbf{x}}, \hat{\Sigma}|\mathcal{A}, \mathcal{B}) \quad \forall (\mathcal{A}, \mathcal{B}) \subset [\mathcal{I} \cup \emptyset] \times [\mathcal{I} \cup \emptyset], \quad \mathcal{A} \cap \mathcal{B} = \emptyset, \quad (9)$$

and the escape clause is minimal in the following sense:

$$W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) > W(\hat{\mathbf{x}}, \hat{\Sigma}|\mathcal{A}, \mathcal{B}) \quad \forall (\mathcal{A}, \mathcal{B}) \subset [\mathcal{I} \cup \emptyset] \times [\mathcal{B}^* \cup \emptyset], \quad \mathcal{A} \cap \mathcal{B} = \emptyset, \quad \mathcal{B} \neq \mathcal{B}^*. \quad (10)$$

The applicability of the stochastic revelation principle The menu of contracts $\mathbf{x}$ is direct by assumption. Bester and Strausz (2001) show that the principal cannot gain anything by extending communication to more general message spaces under a certain set of assumptions. The result applies if the principal contracts with one single agent with private information about his type in a discrete and finite type space, the menu of contracts $\mathbf{x}$ and the agent's reporting strategy Σ maximize the expected surplus of the principal, and the agent communicates its type with the principal only once. Bester and Strausz (2001) do not specifically consider incomplete commitment in the form of an escape clause, nor do they discuss constrained contracting. Still, their *stochastic revelation principle* applies to our model, as we will now demonstrate.

Bester and Strausz (2001) define two metric spaces, X and Y , and a correspondence $F : X \rightrightarrows Y$ between the two. For any $x \in X$, the principal is obliged to implement some $y \in F(x)$. A mechanism (M, x) consists of a metric space M and a decision function $x : M \rightarrow X$ that assigns a decision $x(m) \in X$ to any message $m \in M$ the agent can convey to the principal.

Their formulation has enough flexibility to incorporate incomplete commitment in the shape of an escape clause. An allocation $x = (q, t)$ in our setting consists of output $q \geq 0$ to be supplied by the agent and a transfer $t \geq 0$ from the principal to the agent. Let $X = \mathbb{R}_{++}^2 \cup (0, 0) \cup \emptyset$ and $Y = \mathbb{R}_+^2$. Let the principal be contractually obliged to implement $x(m) = (q(m), t(m))$ if the agent has sent a message m such that $x(m) \in \mathbb{R}_{++}^2 \cup (0, 0)$. Formally, $F(x) = x$ for all $x \in \mathbb{R}_{++}^2 \cup (0, 0)$. Let $M_{++} = \{m \in M | x(m) \in \mathbb{R}_{++}^2\}$ be the subset of messages that commit the principal to implementing strictly positive output and transfer. The agent invokes the escape clause by sending a message m such that $x(m) = \emptyset$. Subsequent to such a message, the principal can offer any contract $y(m) = (q(m), t(m)) \in \mathbb{R}_+^2$. Formally, $F(\emptyset) = Y$. Let $M_\emptyset = \{m \in M | x(m) = \emptyset\}$ be the subset of messages that activate the escape clause. The mechanism features complete commitment if $M_\emptyset = \emptyset$. It then entails pure ex ante contracting in our terminology. Conversely, the mechanism contains an escape clause if $M_\emptyset \neq \emptyset$.

A particular feature of our model is constrained contracting. Define the image $x[M_{++}] = \{a | x(m) = a \text{ some } m \in M_{++}\}$ of the decision function x over M_{++} . Constrained contracting in our framework implies that this image defines a finite set. $|x[M_{++}]| \leq K$ implies that the principal may offer at most K complete commitment contracts with strictly positive output and transfers to the agent.

In Bester and Strausz (2001), a PBE induced by the mechanism (M, x) is characterized by three functions (r, p, y) . The function r is a reporting strategy (a measurable function defined over the Borel σ -algebra of M) that for each potential type maximizes the agent's expected payoff function given y . p is a function defining consistent posterior beliefs by

the principal about the agent's type based on Bayes' rule and on q . y is a function of sequentially rational decisions by the principal based on p . A mechanism (M, x) may induce multiple PBE. Bester and Strausz (2001) assume in this case that the principal implements an equilibrium that maximizes the principal's expected payoff function of the mechanism. This equilibrium property is necessary for a mechanism to be *incentive efficient* in their model.

Bester and Strausz (2001) prove the stochastic revelation principle in three steps. First, their Lemma 1 establishes a necessary condition for a reporting strategy r that is part of a PBE with consistent posterior beliefs p and sequentially rational decisions y to maximize the principal's expected payoff for a given mechanism (M, x) . Second, their Proposition 1 applies their Lemma 1 to establish an alternative reporting strategy r' which is part of a PBE characterized by (r', p, y) under the same mechanism (M, x) . (r', p, y) is payoff equivalent to (r, p, y) given (M, x) . Compared to the more general r , r' has a simplified structure: Each potential agent type randomizes over a finite set of messages $M' = \{m_1, \dots, m_h, \dots, m_{|M'|}\}$, $|M'| \leq I$. The I probability vectors $r'_i = \{r'_i(m_1), \dots, r'_i(m_h), \dots, r'_i(m_{|M'|})\}$ are linearly independent. Because (M, x) is held constant, the results go through even under the constraint $|x[M_{++}]| \leq K$. Hence, Lemma 1 and Proposition 1 in Bester and Strausz (2001) hold under constrained contracting.

Bester and Strausz (2001) apply Proposition 1 to create an alternative mechanism (M', x') with a finite number of decisions $x' = \{x(m_1), \dots, x(m_h), \dots, x(m_{|M'|})\}$. The subset $x' \cap \mathbb{R}_{++}^2$ of decisions with strictly positive output and transfers under the modified mechanism (M', x') is a subset of the decisions with strictly positive output and transfers made under the original mechanism (M, x) . Hence, (M', x') satisfies the requirement of constrained contracting if the original mechanism does so by $|x' \cap \mathbb{R}_{++}^2| \leq |x[M_{++}]| \leq K$. By construction, the PBE (r', p', y') under mechanism (M', x') is payoff equivalent to the PBE (r, p, y) under mechanism (M, x) .

Third, their Proposition 2 applies linear independence of q' to construct a function $C : \mathcal{I} \rightarrow M'$. On the basis of this function, they create a direct mechanism $(\mathcal{I}, \hat{x})$ with the property $\hat{x} = \{x(C(1)), \dots, x(C(i)), \dots, x(C(I))\}$. The subset $\hat{x} \cap \mathbb{R}_{++}^2$ of decisions with strictly positive output and transfers under the direct mechanism $(\mathcal{I}, \hat{x})$ is a subset of the decisions with strictly positive output and transfers under the modified mechanism (M', x') . Hence, $(\mathcal{I}, \hat{x})$ satisfies the requirement of constrained contracting if (M', x') does so by $|\hat{x} \cap \mathbb{R}_{++}^2| \leq |x' \cap \mathbb{R}_{++}^2| \leq K$. By construction, the PBE $(\hat{r}, \hat{p}, \hat{y})$ under mechanism $(\mathcal{I}, \hat{x})$ is payoff equivalent to the PBE (r', p', y') under mechanism (M', x') .

These results confirm that the stochastic revelation principle derived by Bester and Strausz (2001) apply to our model where incomplete commitment occurs in the shape of an escape clause and where ex-ante contracting opportunities are constrained.

Basic properties of incentive feasible and incentive efficient mechanisms The analysis in Tangerås and Gick (2026) rests on some basic properties of incentive feasible and incentive efficient mechanisms under incomplete commitment. The first claim reiterates the standard results that local incentive compatibility, individual rationality of the least efficient cost type and monotonicity of output are necessary and sufficient for incentive compatibility and individual rationality.

Claim 1 *A mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ satisfies individual rationality (2) and incentive compatibility (3) if and only if the following conditions are all met:*

$$U_I(x_I) \geq 0, \quad (11)$$

$$U_i(x_i) \geq U_i(x_{i+1}) \quad \forall i \in \{1, \dots, I-1\}, \quad (12)$$

$$U_i(x_i) \geq U_i(x_{i-1}) \quad \forall i \in \{2, \dots, I\}, \quad (13)$$

$$q_i \geq q_{i+1} \quad \forall i \in \{1, \dots, I-1\}, \quad (14)$$

Proof. Necessity of (11)-(13) is obvious. Local incentive compatibility implies

$$U_i(x_i) \geq U_{i+1}(x_{i+1}) + (\theta_{i+1} - \theta_i)q_{i+1}, \quad U_{i+1}(x_{i+1}) \geq U_i(x_i) - (\theta_{i+1} - \theta_i)q_i \quad \forall i \in \{1, \dots, I-1\}.$$

By rearranging expressions we get

$$(\theta_{i+1} - \theta_i)q_i \geq U_i(x_i) - U_{i+1}(x_{i+1}) \geq (\theta_{i+1} - \theta_i)q_{i+1} \quad \forall i \in \{1, \dots, I-1\}.$$

Hence, output is non-increasing in marginal cost in any incentive compatible mechanism, even if this mechanism features incomplete commitment.

As for sufficiency, the net benefit of truthfully reporting cost θ_i relative to exaggerating it to θ_j , $j \in \{i+1, \dots, I\}$ can be written as

$$U_i(x_i) - U_i(x_j) = \sum_{h=i}^{j-1} [U_h(x_h) - U_h(x_{h+1}) + (\theta_{h+1} - \theta_h)(q_{h+1} - q_j)] \geq 0 \quad \forall i \in \{1, \dots, I-1\}, \quad (15)$$

where non-negativity follows from the assumptions of local downward incentive compatibility (12) and monotonicity (14). The net benefit of truthfully reporting cost θ_i relative to understating it to θ_j , $j \in \{1, \dots, i-1\}$, equals

$$U_i(x_i) - U_i(x_j) = \sum_{h=j+1}^i [U_h(x_h) - U_h(x_{h-1}) + (\theta_h - \theta_{h-1})(q_j - q_{h-1})] \geq 0 \quad \forall i \in \{2, \dots, I\}, \quad (16)$$

where non-negativity follows from the assumptions of local upward incentive compatibility (13) and monotonicity (14).

$$U_i(x_i) \geq U_i(x_I) = U_I(x_I) + (\theta_I - \theta_i)q_I \geq U_I(x_I) \geq 0 \forall i \in \mathcal{I}.$$

then implies individual rationality (2). ■

The following result establishes conditions under which the agent *strictly* prefers to tell the truth rather than deviate to some *non-local downward* contract.

Claim 2 *Let $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ be an incentive feasible mechanism.*

1. *If $U_i(x_i) > U_i(x_{i+1})$, then $U_h(x_h) > U_h(x_j) \forall (h, j) \in \{1, \dots, i\} \times \{i+1, \dots, I\}$.*
2. *If $U_i(x_i) = U_i(x_{i+1})$ and $q_i > q_{i+1}$, then $U_h(x_h) > U_h(x_j) \forall (h, j) \in \{1, \dots, i-1\} \times \{i+1, \dots, I\}$.*

Proof. By (15), the net benefit of truthfully reporting cost θ_i , relative to exaggerating it to θ_j , $j \in \{i+1, \dots, I\}$, satisfies

$$\begin{aligned} U_i(x_i) - U_i(x_j) &= U_i(x_i) - U_i(x_{i+1}) + (\theta_{i+1} - \theta_i)(q_{i+1} - q_j) \\ &\quad + \sum_{h=i+1}^{j-1} [U_h(x_h) - U_h(x_{h+1}) + (\theta_{h+1} - \theta_h)(q_{h+1} - q_j)] > 0 \end{aligned}$$

if $U_i(x_i) > U_i(x_{i+1})$. Similarly, the net benefit of truthfully reporting cost θ_h , $h \in \{1, \dots, i-1\}$, relative to exaggerating it to θ_j , $j \in \{i+1, \dots, I\}$, satisfies

$$\begin{aligned} U_h(x_h) - U_h(x_j) &= \sum_{l=h}^{i-1} [U_l(x_l) - U_l(x_{l+1})] + \sum_{l=h}^{i-2} (\theta_{l+1} - \theta_l)(q_{l+1} - q_j) \\ &\quad + U_i(x_i) - U_i(x_{i+1}) + (\theta_i - \theta_{i-1})(q_i - q_{i+1}) + (\theta_{i+1} - \theta_{i-1})(q_{i+1} - q_j) \\ &\quad + \sum_{l=i+1}^{j-1} [U_l(x_l) - U_l(x_{l+1}) + (\theta_{l+1} - \theta_l)(q_{l+1} - q_j)] \\ &\geq U_i(x_i) - U_i(x_{i+1}) + (\theta_i - \theta_{i-1})(q_i - q_{i+1}) > 0 \end{aligned}$$

if either $U_i(x_i) > U_i(x_{i+1})$, or $U_i(x_i) = U_i(x_{i+1})$ and $q_i > q_{i+1}$. ■

The next result establishes conditions under which the agent *strictly* prefers to tell the truth rather than deviate to some *non-local upward* contract.

Claim 3 *Let $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ be an incentive feasible mechanism.*

1. *If $U_{i+1}(x_{i+1}) > U_{i+1}(x_i)$, then $U_h(x_h) > U_h(x_j) \forall (h, j) \in \{i+1, \dots, I\} \times \{1, \dots, i\}$.*

2. If $U_{i+1}(x_{i+1}) = U_{i+1}(x_i)$ and $q_i > q_{i+1}$, then $U_h(x_h) > U_h(x_j) \forall (h, j) \in \{i+2, \dots, I\} \times \{1, \dots, i\}$.

Proof. By (16), the net benefit of truthfully reporting cost θ_{i+1} , relative to understating it to θ_j , $j \in \{1, \dots, i\}$, satisfies

$$\begin{aligned} U_{i+1}(x_{i+1}) - U_{i+1}(x_j) &= U_{i+1}(x_{i+1}) - U_{i+1}(x_i) + (\theta_{i+1} - \theta_i)(q_j - q_i) \\ &\quad + \sum_{h=j+1}^i [U_h(x_h) - U_h(x_{h-1}) + (\theta_h - \theta_{h-1})(q_j - q_{h-1})] > 0, \end{aligned}$$

by the assumption that $U_{i+1}(x_{i+1}) > U_{i+1}(x_i)$. The net benefit of truthfully reporting cost θ_h , $h \in \{i+2, \dots, I\}$, relative to understating it to θ_j , $j \in \{1, \dots, i\}$, satisfies

$$\begin{aligned} U_h(x_h) - U_i(x_j) &= \sum_{l=i+2}^h [U_l(x_l) - U_l(x_{l-1})] + \sum_{l=i+3}^h (\theta_l - \theta_{l-1})(q_j - q_{l-1}) \\ &\quad + U_{i+1}(x_{i+1}) - U_{i+1}(x_i) + (\theta_{i+2} - \theta_{i+1})(q_i - q_{i+1}) + (\theta_{i+2} - \theta_i)(q_j - q_i) \\ &\quad + \sum_{l=j+1}^i [U_l(x_l) - U_l(x_{l-1}) + (\theta_l - \theta_{l-1})(q_j - q_{l-1})] \\ &\geq U_{i+1}(x_{i+1}) - U_{i+1}(x_i) + (\theta_{i+2} - \theta_{i+1})(q_i - q_{i+1}) > 0 \end{aligned}$$

if either $U_{i+1}(x_{i+1}) > U_{i+1}(x_i)$, or $U_{i+1}(x_{i+1}) = U_{i+1}(x_i)$ and $q_i > q_{i+1}$. ■

The fourth result is that two contracts have the same transfer if they have the same output:

Claim 4 Let $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ be an incentive feasible mechanism. If $q_i = q_j$, then $x_i = x_j$.

Proof. Incentive compatibility (3) implies

$$t_i - \theta_i q_i \geq t_j - \theta_i q_j, \quad t_j - \theta_j q_j \geq t_i - \theta_j q_i \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I}.$$

Rearranging the two expressions yields

$$\theta_i(q_j - q_i) \geq t_j - t_i \geq \theta_j(q_j - q_i) \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I}.$$

If $q_i = q_j$, then $t_i = t_j$ and therefore $x_i = x_j$. ■

Our next intermediary result concerns incentive efficient mechanisms.

Claim 5 A mechanism $(\hat{\mathbf{x}}, \hat{\Sigma} | \mathcal{A}, \mathcal{B})$, $\mathcal{A} = \{1, \dots, A\} \neq \emptyset$, is incentive efficient only if

$$U_i(\hat{x}_i) = U_i(\hat{x}_{i+1}) \quad \forall i \in \{1, \dots, A-1\}, \quad A \geq 2. \quad (17)$$

If $\mathcal{B} = \emptyset$, then incentive efficiency requires $U_A(\hat{x}_A) = 0$.

Proof. We first show that the mechanism is incentive efficient only if

$$[U_i(\hat{x}_i) - U_i(\hat{x}_{i+1})][U_{i+1}(\hat{x}_{i+1}) - U_{i+1}(\hat{x}_i)] = 0 \quad \forall i \in \{1, \dots, \min\{A, I-1\}\}. \quad (18)$$

Suppose both local IC constraints are slack for some $i \in \{1, \dots, \min\{A, I-1\}\}$. By Claim 2, the downward IC constraints are slack, and therefore $\hat{\sigma}_{ji} = 0$, for any cost θ_i and cost report θ_j such that $(i, j) \in \{1, \dots, \min\{A, I-1\}\} \times \{\min\{A+1, I\}, \dots, I\}$. By Claim 3, the upward IC constraints are slack, and therefore $\hat{\sigma}_{ji} = 0$, for any cost θ_i and cost report θ_j such that $(i, j) \in \{\min\{A+1, I\}, \dots, I\} \times \{1, \dots, \min\{A, I-1\}\}$. A marginal reduction in the transfer payment $\hat{t}_j$ by a small amount ϵ for all types $j \in \{1, \dots, \min\{A, I-1\}\}$ then increases the principal's expected surplus while maintaining incentive feasibility. Consequently, the proposed mechanism cannot be incentive efficient.

Suppose the local upward IC constraint in (18) is binding, but the local downward IC constraint is slack for some $i \in \{1, \dots, A-1\}$, $A \geq 2$. By implication, $\hat{q}_i > \hat{q}_{i+1}$. For if $\hat{q}_i = \hat{q}_{i+1}$, then $\hat{x}_i = \hat{x}_{i+1}$ by Claim 4 and therefore $U_i(\hat{x}_i) = U_i(\hat{x}_{i+1})$. An agent with marginal cost equal to θ_i or below strictly prefers to truthfully report its cost rather than exaggerate it to θ_{i+1} or above by Claim 2. By $\hat{q}_i > \hat{q}_{i+1}$ and Claim 3, an agent with marginal cost equal to θ_{i+2} or above strictly prefers to truthfully report its cost rather than understate it to θ_i or below. Construct a perturbed mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ by setting $t_j = \hat{t}_j - \epsilon$, $\epsilon > 0$, for all $j \in \{1, \dots, i\}$ and setting $\sigma_{i+1} = \sum_{j=1}^i \hat{\sigma}_{j(i+1)} + \hat{\sigma}_{i+1}$. Everything else is held equal to the original mechanism. This perturbed mechanism is incentive feasible for all ϵ sufficiently small. The difference in expected surplus between the two mechanisms is:

$$W(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B}) - W(\hat{\mathbf{x}}, \hat{\Sigma} | \mathcal{A}, \mathcal{B}) = \sum_{j=1}^i \nu_j [(1-\alpha)\epsilon + \hat{\sigma}_{j(i+1)}(W_{i+1}(\hat{x}_{i+1}) - W_{i+1}(\hat{x}_j))] > 0.$$

If $W_{i+1}(\hat{x}_j) > W_{i+1}(\hat{x}_{i+1})$ for some $j \in \{1, \dots, i\}$, then $\hat{\sigma}_{i+1} = 0$, which would violate incentive feasibility. Hence, $W_{i+1}(\hat{x}_{i+1}) \geq W_{i+1}(\hat{x}_j)$ for all $j \in \{1, \dots, i\}$. We conclude that an incentive efficient mechanism satisfies (18) only if $U_i(\hat{x}_i) = U_i(\hat{x}_{i+1})$ for all $i \in \{1, \dots, A-1\}$, $A \geq 2$.

To prove the second part of the claim, let $\mathcal{B} = \emptyset$. If $A \leq I-1$, then $A+1 = \underline{C}$. Suppose $U_{A+1}(\hat{x}_{A+1}) = U_{A+1}(\hat{x}_A)$ and $U_A(\hat{x}_A) > U_A(\hat{x}_{A+1})$. Then, the principal can increase expected surplus by reducing the transfer by the same amount for all cost types equal to or below A , and through an additional change of the reporting strategy of type $A+1$; see the previous analysis. Hence, $U_A(\hat{x}_A) = U_A(\hat{x}_{A+1}) = U_A(x_0) = 0$ because the

IC constraint of type A is locally downward-binding by (18). If $A = I$ and $U_I(\hat{x}_I) > 0$, then the principal could reduce the transfer for all cost types without violating incentive feasibility. Hence, $U_A(\hat{x}_A) = 0$ also in this final case. ■

3 Maximally constrained contracting

Consider the polar extreme case of maximally constrained contracting where the principal can offer at most one contract upfront to the agent, that is, $K = 1$. Assume first that the principal implements a pure ex-ante contract that only an agent with marginal cost equal to or below θ_A , $A \in \mathcal{I}$, accepts. The rent-minimizing transfer by the principal equals $t = \theta_A q$ for arbitrary quantity q . Maximization of the principal's expected surplus

$$\tilde{W}_A(q) = \sum_{i=1}^A \nu_i [S(q) - (\theta_i + (1 - \alpha)(\theta_A - \theta_i))q]$$

over q yields the associated output requirement $\hat{q}_A$ as the solution to

$$S'(\hat{q}_A) = \sum_{i=1}^A \frac{\nu_i}{G_A} (\theta_i + (1 - \alpha)(\theta_A - \theta_i)) \leq \theta_A = S'(q_A^{fb}).$$

The inequality is strict if $A \geq 2$, which implies $\hat{q}_A > q_A^{fb}$ in that case. We denote the incentive efficient ex-ante contract that yields a cut-off at θ_A by $\hat{x}_A = (\hat{q}_A, \theta_A \hat{q}_A)$. Let $\theta_{\hat{A}}$ be the cut-off that maximizes the principal's expected surplus among all potential cut-offs θ_A , $A \in \mathcal{I} \cup \emptyset$. Assume that $\hat{A} \geq A$, and let the maximum be strict:

$$\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}}) > \tilde{W}_A(\hat{q}_A) \quad \forall A \in \mathcal{I} \cup \emptyset, \quad A \neq \hat{A}.$$

Consider pure ex-post contracting. The principal does not offer any contract upfront, but states an upper bound θ_B to the marginal cost report above which there will be no contract with the agent. If the agent reports marginal cost $\theta_j \leq \theta_B$, then the principal offers a contract $x_j = (q_j, t_j)$ based on this cost report and the beliefs inferred about the agent's true cost.

Let $B = \hat{A}$. Assume that the agent randomizes uniformly across all cost reports θ_j , $j \in \{1, \dots, \hat{A} - 1\}$ if it has marginal cost $\theta_i < \theta_{\hat{A}}$. Let $\sigma_{\hat{A}} \in (0, 1)$ be the probability that an agent with marginal cost $\theta_{\hat{A}}$ truthfully reports its cost, and assume that this agent reports marginal cost θ_j with probability $\frac{1 - \sigma_{\hat{A}}}{\hat{A} - 1}$ for all $j < \hat{A}$. Assume that the agent rejects the contract if it has marginal cost $\theta_i > \theta_{\hat{A}}$ and $\hat{A} + 1 \leq I$.

The sequentially rational contract offer by the principal subsequent to the cost report $\theta_{\hat{A}}$ is $x_{\hat{A}}^{fb}$ because $w_{\hat{A}}^{fb} > 0$ and the principal attaches posterior probability $\mu_{\hat{A}} = 1$ to

the event that the agent was truthful upon receiving the report $\theta_{\hat{A}}$. Subsequent to every marginal cost report $\theta_j < \theta_{\hat{A}}$, the principal offers the transfer $t = \theta_{\hat{A}}q$ under the assumption that the principal wants the agent to produce q for all marginal cost realizations equal to or below $\theta_{\hat{A}}$. The expected ex-post surplus of the principal equals

$$\tilde{\Omega}_{\hat{A}}(q, \sigma_{\hat{A}}) = \sum_{i=1}^{\hat{A}-1} \nu_i [S(q) - (\theta_i + (1 - \alpha)(\theta_{\hat{A}} - \theta_i))q] + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})[S(q) - \theta_{\hat{A}}q] \quad (19)$$

divided by $G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})$. The output target $q_1 > q_{\hat{A}}^{fb}$ solves

$$S'(q_1) = \frac{\sum_{i=1}^{\hat{A}} \nu_i (\theta_i + (1 - \alpha)(\theta_{\hat{A}} - \theta_i)) - \nu_{\hat{A}} \sigma_{\hat{A}} \theta_{\hat{A}}}{G_{\hat{A}} - \nu_{\hat{A}} \sigma_{\hat{A}}} < \theta_{\hat{A}} = S'(q_{\hat{A}}^{fb}). \quad (20)$$

The ex-post contract for any cost report θ_j , $j \in \{1, \dots, \hat{A} - 1\}$ is $x_1 = (q_1, \theta_{\hat{A}}q_1)$ if the principal wants the agent to produce for all cost realizations equal to or below $\theta_{\hat{A}}$.

The ex ante expected surplus of the principal equals

$$\Omega_{\hat{A}}(q_1, \sigma_{\hat{A}}) = \tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}}) + \nu_{\hat{A}} \sigma_{\hat{A}} w_{\hat{A}}^{fb} \quad (21)$$

under pure ex-post contracting. Tangerås and Gick (2026) establish the following result:

Proposition 1 *The principal strictly prefers pure ex-post over pure ex-ante contracting if contracting is maximally constrained ($K = 1$), and the principal's surplus-maximizing pure ex-ante mechanism involves some pooling of cost types ($\hat{A} \geq 2$).*

The principal can do even better by adding an escape clause:

Proposition 2 *The incentive optimal mechanism contains an escape clause if contracting is maximally constrained ($K = 1$), and the principal's surplus-maximizing pure ex-ante mechanism involves some pooling of cost types ($\hat{A} \geq 2$).*

Proof. The first and most lengthy part of the proof is to derive an upper bound on the principal's expected surplus in mechanisms with pure ex-post contracting. The second part shows that adding an escape clause strictly increases expected surplus.

Let $z_j \in \{j, \dots, I\}$ be the maximal type that reports θ_j with positive probability in an incentive feasible mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$, that is, $\sigma_{jz_j} > 0$ and $\sigma_{ji} = 0$ for all $i \in \{z_j + 1, \dots, I\}$ if $z_j \leq I - 1$. The type z_j exists by $\sigma_j > 0$.

Claim 6 *A mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ with incomplete commitment ($\mathcal{B} \neq \emptyset$) is incentive feasible only if $t_j = \theta_{z_j}q_j > 0 \ \forall j \in \mathcal{B}$.*

Proof. Consider the principal's optimal choice t_j in Stage 4 after some cost report θ_j , $j \in \mathcal{B}$. If $t_j > \theta_{z_j} q_j$, then the principal can strictly reduce the transfer and thereby save on information rent without violating the individual rationality constraint for any type $i \in \mathcal{I}$ that also reports to be of type θ_j with positive probability. If $t_j < \theta_{z_j} q_j$, then $U_{z_j}(x_j) < 0 \leq U_{z_j}(x_{z_j})$ and therefore $\sigma_{jz_j} = 0$ by (4), which contradicts the assumption that $\sigma_{jz_j} > 0$. This leaves $t_j = \theta_{z_j} q_j$ as the only remaining possibility. Substituting t_j into (5) and maximizing over q_j leads to

$$S'(q_j) = \sum_{i=1}^I \mu_{ji}(\theta_i + (1 - \alpha)(\theta_{z_j} - \theta_i)) \leq \theta_I = S'(q_I^{fb}).$$

Strict concavity of $S(q)$ implies $q_j \geq q_I^{fb} > 0$, where $q_I^{fb} > 0$ by assumption. ■

Let $\underline{B}$ be the minimal and B the maximal cost type in $\mathcal{B}$ if this set is nonempty. By monotonicity of output, $\mathcal{B} = \{1, \dots, B\}$ in any incentive feasible mechanism $(\mathbf{x}, \Sigma | \emptyset, \mathcal{B})$ with pure ex-post contracting. Hence, $\underline{B} = 1$. Moreover, $\mathcal{C} = \{B + 1, \dots, I\}$ if $B \leq I - 1$.

Claim 7 *A mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ with incomplete commitment ($\mathcal{B} \neq \emptyset$) is incentive feasible only if $z_j = z \geq B \forall j \in \mathcal{B}$. Incentive feasibility further implies:*

1. $x_j = x_{\underline{B}} \forall j \in \{\underline{B}, \dots, z - 1\}$ if either $\underline{B} \leq B - 1$ or $z \geq B + 1$.
2. $x_j = x_0 \forall j \in \{z + 1, \dots, I\}$ if $z \leq I - 1$.

Proof. We first show that $z_j = z \geq B$ for all $j \in \mathcal{B}$. Observe that $U_{z_j}(x_{z_j}) = U_{z_j}(x_j) = t_j - \theta_{z_j} q_j = 0$ by $\sigma_{jz_j} > 0$ and Claim 6. Suppose $z_j < z_h$ for some $(j, h) \in \mathcal{B}^2$. In this case, $U_{z_j}(x_{z_j}) = 0 < (\theta_{z_h} - \theta_{z_j}) q_{z_h} = U_{z_j}(x_{z_h})$, which violates incentive compatibility (3). By implication, $z_j = z_B \geq B$ for all $j \in \mathcal{B}$.

Consider Item 1. Note that $z \geq \underline{B} + 1$ by $z \geq B$ and the assumptions of the claim. By (3),

$$U_z(x_{z-1}) = U_{z-1}(x_{z-1}) - (\theta_z - \theta_{z-1}) q_{z-1} \leq U_z(x_z).$$

Moreover, $U_z(x_z) = U_z(x_{\underline{B}}) = t_{\underline{B}} - \theta_z q_{\underline{B}} = 0$ by $\sigma_{\underline{B}z} > 0$ and Claim 6. Invoking incentive compatibility and Claim 6 again delivers,

$$U_{z-1}(x_{z-1}) \geq U_{z-1}(x_{\underline{B}}) = t_{\underline{B}} - \theta_{z-1} q_{\underline{B}} = (\theta_z - \theta_{z-1}) q_{\underline{B}}.$$

Combining these two inequalities produces

$$(\theta_z - \theta_{z-1}) q_{\underline{B}} \leq U_{z-1}(x_{z-1}) \leq (\theta_z - \theta_{z-1}) q_{z-1},$$

and therefore $q_{\underline{B}} \leq q_{z-1}$. By monotonicity, it must also be the case that $q_{\underline{B}} \geq q_{z-1}$. Hence, $q_{\underline{B}} = q_{z-1}$. Applying monotonicity again yields $q_j = q_{\underline{B}} = q_{z-1}$ for all $j \in \{\underline{B}, \dots, z-1\}$. We can now invoke Claim 4 to obtain $x_j = x_{\underline{B}}$ for all $j \in \{\underline{B}, \dots, z-1\}$.

Consider Item 2 of the claim. Assume that $z \leq I-1$, and suppose either $q_j > 0$ or $q_j = 0$ and $t_j > 0$ for some $j \in \{z+1, \dots, I\}$. In this case, $U_z(x_j) = U_j(x_j) + (\theta_j - \theta_z)q_j > 0 = U_z(x_z)$, which violates incentive compatibility. By necessity, $x_j = (0, 0) = x_0$ for all $j \in \{z+1, \dots, I\}$. ■

Claim 8 *Any incentive feasible mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ with incomplete commitment ($\mathcal{B} \neq \emptyset$) has the property that $|\mathbf{x}_{\mathcal{B}}| \in \{1, 2\}$. Moreover, $x_j = x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}}) \forall j \in \{\underline{B}, \dots, B-1\}$ and $x_B = x_B^{fb}$ if $|\mathbf{x}_{\mathcal{B}}| = 2$.*

Proof. Trivially, $|\mathbf{x}_{\mathcal{B}}| = 1$ if $\underline{B} = B$. Let $\underline{B} \leq B-1$. By Claim 7, $\underline{B} \leq B-1 \leq z-1$, and then all contracts $x_j, j \in \{\underline{B}, \dots, B-1\}$ are identical and equal to $x_{\underline{B}}$. Seeing as $\mathcal{B} \subset \{\underline{B}, \dots, B\}$, $|\mathbf{x}_{\mathcal{B}}| \in \{1, 2\}$ if $\underline{B} \leq B-1$. Consider the second part of the claim. By Claim 7, $x_j = x_{\underline{B}}$ for all $j \in \{\underline{B}, \dots, B\}$ if $z \geq B+1$. In this case, $|\mathbf{x}_{\mathcal{B}}| = 1$ by $\mathcal{B} \subset \{\underline{B}, \dots, B\}$. Hence, $|\mathbf{x}_{\mathcal{B}}| = 2$ implies $z = B$. Invoking Claim 6, produces $x_j = x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}}) \forall j \in \{\underline{B}, \dots, B-1\}$ if $|\mathbf{x}_{\mathcal{B}}| = 2$. To establish $x_B = x_B^{fb}$ if $|\mathbf{x}_{\mathcal{B}}| = 2$, observe that

$$U_i(x_i) \geq U_i(x_{\underline{B}}) = (\theta_B - \theta_i)q_{\underline{B}} > (\theta_B - \theta_i)q_B = U_i(x_B) \forall i \in \{1, \dots, B-1\}.$$

The first (weak) inequality follows from incentive compatibility, the second (strict) inequality from $q_{\underline{B}} > q_B$ by $x_{\underline{B}} \neq x_B$ and monotonicity of output. Hence, $\sigma_{Bi} = 0$ for all $i \in \{1, \dots, B-1\}$. Furthermore, $\sigma_{Bi} = 0$ for all $i \in \{B+1, \dots, I\}$ by $z = B$. Hence, $\mu_B = 1$ by (6) if $|\mathbf{x}_{\mathcal{B}}| = 2$. Upon observing cost report θ_B , the principal attaches posterior probability equal to one that the agent in fact has marginal cost θ_B . The sequentially rational choice for the principal is then to offer x_B^{fb} and obtain ex-post surplus $w_B^{fb} > 0$. ■

Our next result shows that it is sufficient to evaluate pure ex-post mechanisms with $|\mathbf{x}_{\mathcal{B}}| = 2$ if $\hat{A} \geq 2$.

Claim 9 *Assume that $\hat{A} \geq 2$. For any incentive feasible pure ex-post mechanism $(\mathbf{x}, \Sigma | \emptyset, \mathcal{B})$, $|\mathbf{x}_{\mathcal{B}}| = 1$, there exists an alternative incentive feasible pure ex-post mechanism $(\mathbf{x}', \Sigma' | \emptyset, \mathcal{B}')$, $|\mathbf{x}'_{\mathcal{B}'}| = 2$, such that $W(\mathbf{x}', \Sigma' | \emptyset, \mathcal{B}') > W(\mathbf{x}, \Sigma | \emptyset, \mathcal{B})$.*

Proof. For arbitrary incentive feasible mechanism $(\mathbf{x}, \Sigma | \emptyset, \mathcal{B})$, $|\mathbf{x}_{\mathcal{B}}| = 1$, the pure ex-ante mechanism $(\mathbf{x}, \Sigma | \mathcal{B}, \emptyset)$ is incentive feasible and delivers expected welfare $W(\mathbf{x}, \Sigma | \emptyset, \mathcal{B}) = W(\mathbf{x}, \Sigma | \mathcal{B}, \emptyset) \leq \tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}})$, where $\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}})$ maximizes the expected welfare under pure ex-ante contracting, subject to $K = 1$. Proposition 1 shows that there exists an incentive

feasible mechanism $(\mathbf{x}', \Sigma' | \emptyset, \mathcal{B}')$, $|\mathbf{x}_{\mathcal{B}'}| = 2$, such that $W(\mathbf{x}', \Sigma' | \emptyset, \mathcal{B}') > \tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}})$, if $\hat{A} \geq 2$.

■

Let the mechanism $(\mathbf{x}, \Sigma | \emptyset, \mathcal{B})$ be incentive feasible, and assume that $|\mathbf{x}_{\mathcal{B}}| = 2$. We know from the proof of Claim 8 that $z = B$ in this case. Hence, $\sigma_{ji} = 0$ for all $(i, j) \in \{B + 1, \dots, I\} \times \mathcal{B}$ if $B \leq I - 1$. By implication, $q_{\underline{B}}$ solves

$$S'(q_{\underline{B}}) = \sum_{i=1}^B \frac{\nu_i \sigma_{ji}}{\sum_{h=1}^B \nu_h \sigma_{jh}} (\theta_i + (1 - \alpha)(\theta_B - \theta_i)) \quad \forall j \in \{1, \dots, B - 1\}.$$

Multiply both sides of the equation by the denominator on the right-hand side, and sum up across all $j \in \{1, \dots, B - 1\}$ to get

$$S'(q_{\underline{B}}) = \frac{\sum_{i=1}^B \nu_i (\theta_i + (1 - \alpha)(\theta_B - \theta_i)) - \nu_B \theta_B \sum_{j=B}^I \sigma_{jB}}{\sum_{i=1}^B \nu_i - \nu_B \sum_{j=B}^I \sigma_{jB}}$$

after simplification. To arrive at this expression, we used $\sum_{j=1}^{B-1} \sigma_{ji} = 1$ for all $i \in \{1, \dots, B - 1\}$ because $U_i(x_i) = (\theta_B - \theta_i)q_{\underline{B}} > (\theta_B - \theta_i)q_B = U_i(x_B) > 0 = U_i(x_j)$ for all $i \in \{1, \dots, B - 1\}$ and $j \in \{B + 1, \dots, I\}$ if $B \leq I - 1$. The expected surplus of this mechanism can be written as

$$W(\mathbf{x}, \Sigma | \emptyset, \mathcal{B}) = \Omega_B(q_{\underline{B}}, \sigma_B) - \nu_B W_B^{fb}(q_{\underline{B}}) \sum_{j \in \mathcal{C}} \sigma_{jB},$$

see (19) and (21) for definitions of the expressions on the right-hand side. Seeing as $w_B^{fb} > W_B^{fb}(q_{\underline{B}})$, the principal can increase expected surplus by increasing the probability of a truthful report by type B to $\tilde{\sigma}_B = \sum_{j=B}^I \sigma_{jB}$. This change has no implications for $x_{\underline{B}}$. The principal then maximizes $\Omega_B(q_{\underline{B}}, \tilde{\sigma}_B)$ over $\tilde{\sigma}_B$. The ratchet effect renders the mechanism incentive infeasible above an upper bound $\bar{\sigma}_B < 1$; see Tangerås and Gick (2026). Hence, $\Omega_B(q_{\underline{B}}, \bar{\sigma}_B)$ is the maximal expected surplus that the principal can achieve in an incentive feasible pure ex-post mechanism where $|\mathbf{x}_{\mathcal{B}}| = 2$.

Construct a mechanism $(\mathbf{x}, \Sigma' | \mathcal{A}', \mathcal{B}')$ featuring ex-ante contracting complemented by an escape clause, $\mathcal{A}' = \{1, \dots, B - 1\}$ and $\mathcal{B}' = \{B\}$. Assume that all cost types i in $\mathcal{A}'$ and $\mathcal{C}$ truthfully report their marginal cost, $\sigma'_i = 1$, and that type B truthfully reports marginal cost θ_B with probability $\sigma'_B = \tilde{\sigma}_B$ and understates the marginal cost to θ_{B+1} with probability $1 - \sigma'_B$. This alternative mechanism satisfies (2)-(6) by incentive feasibility of the pure ex-post mechanism. By $|\mathbf{x}_{\mathcal{A}'}| = 1 = K$, the modified mechanism also meets the contract constraint (7). It delivers expected surplus $\Omega_B(q_{\underline{B}}, \tilde{\sigma}_B)$, which is strictly increasing in $\tilde{\sigma}_B$. However, the principal is no longer bound by the sequential rationality constraint $\tilde{\sigma}_B \leq \bar{\sigma}_B$ because of ex-ante commitment to $x'_j = x_{\underline{B}}$ for all $j \in \{1, \dots, B - 1\}$.

By maximization of the agent's truthfulness, $\tilde{\sigma}_B = 1$. $W(\mathbf{x}, \mathbf{I}|\mathcal{A}', \mathcal{B}') = \Omega_B(q_B, 1) > \Omega_B(q_B, \tilde{\sigma}_B) \geq W(\mathbf{x}, \Sigma|\emptyset, \mathcal{B})$ implies that the principal can find a mechanism stipulating an escape clause which strictly outperforms any incentive feasible mechanism with pure ex-post contracting contingent on $|\mathbf{x}_B| = 2$. ■

4 Constrained contracting

We now turn to the general case $K \geq 1$ of constrained contracting. Section 4.1 analyzes constrained mechanisms under complete commitment, which represent the appropriate benchmark against which to evaluate incomplete commitment mechanisms. In Section 4.2, we derive general properties of mechanisms under incomplete commitment. Section 4.3 then establishes circumstances under which an escape clause increases the principal's expected surplus. Finally, Section 4.4 characterizes a fundamental trade-off underlying an incentive optimal escape clause.

4.1 Constrained contracting with complete commitment

This section contains a full analysis of incentive efficient mechanisms with complete commitment. The principal commits to an ex-ante menu of contracts ($\mathcal{A} \neq \emptyset$), but does not engage in ex-post contracting ($\mathcal{B} = \emptyset$). By the revelation principle, we can restrict attention to truthful mechanisms, i.e. incentive feasible mechanisms for which $\Sigma = \mathbf{I}$, where $\mathbf{I}$ is the I -dimensional identity matrix. The principal maximizes (8) over $\mathbf{x}$ subject to (2), (3) and (7).

We let $K \geq 1$, but do not necessarily assume that contracting is constrained. The incentive efficient mechanism under complete commitment consists of a partitioning $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_k, \dots, \mathcal{A}_{\tilde{K}}\}$ of $\mathcal{A} = \{1, \dots, A\}$ into $\tilde{K} \leq \min\{K; I\}$ non-empty *cost groups*. Let $\theta_{\underline{A}_k}$ be the lowest marginal cost and θ_{A_k} the highest marginal cost in cost group k . $x_{A_k} = (q_{A_k}, t_{A_k})$ is the contract awarded to the agent for cost reports θ_j , $j \in \mathcal{A}_k$. Each cost group is a convex set because incentive compatibility implies that output is non-increasing in the agent's marginal cost. By monotonicity of output, the set of cost types for which there is ex-ante contracting is convex and contains the most efficient cost types: $\mathcal{A} = \{1, \dots, A\}$, where θ_A is the marginal cost of the least efficient agent that produces positive output in the mechanism. The mechanism features full participation if $A = I$. Otherwise, $\mathcal{C} = \{A + 1, \dots, I\}$. We rank cost groups such that the agent is more efficient if it has marginal cost in cost group k than $k + 1$. Observe also that $x_{A_K} = x_A$ because A is the least efficient cost type contained in $\mathcal{A}_K$.

Claim 10 Assume that the mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\hat{A}, \emptyset)$ maximizes the principal's expected surplus under complete commitment. Let $1 \leq \hat{K} \leq \min\{K; I\}$ be the number of non-empty cost groups in this mechanism.

1. Output $\hat{q}_{\hat{A}_k}$ in non-empty cost group $\hat{A}_k$ is characterized by

$$S'(\hat{q}_{\hat{A}_k}) = \sum_{i \in \mathcal{A}_k} \frac{\nu_i}{\nu_{\mathcal{A}_k}} (\theta_i + \frac{G_{-i}}{\nu_i} (1 - \alpha)(\theta_i - \theta_{-i})), \quad (22)$$

where $\nu_{\mathcal{A}_k} = \sum_{i \in \mathcal{A}_k} \nu_i$ measures the ex-ante probability that the agent's marginal cost is contained in $\mathcal{A}_k$.

2. If cost group k features pooling, $|\hat{A}_k| \geq 2$, then output is upward- [downward-] distorted relative to the second-best efficient output of the least [most] efficient cost type in the cost group, $q_{\hat{A}_k}^{sb} < \hat{q}_{\hat{A}_k} < q_{\underline{\hat{A}}_k}^{sb}$

3. If $\hat{K} \geq 2$, then the upper boundary type $\hat{A}_k$ in interior cost group $k \in \{1, \dots, \hat{K} - 1\}$, satisfies the principal's local incentive compatibility constraint:

$$W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}}) \geq 0 \geq W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_{k+1}}). \quad (23)$$

4. The upper bound $\hat{A}$ to ex-ante contracting under partial participation, $\hat{A} \leq I - 1$, satisfies the principal's individual rationality constraint:

$$W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) \geq 0 \geq W_{\hat{A}+1}^{sb}(\hat{q}_{\hat{A}}). \quad (24)$$

Proof. We first establish Item 1 of Claim 10. In the process we also verify Item 2 of the claim. Using Claim 5, we can derive

$$U_i(x_i) = \sum_{j=i}^{I-1} (\theta_{j+1} - \theta_j) q_{j+1} \quad \forall i \in \{1, \dots, I-1\}, \quad U_I(x_I) = 0. \quad (25)$$

for agency rent in any incentive efficient mechanism with complete commitment. The expected welfare function

$$\sum_{i=1}^I \nu_i W_i^{sb}(q_i), \quad W_i^{sb}(q_i) = S(q_i) - (\theta_i + \frac{G_{i-1}}{\nu_i} (1 - \alpha)(\theta_i - \theta_{i-1})) q_i, \quad (26)$$

follows from substituting (25) and $\Sigma = \mathbf{I}$ into (8) and applying a summation by parts:

$$G_{A_k} \theta_{A_k} - G_{A_{k-1}} \theta_{A_{k-1}} = \sum_{i \in \mathcal{A}_k} [\nu_i \theta_i + G_{i-1} (\theta_i - \theta_{i-1})].$$

Maximization of the expected welfare function

$$W_{\mathcal{A}_k}(q) = \sum_{i \in \mathcal{A}_k} \nu_i W_i^{sb}(q) \quad (27)$$

of cost group k over q yields the incentive efficient output $\hat{q}_{A_k}$ in cost group k as the solution to (22).

The welfare function (26) is additively separable across cost groups, and the objective function (27) is strictly concave in output. Hence, a mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ with transfer payments that yield (25) and where output is characterized by (22) for all $\hat{K}$ cost groups, is incentive efficient if this mechanism is also incentive feasible. The mechanism satisfies (7) by $\hat{K} \leq K$. We then need to verify individual rationality (2) and incentive compatibility (3).

We demonstrate monotonicity of output. The second-best quantity q_i^{sb} of an agent with marginal cost θ_i solves

$$S'(q_i^{sb}) = \theta_i + \frac{G_{i-1}}{\nu_i} (1 - \alpha)(\theta_i - \theta_{i-1}) = \theta_i^{sb}. \quad (28)$$

Comparing $\hat{q}_{\hat{\mathcal{A}}_k}$ with $\hat{q}_{\hat{\mathcal{A}}_k}^{sb}$ delivers

$$\nu_{\hat{\mathcal{A}}_k} [S'(\hat{q}_{\hat{\mathcal{A}}_k}) - S'(q_{\hat{\mathcal{A}}_k}^{sb})] = \sum_{i \in \hat{\mathcal{A}}_k} \nu_i [\theta_i + \frac{G_{-i}}{\nu_i} (1 - \alpha)(\theta_i - \theta_{-i}) - \theta_{\hat{\mathcal{A}}_k} - \frac{G_{\hat{\mathcal{A}}_k-1}}{\nu_{\hat{\mathcal{A}}_k}} (1 - \alpha)(\theta_{\hat{\mathcal{A}}_k} - \theta_{\hat{\mathcal{A}}_k-1})] \geq 0$$

by assumption (1) of increasing virtual marginal production cost. Strict concavity of $S(q)$ implies $q_{\hat{\mathcal{A}}_k}^{sb} \geq \hat{q}_{\hat{\mathcal{A}}_k}$ with strict inequality if $|\hat{\mathcal{A}}_k| \geq 2$ (the second inequality of Item 2). Next,

$$\nu_{\hat{\mathcal{A}}_k} [S'(q_{\hat{\mathcal{A}}_k}^{sb}) - S'(\hat{q}_{\hat{\mathcal{A}}_k})] = \sum_{i \in \hat{\mathcal{A}}_k} \nu_i [\theta_{\hat{\mathcal{A}}_k} + \frac{G_{\hat{\mathcal{A}}_k-1}}{\nu_{\hat{\mathcal{A}}_k}} (1 - \alpha)(\theta_{\hat{\mathcal{A}}_k} - \theta_{\hat{\mathcal{A}}_k-1}) - \theta_i - \frac{G_{-i}}{\nu_i} (1 - \alpha)(\theta_i - \theta_{-i})] \geq 0$$

implies $\hat{q}_{\hat{\mathcal{A}}_k} \geq q_{\hat{\mathcal{A}}_k}^{sb}$ with strict inequality if $|\hat{\mathcal{A}}_k| \geq 2$ (the first inequality of Item 2). Combining inequalities yields $\hat{q}_{\hat{\mathcal{A}}_k} \geq q_{\hat{\mathcal{A}}_k}^{sb} > q_{\hat{\mathcal{A}}_{k+1}}^{sb} \geq \hat{q}_{\hat{\mathcal{A}}_{k+1}}$. This property establishes monotonicity of output, $\hat{q}_i \geq \hat{q}_{i+1}$ for all $i \in \{1, \dots, I-1\}$. Locally downward-binding incentive compatibility, $U_i(\hat{x}_i) = U_i(\hat{x}_{i+1})$ for all $i \in \{1, \dots, I-1\}$, the zero rent condition $U_A(\hat{x}_A) = 0$, and monotonicity of output imply $U_i(\hat{x}_i) \geq U_i(\hat{x}_{i-1})$ for all $i \in \{2, \dots, I\}$. Hence, $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ satisfies (2) and (3) by Claim 1. This completes the proof of Item 1 of the Lemma.

We establish Item 3 by verifying that any incentive efficient mechanism features $W_{\hat{\mathcal{A}}_k}^{sb}(\hat{q}_{\hat{\mathcal{A}}_k}) \geq$

$W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}})$ for all $k \in \{1, \dots, \hat{K} - 1\}$ if the number $\hat{K} \leq \min\{K; I\}$ of non-empty cost groups satisfies $\hat{K} \geq 2$. This holds trivially if $|\hat{\mathcal{A}}_k| = 1$ because then $W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) = w_{\hat{A}_k}^{sb} \geq W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}})$. Assume that $|\hat{\mathcal{A}}_k| \geq 2$. Compare expected surplus $W(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ to what the principal could achieve under a modified mechanism $(\mathbf{x}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ where an agent with marginal cost $\theta_{\hat{A}_k}$ is transferred to a less efficient cost group: $\mathcal{A}_k = \hat{\mathcal{A}}_k \setminus \hat{A}_k$ and $\mathcal{A}_{k+1} = \hat{\mathcal{A}}_{k+1} \cup \hat{A}_k$, $k \leq \hat{K} - 1$. All other cost groups remain unchanged if $\hat{K} \geq 3$. The menu of contracts has the following properties: $x_j = (\hat{q}_j, t_j)$, $t_j = \hat{t}_j - (\theta_{\hat{A}_k} - \theta_{\hat{A}_{k-1}})(q_{\hat{A}_k} - q_{\hat{A}_{k+1}})$ for all $j \in \{1, \dots, \hat{A}_k - 1\}$, $x_{\hat{A}_k} = \hat{x}_{\hat{A}_{k+1}}$, and $x_j = \hat{x}_j$ for all $j \in \{\hat{A}_k + 1, \dots, I\}$. The modified mechanism is incentive feasible. First, it satisfies $U_i(x_i) - U_i(x_{i+1}) = U_i(\hat{x}_i) - U_i(\hat{x}_{i+1}) = 0$ for all $i \in \{1, \dots, \hat{A}_k - 2\}$ if $\hat{A}_k \geq 3$ and for all $i \in \{\hat{A}_k, \dots, I - 1\}$. Moreover, $U_{\hat{A}_{k-1}}(x_{\hat{A}_{k-1}}) = U_{\hat{A}_{k-1}}(\hat{x}_{\hat{A}_k})$ and $U_I(x_I) = U_I(\hat{x}_I) = 0$. Output is monotonic by $q_i = \hat{q}_i$ for all $i \in \mathcal{I}$. These properties imply (2) and (3) by Claim 1. Moreover, $|\mathbf{x}_{\mathcal{A}}| = |\hat{\mathbf{x}}_{\hat{\mathcal{A}}}| = \hat{K} \leq K$. These results verify incentive feasibility of $(\mathbf{x}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$. The difference in expected surplus between the two mechanisms simplifies to

$$W(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset) - W(\mathbf{x}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset) = \nu_{\hat{A}_k} [W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}})] \geq 0,$$

where the inequality follows from the assumed incentive efficiency of $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$. Hence, incentive efficiency implies $W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) \geq W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}})$ also for $|\hat{\mathcal{A}}_k| \geq 2$. By analogous arguments, an agent with marginal cost $\theta_{\hat{A}_{k+1}}$ optimally belongs in cost group $\hat{\mathcal{A}}_{k+1}$ only if $W_{\hat{A}_{k+1}}^{sb}(\hat{q}_{\hat{A}_{k+1}}) \geq W_{\hat{A}_{k+1}}^{sb}(\hat{q}_{\hat{A}_k})$.

One can use the same recipe as above to establish Item 4. ■

Notice the similarity between (23) and (24) and the local downward, upward IC constraints and IR constraint of the agent. The difference is that the constraints of the principal are evaluated using the second-best welfare function $W_i^{sb}(q)$ because this is the correct welfare metric with which to evaluate the trade-off between rent and efficiency.

4.2 Contracting with incomplete commitment

For any incentive feasible mechanism $(\mathbf{x}, \Sigma|\mathcal{A}, \mathcal{B})$ that features ex-post contracting, $\mathcal{B} \neq \emptyset$, recall the definition of $\underline{B}$ and B as the minimal and maximal types, respectively, contained in $\mathcal{B}$. Hence, $\mathcal{B} \subset \{\underline{B}, \dots, B\}$. We do not assume that $\mathcal{B}$ is convex. $\underline{C}$ represents the minimal type contained in $\mathcal{C}$ if the mechanism features partial participation, $\mathcal{C} \neq \emptyset$. We add an asterisk “*” to this notation if the mechanism $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ in question is incentive optimal.

We first establish general results concerning contracts that satisfy conditions (2)-(7) in mechanisms with incomplete commitment.

Lemma 1 *Any incentive feasible mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ that features incomplete commitment ($\mathcal{B} \neq \emptyset$) has the following properties:*

1. $\mathcal{B}$ has at most two unique contracts: $|\mathbf{x}_{\mathcal{B}}| \in \{1, 2\}$.
2. If $|\mathbf{x}_{\mathcal{B}}| = 2$, then:
 - (a) All cost reports θ_j , $j \in \{\underline{B}, \dots, B-1\}$, yield the same contract $x_{\underline{B}} = (q_{\underline{B}}, \theta_{\underline{B}} q_{\underline{B}})$.
 - (b) Cost report θ_B yields contract x_B^{fb} .

Proof. This is a restatement of Claim 8. ■

We now establish some fundamental results of contracts in an incentive optimal mechanism, i.e. one chosen by the principal.

Lemma 2 *Any incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) has the following properties:*

1. The mechanism exploits all available ex-ante contractual flexibility, $|\mathbf{x}_{\mathcal{A}^*}^*| = K$.
2. All cost types contained in $\mathcal{A}^*$ are more efficient than those contained in $\mathcal{B}^*$, and all ex-ante contracts have strictly higher output than all ex-post contracts, $\theta_{A^*} = \theta_{\underline{B}^*-1}$ and $q_{A^*}^* > q_{\underline{B}^*}^*$.
3. All cost types contained in $\mathcal{B}^*$ are more efficient than those contained in $\mathcal{C}^*$ if the mechanism features partial participation, $\theta_{\underline{C}^*} = \theta_{B^*+1}$ if $\mathcal{C}^* \neq \emptyset$.

Proof. The following claim proves Item 1 of the lemma.

Claim 11 *A mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) is incentive optimal only if $|\mathbf{x}_{\mathcal{A}^*}^*| = K$ and $q_j^* \notin \{q_{\underline{B}^*}^*, q_{B^*}^*\}$ for all $j \in \mathcal{A}^*$.*

Proof. Suppose $|\mathbf{x}_{\mathcal{A}^*}^*| < K$, and denote the corresponding number of cost groups by $K^* \leq K-1$. Construct a modified mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}, \mathcal{B})$ as follows: $\mathcal{A}_l = \mathcal{A}_l^*$ for all $l \leq K^*$ if $K^* \geq 1$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then $\mathcal{A}_{K^*+1} = \mathcal{B}^* \setminus B^*$ and $\mathcal{B} = B^*$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, then $\mathcal{A}_{K^*+1} = \mathcal{B}^*$ and $\mathcal{B} = \emptyset$. The modified mechanism is incentive feasible since the menu of contracts and reporting strategies are the same as in the initial mechanism. Both mechanisms also yield the same expected surplus to the principal. Seeing as $\mathcal{B} \subset \mathcal{B}^* \cup \emptyset$, $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ is not minimal in the sense of (10), and cannot be incentive optimal.

Assume next that $|\mathbf{x}_{\mathcal{A}^*}^*| = K$, but $q_j^* \in \{q_{\underline{B}^*}^*, q_{B^*}^*\}$ for some $j \in \mathcal{A}_k^*$. Then $x_{\mathcal{A}_k^*}^* \in \{x_{\underline{B}^*}^*, x_{B^*}^*\}$ by Claim 4. Construct a modified mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}, \mathcal{B})$ as follows: $\mathcal{A}_l = \mathcal{A}_l^*$ for all $l \neq k$ if $K \geq 2$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $x_{\mathcal{A}_k^*}^* = x_{\underline{B}^*}^*$, then $\mathcal{A}_k = \mathcal{A}_k^* \cup \mathcal{B}^* \setminus B^*$ and $\mathcal{B} = B^*$.

If $|\mathbf{x}_{B^*}^*| = 2$ and $x_{\mathcal{A}_k}^* = x_{B^*}^*$, then $\mathcal{A}_k = \mathcal{A}_k^* \cup B^*$ and $\mathcal{B} = \mathcal{B}^* \setminus B^*$. If $|\mathbf{x}_{B^*}^*| = 1$, then $\mathcal{A}_k = \mathcal{A}_k^* \cup B^*$ and $\mathcal{B} = \emptyset$. The proposed mechanism is not minimal in the sense of (10), and therefore cannot be incentive optimal. ■

An immediate implication of Claim 11 is convexity of the incentive optimal escape clause: $\mathcal{B}^* = \{\underline{B}^*, \dots, B^*\}$. This property holds trivially if either $\underline{B}^* = B^* - 1$ or $\underline{B}^* = B^*$. Assume that $\underline{B}^* \leq B^* - 2$. If $j \in \mathcal{A}^*$ for some $\underline{B}^* < j < B^*$, then $q_j^* = q_{\underline{B}^*}^*$ by Claim 7, which violates Claim 11. If $j \in \mathcal{C}^*$ for some $\underline{B}^* < j < B^*$, then $q_j^* = 0 < q_{\underline{B}^*}^*$, which violates monotonicity of output.

Let z^* be the maximal cost type that with positive probability invokes the escape clause by reporting cost θ_j , $j \in \mathcal{B}^*$, in an incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$).

Claim 12 *Assume that the incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ features incomplete commitment ($\mathcal{B}^* \neq \emptyset$), where $B^* \leq I - 1$. If $q_{B^*+1}^* > 0$, then $z^* = B^* + 1$.*

Proof. We first demonstrate that $z^* \leq B^* + 1$ generally holds. This is obviously true if $B^* \in \{I - 1, I\}$, but the result holds also for $B^* \leq I - 2$. For if $z^* \geq B^* + 2$, then $q_j^* = q_{\underline{B}^*}^*$ for all $j \in \{\underline{B}^*, \dots, z^* - 1\}$ by Claim 7. In particular, $q_{B^*+1}^* = q_{\underline{B}^*}^*$, which violates the necessary condition of incentive optimality established in Claim 11. Invoking Claim 7 delivers $z^* \in \{B^*, B^* + 1\}$. Assume that $q_{B^*+1}^* > 0$. If $z^* = B^*$, then $U_{B^*}(x_{B^*}^*) = 0 < (\theta_{B^*+1} - \theta_{B^*})q_{B^*+1}^* = U_{B^*}(x_{B^*+1}^*)$, which violates incentive compatibility. This leaves $z^* = B^* + 1$ as the only remaining possibility if $q_{B^*+1}^* > 0$. ■

Claim 13 *A mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) is incentive optimal only if $z^* = B^*$.*

Proof. The result follows directly if $B^* = I$ since we already established $z^* \geq B^*$ in Claim 7. Let $B^* \leq I - 1$. The proof proceeds as follows: We first show that $x_{B^*+1}^* = x_{B^*+1}^{fb}$ if $q_{B^*+1}^* > 0$. We then show that the principal in that case can obtain strictly higher expected surplus than in the proposed mechanism by modifying the escape clause. Hence, incentive optimality implies $q_{B^*+1}^* = 0$. We already showed in the proof of Claim 12 that $z^* \in \{B^*, B^* + 1\}$. The final part of the proof establishes that $z^* \neq B^* + 1$ if $q_{B^*+1}^* = 0$. This leaves $z^* = B^*$ as the only remaining possibility for $B^* \leq I - 1$.

It cannot be the case that $q_{B^*+1}^* = q_{B^*}^*$, because this would violate Claim 11. If $q_{B^*+1}^* \in (0, q_{B^*}^*)$, then $z^* = B^* + 1$ by Claim 12. Hence, $x_j^* = x_{\underline{B}^*}^*$ for all $j \in \mathcal{B}^*$ by Claim 7. Moreover, $\mathcal{A}_K^* = \{B^* + 1\}$ identifies the maximal cost group in $\mathcal{A}^*$ because $x_j^* = x_0$ for all $j \in \{B^* + 2, \dots, I\}$ if $B^* \leq I - 2$; see Claim 7. The local downward incentive compatibility constraint $U_{B^*}(x_{B^*}^*) \geq U_{B^*}(x_{B^*+1}^*)$ is slack because $U_{B^*+1}(x_{B^*+1}^*) = U_{B^*+1}(x_{B^*}^*)$ and $q_{B^*}^* >$

$q_{B^*+1}^*$. By Claim 2, it follows that $U_i(x_i^*) > U_i(x_j^*)$, and therefore $\sigma_{ji}^* = 0$, for all $(i, j) \in \{1, \dots, B^*\} \times \{B^*+1, \dots, I\}$. If $B^* \leq I-2$, then upward-binding IC and strict monotonicity also imply $\sigma_{ji}^* = 0$ for all $(i, j) \in \{B^*+2, \dots, I\} \times \{1, \dots, B^*\}$ by Claim 3. Moreover, $U_i(x_i^*) = U_i(x_0) = 0 > -(\theta_i - \theta_{B^*+1})q_{B^*+1}^* = U_i(x_{B^*+1}^*)$ imply $\sigma_{(B^*+1)i}^* = 0$ for all $i \in \{B^*+2, \dots, I\}$. In conclusion, $U_i(x_i^*) > U_i(x_{B^*+1}^*)$ for all $i \neq B^*+1$ if $q_{B^*+1}^* \in (0, q_{B^*}^*)$. As the principal cannot reduce the information rent by distorting $q_{B^*+1}^*$, it follows that $x_{B^*+1}^* = x_{B^*+1}^{fb}$. Finally, $\sigma_{j(B^*+1)}^* = 0$ for all $j \in \{1, \dots, B^*-1\}$ if $B^* \geq 2$ by $U_{B^*}(x_{B^*}^*) \geq U_{B^*}(x_{B^*-1}^*)$, $q_{B^*-1}^* > q_{B^*}^*$ and Claim 3. Based on this information, we can write the principal's expected surplus of the proposed incentive optimal mechanism as:

$$W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{j=1}^{B^*} \sum_{i=1}^{B^*} \nu_i \sigma_{ji}^* W_i(x_j^*) + \nu_{B^*+1} \left[\sum_{j \in B^*} \sigma_{j(B^*+1)}^* W_{B^*+1}(x_{B^*}^*) + \sigma_{B^*+1}^* w_{B^*+1}^{fb} \right].$$

Consider the alternative mechanism $(\mathbf{x}^*, \Sigma | \mathcal{A}, \mathcal{B})$, where $\mathcal{A}_l = \mathcal{A}_l^*$ for all $l \in \{1, \dots, K-1\}$, if $K \geq 2$, $\mathcal{A}_K = \mathcal{B}^*$ and $\mathcal{B} = \{B^*+1\}$. Also, let $\sigma_{B^*+1} = 1$. Reporting strategies remain unchanged otherwise. Setting $x_{B^*+1} = x_{B^*+1}^{fb} = x_{B^*+1}^*$ is sequentially rational following the cost report θ_{B^*+1} in the modified mechanism: $U_i(x_i^*) > U_i(x_{B^*+1}^*)$ for all $i \neq B^*+1$ implies $\sigma_{(B^*+1)i} = 0$ for all $i \neq B^*+1$, which in turn implies that the principal attaches posterior probability equal to 1 to the event that the agent has cost θ_{B^*+1} after observing that particular cost report. The expected surplus of the modified mechanism equals

$$W(\mathbf{x}^*, \Sigma | \mathcal{A}, \mathcal{B}) = \sum_{j=1}^{B^*} \sum_{i=1}^{B^*} \nu_i \sigma_{ji}^* W_i(x_j^*) + \nu_{B^*+1} w_{B^*+1}^{fb}.$$

The difference

$$\begin{aligned} W(\mathbf{x}^*, \Sigma | \mathcal{A}, \mathcal{B}) - W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) &= \nu_{B^*+1} (1 - \sigma_{B^*+1}^* - \sum_{j \in B^*} \sigma_{j(B^*+1)}^*) w_{B^*+1}^{fb} \\ &\quad + \nu_{B^*+1} \sum_{j \in B^*} \sigma_{j(B^*+1)}^* [w_{B^*+1}^{fb} - W_{B^*+1}(x_{B^*}^*)] \end{aligned}$$

in expected surplus between the two mechanisms is strictly positive by $x_{B^*}^* \neq x_{B^*+1}^* = x_{B^*+1}^{fb}$, and because $z^* = B^*+1$ it follows that $\sum_{j \in B^*} \sigma_{j(B^*+1)}^* > 0$. Having eliminated all other possibilities, it follows that $q_{B^*+1}^* = 0$.

We next establish $z^* \neq B^*+1$ if $q_{B^*+1}^* = 0$. Suppose $z^* = B^*+1$. Everything is nearly the same as in the previous part of the proof, except now $x_{B^*+1}^* = x_0$ instead of $x_{B^*+1}^* = x_{B^*+1}^{fb}$. In particular, the expected surplus of the proposed incentive optimal

mechanism is:

$$W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{j=1}^{B^*} \sum_{i=1}^{B^*} \nu_i \sigma_{ji}^* W_i(x_j^*) + \nu_{B^*+1} \sum_{j \in B^*} \sigma_{j(I_{B^*+1})}^* W_{I_{B^*+1}}(x_{\underline{B}^*}^*).$$

Consider a modified mechanism $(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B})$, where $\mathcal{B} = \{\underline{B}^*, \dots, B^* + 1\}$, $x_{B^*+1} = x_{B^*+1}^{fb}$ and $\sigma_i = \sigma_i^* + \sigma_{(B^*+1)i}^*$ for all $i \in \{B^* + 2, \dots, I\}$ if $B^* \leq I - 2$. All other contracts and reporting strategies remain the same as in the initial mechanism. Even this mechanism is locally upward-binding at θ_{B^*+1} , $U_{B^*+1}(x_{B^*+1}^{fb}) = U_{B^*+1}(x_{\underline{B}^*}^*) = 0$, and is incentive feasible if $q_{\underline{B}^*}^* > q_{B^*+1}^{fb}$.

We now demonstrate $q_{\underline{B}^*}^* > q_{B^*+1}^{fb}$. On the basis of the locally upward-binding IC constraint $U_{B^*+1}(x_{B^*+1}^*) = U_{B^*+1}(x_{\underline{B}^*}^*)$, monotonicity $q_{\underline{B}^*}^* > 0 = q_{B^*+1}^*$ and Claim 3, we obtain $\sigma_{ji}^* = 0$ for all $(i, j) \in \{B^* + 2, \dots, I\} \times \{1, \dots, B^*\}$ if $B^* \leq I - 2$. Upon observing a cost report θ_j , $j \in \mathcal{B}^*$, the principal therefore obtains the expected ex-post surplus

$$S(q_{\underline{B}^*}^*) - \sum_{i=1}^{B^*+1} \mu_{ji}^* [\theta_i + (1 - \alpha)(\theta_{B^*+1} - \theta_i)] q_{\underline{B}^*}^*, \quad \mu_{ji}^* = \frac{\nu_i \sigma_{ji}^*}{\sum_{h=1}^{B^*+1} \nu_h \sigma_{jh}^*},$$

of offering the contract $x_{\underline{B}^*}^* = (q_{\underline{B}^*}^*, \theta_{B^*+1} q_{\underline{B}^*}^*)$. The equilibrium quantity $q_{\underline{B}^*}^*$ is then characterized by

$$S'(q_{\underline{B}^*}^*) = \sum_{i=1}^{B^*+1} \mu_{ji}^* [\theta_i + (1 - \alpha)(\theta_{B^*+1} - \theta_i)] < \theta_{B^*+1} = S'(q_{B^*+1}^{fb}),$$

where the inequality follows from

$$\theta_{B^*+1} - \sum_{i=1}^{B^*+1} \mu_{ji}^* [\theta_i + (1 - \alpha)(\theta_{B^*+1} - \theta_i)] = \alpha \frac{\sum_{i=1}^{B^*} \nu_i \sigma_{ji}^* (\theta_{B^*+1} - \theta_i)}{\sum_{h=1}^{B^*+1} \nu_h \sigma_{jh}^*} > 0.$$

Strict concavity of $S(q)$ then implies $q_{\underline{B}^*}^* > q_{B^*+1}^{fb}$. The expected surplus in the modified mechanism is

$$W(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B}) = \sum_{j=1}^{B^*} \sum_{i=1}^{B^*} \nu_i \sigma_{ji}^* W_i(x_j^*) + \nu_{B^*+1} \left[\sum_{j \in B^*} \sigma_{j(I_{B^*+1})}^* W_{I_{B^*+1}}(x_{\underline{B}^*}^*) + \sigma_{B^*+1}^* w_{B^*+1}^{fb} \right].$$

The difference in expected surplus between the two mechanisms is

$$W(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B}) - W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \nu_{B^*+1} \sigma_{B^*+1}^* w_{B^*+1}^{fb} > 0,$$

which contradicts the assumed incentive optimality of $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$. We conclude that

$q_{B^*+1}^* = 0$ implies $z^* \neq B^* + 1$. ■

We can now draw conclusions about incentive optimal mechanisms with incomplete commitment. Claim 11 proves Item 1 of Lemma 2. By way of $z^* = B^*$ and Item 2 of Claim 7, $\mathcal{C}^* = \{B^* + 1, \dots, I\}$ if $B^* \leq I - 1$. This proves Item 3 of Lemma 2. Moreover, $\mathcal{B}^* \cup \mathcal{C}^* = \{\underline{B}^*, \dots, I\}$. By $\mathcal{A}^* \neq \emptyset$, and since $\mathcal{A}^*$, $\mathcal{B}^*$ and $\mathcal{C}^*$ partition $\mathcal{I} \cup \emptyset$, it follows that $\underline{B}^* \geq 2$ and $\mathcal{A}^* = \{1, \dots, A^*\}$, where $A^* = \underline{B}^* - 1$. This proves Item 2 of Lemma 2. ■

Consider next the incentive optimal reporting strategy Σ^* by the agent. Denote by $\underline{\mathcal{B}}^*$ the set of types such that any marginal cost report θ_j , $j \in \underline{\mathcal{B}}^*$, induces the principal to offer the ex-post contract $x_{\underline{B}^*}^*$.¹ Lemma 3 in Tangerås and Gick (2026) derives fundamental properties of incentive optimal reporting strategies under the simplifying assumption that $|\mathcal{B}^*| \geq 3$. The following lemma provides a characterization for the general case of $\mathcal{B}^* \neq \emptyset$.

Lemma 3 *For any incentive optimal mechanism $(\mathbf{x}^*, \Sigma^{**} | \mathcal{A}^*, \mathcal{B}^*)$ featuring incomplete commitment ($\mathcal{B}^* \neq \emptyset$), there exists an incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ and reporting strategy Σ^* with the following properties:*

1. *With respect to cost types in $\mathcal{A}^*$:*

- (a) *All but the upper boundary type A^* truthfully report their cost, $\sigma_i^* = 1 \ \forall i \in \{1, \dots, A^* - 1\}$ if $|\mathcal{A}^*| \geq 2$.*
- (b) *The upper boundary type A^* may invoke the escape clause, and then randomizes uniformly across all cost types contained in $\underline{\mathcal{B}}^*$, $\sigma_{A^*}^* > 0$ and $\sigma_{jA^*}^* = \frac{1 - \sigma_{A^*}^*}{|\underline{\mathcal{B}}^*|} \ \forall j \in \underline{\mathcal{B}}^*$.*

2. *With respect to cost types in $\mathcal{B}^*$:*

- (a) *The lower boundary type $\underline{B}^*$:*
 - i. *truthfully reveals its cost if the escape clause contains one type, $\sigma_{\underline{B}^*}^* = 1$ if $|\mathcal{B}^*| = 1$;*
 - ii. *may choose an ex-ante contract if the escape clause contains two types and two distinct contracts, $1 > \sigma_{A^* \underline{B}^*}^* = 1 - \sigma_{\underline{B}^*}^* \geq 0$ if $|\mathcal{B}^*| = 2$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$. In that case, the A^* type invokes the escape clause with zero probability, $(1 - \sigma_{\underline{B}^*}^*)(1 - \sigma_{A^*}^*) = 0$;*
 - iii. *uniformly randomizes across all types in $\underline{\mathcal{B}}^*$ otherwise, $\sigma_{j \underline{B}^*}^* = \frac{1}{|\underline{\mathcal{B}}^*|} \ \forall j \in \underline{\mathcal{B}}^*$ if $|\mathcal{B}^*| = 2$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$ or if $|\mathcal{B}^*| \geq 3$.*
- (b) *Intermediary types randomize uniformly across all types in $\underline{\mathcal{B}}^*$, $\sigma_{ji}^* = \frac{1}{|\underline{\mathcal{B}}^*|} \ \forall (i, j) \in \{B^* + 1, B^* - 1\} \times \underline{\mathcal{B}}^*$ if $|\mathcal{B}^*| \geq 3$.*

¹Formally, $\underline{\mathcal{B}}^* = \{\underline{B}^*, \dots, B^* - 1\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, and $\underline{\mathcal{B}}^* = \mathcal{B}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$; see Lemma 2.

(c) The upper boundary type B^* randomizes between all types of cost reports that yield ex-post contracting:

- i. $0 < \sigma_{B^*}^* < 1$ and $\sigma_{jB^*}^* = \frac{1-\sigma_{B^*}^*}{|\underline{\mathcal{B}}^*|} \forall j \in \underline{\mathcal{B}}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$;
- ii. $\sigma_{jB^*}^* = \frac{1}{|\mathcal{B}^*|} \forall j \in \mathcal{B}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$.

3. Cost types in $\mathcal{C}^*$ truthfully report their cost if the mechanism features partial participation, $\sigma_i^* = 1 \forall i \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$.

Proof. We first demonstrate some general properties of Σ^* in incentive optimal mechanisms $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ with incomplete commitment ($\mathcal{B}^* \neq \emptyset$). This is done in 6 claims. Claim 14 and Claim 15 establish combinations of cost types and cost reports that *cannot* occur in an incentive optimal mechanism.

Claim 14 Consider a mechanism $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$). This mechanism is incentive optimal only if the following conditions are all met:

- 1. $\sigma_{ji}^* = 0 \forall (i, j) \in \{1, \dots, A^* - 1\} \times \{\underline{B}^*, \dots, I\}$ if $A^* \geq 2$.
- 2. $\sigma_{ji}^* = 0 \forall (i, j) \in \{1, \dots, B^* - 1\} \times \{B^*, \dots, I\}$ if $q_{\underline{B}^*}^* > q_{B^*}^*$.
- 3. $\sigma_{ji}^* = 0 \forall (i, j) \in \{1, \dots, B^* - 1\} \times \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$.
- 4. $\sigma_{ji}^* = 0 \forall (i, j) \in \{\underline{B}^* + 1, \dots, I\} \times \mathcal{A}^*$ if $\underline{B}^* \leq I - 1$.
- 5. $\sigma_{j\underline{B}^*}^* = 0 \forall j \in \{1, \dots, A^* - 1\}$ such that $q_j^* > q_{A^*}^*$, if $A^* \geq 2$.
- 6. $\sigma_{ji}^* = 0 \forall (i, j) \in \mathcal{C}^* \times (\mathcal{A}^* \cup \mathcal{B}^*)$ if $\mathcal{C}^* \neq \emptyset$.

Proof. By combining incentive compatibility conditions:

$$U_i(x_i^*) - U_i(x_j^*) = U_i(x_i^*) - U_i(x_h^*) + U_h(x_h^*) - U_h(x_j^*) + (\theta_h - \theta_i)(q_h^* - q_j^*) \geq (\theta_h - \theta_i)(q_h^* - q_j^*). \quad (29)$$

Hence, $U_i(x_i^*) > U_i(x_j^*)$ and therefore $\sigma_{ji}^* = 0$ if $(\theta_h - \theta_i)(q_h^* - q_j^*) > 0$ for some $h \in \mathcal{I}$.

Item 1: If $h = A^*$, then the rightmost expression in (29) is strictly positive for all $(i, j) \in \{1, \dots, A^* - 1\} \times \{\underline{B}^*, \dots, I\}$ by $q_{A^*}^* > q_{\underline{B}^*}^* \geq q_j^*$ for all $j \in \{\underline{B}^*, \dots, I\}$; see Lemma 2.

Item 2: If $h = B^* - 1$, then the rightmost expression in (29) is strictly positive for all $(i, j) \in \{1, \dots, B^* - 2\} \times \{B^*, \dots, I\}$ by $q_{B^*-1}^* > q_{B^*}^* \geq q_j^*$ for all $j \in \{B^*, \dots, I\}$.

Item 3: If $h = B^*$, then the rightmost expression in (29) is strictly positive for all $(i, j) \in \{1, \dots, B^* - 1\} \times \mathcal{C}^*$ by $q_{B^*}^* > 0$.

Item 4: If $h = \underline{B}^*$, then the rightmost expression in (29) is strictly positive for all $(i, j) \in \{\underline{B}^* + 1, \dots, I\} \times \mathcal{A}^*$ by $q_j^* \geq q_{A^*}^* > q_{\underline{B}^*}^*$ for all $j \in \mathcal{A}^*$.

Item 5: If $h = A^*$, then the rightmost expression in (29) is strictly positive for all $j \in \{1, \dots, A^* - 1\}$ that satisfy $q_j^* > q_{A^*}^*$.

Item 6: If $h = B^*$, then the rightmost expression in (29) is strictly positive for all $(i, j) \in \mathcal{C}^* \times \mathcal{A}^*$ by $q_j^* \geq q_{A^*}^* > q_{\underline{B}^*}^* \geq q_{B^*}^*$ for all $j \in \mathcal{A}^*$. $U_i(x_i^*) = 0 > -(\theta_i - \theta_{B^*})q_j^* = U_i(x_j^*)$ for all $(i, j) \in \mathcal{C}^* \times \mathcal{B}^*$ completes the proof. ■

Claim 15 *A mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) and partial participation ($\mathcal{C}^* \neq \emptyset$), is incentive optimal only if $\sum_{j \in \mathcal{C}^*} \sigma_{jB^*}^* = 0$.*

Proof. The method of proof is to show that an incentive feasible mechanism which strictly outperforms $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ exists if $\sum_{j \in \mathcal{C}^*} \sigma_{jB^*}^* > 0$. We consider two cases separately.

In case one, $q_{\underline{B}^*}^* > q_{B^*}^*$. By Claim 14, $\sigma_{B^*i}^* = 0$ for all $i \neq B^*$. Upon observing θ_{B^*} , the principal therefore deduces that the agent with probability one has marginal cost θ_{B^*} . The sequentially rational ex-post contract then equals $x_{B^*}^* = x_{B^*}^{fb}$. This holds for any $\sigma_{B^*}^* > 0$. The expected surplus of the principal equals

$$W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{i=1}^{B^*-1} \sum_{j=1}^{B^*-1} \nu_i \sigma_{ji}^* W_i(x_j^*) + \sum_{j=\underline{B}^*}^{B^*-1} \nu_{B^*} \sigma_{jB^*}^* W_{B^*}(x_{\underline{B}^*}^*) + \nu_{B^*} \sigma_{B^*}^* w_{B^*}^{fb}.$$

Let a modified mechanism $(\mathbf{x}^*, \Sigma | \mathcal{A}^*, \mathcal{B}^*)$ differ from the previous mechanism only by $\sigma_{B^*} = \sum_{j=B^*}^I \sigma_{jB^*}^*$. The principal can implement $\mathbf{x}^*$ also under the modified reporting strategy because the change from Σ^* to Σ does not affect posterior beliefs about the agent's true cost type θ_i upon observing cost report θ_j , $j \in \mathcal{B}^*$. The difference

$$W(\mathbf{x}^*, \Sigma | \mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{j \in \mathcal{C}^*} \nu_{B^*} \sigma_{jB^*}^* w_{B^*}^{fb},$$

in the principal's expected surplus is strictly positive if $\sum_{j \in \mathcal{C}^*} \sigma_{jB^*}^* > 0$, which would contradict the assumed incentive optimality of $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$.

In case two, $q_{\underline{B}^*}^* = q_{B^*}^*$, so that $x_j^* = x_{\underline{B}^*}^*$ for all $j \in \mathcal{B}^*$. We now introduce some notation that will be useful later. Recall the definition $\underline{\mathcal{B}}^* = \{\underline{B}^*, \dots, B^* - 1\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $\underline{\mathcal{B}}^* = \mathcal{B}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$. After observing a cost report $j \in \underline{\mathcal{B}}^*$, the principal's option is whether to offer the contract $x_{\underline{B}^*}^*$ or save on information rent by excluding one or more of the least efficient cost types. The maximal surplus the principal can achieve by offering a deviation contract $x_{jh}^{*d} = (q_{jh}^{*d}, \theta_h q_{jh}^{*d})$ in Stage 4 that leaves an agent of cost type $h \in \{A^*, \dots, B^* - 1\}$ indifferent between accepting or rejecting the ex-post contract, equals

$\frac{\Omega_{jh}(\sigma_{jh}^*)}{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^*}$, where

$$\Omega_{jh}(\sigma_{jh}) = \sum_{i=A^*}^h \nu_i \sigma_{ji} W_i(x_{jh}^d) = \sum_{i=A^*}^h \nu_i \sigma_{ji} [S(q_{jh}^d) - (\theta_i + (1 - \alpha)(\theta_h - \theta_i))q_{jh}^d], \quad (30)$$

$x_{jh}^d = (q_{jh}^d, \theta_h q_{jh}^d)$ is the ex-post contract offered by the principal in that case, and

$$S'(q_{jh}^d) = \frac{\sum_{i=A^*}^h \nu_i \sigma_{ji} (\theta_i + (1 - \alpha)(\theta_h - \theta_i))}{\sum_{i=A^*}^h \nu_i \sigma_{ji}}, \quad (31)$$

characterizes the optimal output given the reporting strategy $\sigma_{jh} = (\sigma_{jA^*}, \dots, \sigma_{jh})$. The contract x_{jh}^{*d} results from replacing σ_{jh} by σ_{jh}^* in (30) and (31). The Stage 4 expected surplus of offering $x_{\underline{B}^*}^*$ subsequent to a cost report θ_j , $j \in \underline{B}^*$, equals $\frac{\Omega_{jB^*}^*}{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^*}$, where

$$\Omega_{jB^*}^* = \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* W_i(x_{\underline{B}^*}^*) = \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* [S(q_{\underline{B}^*}^*) - (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i))q_{\underline{B}^*}^*]$$

By these definitions, $x_{\underline{B}^*}^*$ is sequentially rational if and only if

$$\Omega_{jB^*}^* \geq \Omega_{jh}(\sigma_{jh}^*) \quad \forall (j, h) \in \underline{B}^* \times \{A^*, \dots, B^* - 1\}. \quad (32)$$

In particular, $x_{\underline{B}^*}^*$ is sequentially rational only if $W_{B^*}(x_{\underline{B}^*}^*) \geq 0$. Otherwise, the principal would be strictly better off by excluding the least efficient cost type under ex-post contracting and offering instead a deviation contract.

Let us continue with the specific case of $q_{\underline{B}^*}^* = q_{B^*}^*$, so that $|\mathbf{x}_{B^*}^*| = 1$. Suppose $\sigma_{lB^*}^* > 0$ for some $l \in \mathcal{C}^*$. Construct a modified mechanism $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ where $\sigma_{lB^*} = \sigma_{lB^*}^* - \epsilon \geq 0$, $\epsilon > 0$, and

$$\sigma_{jB^*} = \sigma_{jB^*}^* + \frac{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* (\theta_{B^*} - \theta_i)}{\sum_{j' \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{j'i}^* (\theta_{B^*} - \theta_i)} \epsilon \quad \forall j \in \mathcal{B}^*.$$

All other reporting strategies remain the same as before. By this construction, $\sum_{j \in \mathcal{B}^*} (\sigma_{jB^*} - \sigma_{jB^*}^*) = \epsilon$. Also, the contract $x_{B^*} = (q_{B^*}, \theta_{B^*} q_{B^*})$, where

$$S'(q_{B^*}) = \frac{\sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) + \nu_{B^*} \theta_{B^*} \epsilon}{\sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* + \nu_{B^*} \epsilon} < \theta_{B^*}$$

is sequentially rational for all cost reports $j \in \mathcal{B}^*$ if and only if

$$\Omega_{jB^*}(\sigma_{jB^*}) \geq \Omega_{jh}(\sigma_{jh}^*) \quad \forall (j, h) \in \mathcal{B}^* \times \{A^*, \dots, B^* - 1\}. \quad (33)$$

A marginal increase in ϵ has no effect on the right-hand side of (33). The marginal effect of ϵ on x_{B^*} has only a second-order effect on the principal's surplus, i.e. $\frac{\partial \Omega_{jB^*}}{\partial \epsilon} = \nu_{B^*} \frac{\partial \sigma_{jB^*}}{\partial \epsilon} W_{B^*}(x_{B^*})$. The derivative

$$\frac{\partial W_{B^*}(x_{B^*})}{\partial \epsilon} = (S'(q_{B^*}) - \theta_{B^*}) \frac{\partial q_{B^*}}{\partial \epsilon} = -\alpha \frac{\sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* (\theta_{B^*} - \theta_i)}{\sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* + \nu_{B^*} \epsilon} \frac{\partial q_{B^*}}{\partial \epsilon}$$

is strictly positive by

$$\frac{\partial q_{B^*}}{\partial \epsilon} = \frac{\alpha \nu_{B^*} \sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* (\theta_{B^*} - \theta_i)}{S''(q_{B^*}) (\sum_{j \in \mathcal{B}^*} \sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^* + \nu_{B^*} \epsilon)^2} < 0.$$

Moreover, $\frac{\partial \Omega_{jB^*}}{\partial \epsilon} > \nu_{B^*} \frac{\partial \sigma_{jB^*}}{\partial \epsilon} W_{B^*}(x_{B^*}^*) \geq 0$ for all $\epsilon > 0$ by $W_{B^*}(x_{B^*}) - W_{B^*}(x_{B^*}^*) = W_{B^*}^{fb}(q_{B^*}) - W_{B^*}^{fb}(q_{B^*}^*)$, $q_{B^*}^{fb} < q_{B^*} < q_{B^*}^*$ for $\epsilon > 0$, and $W_{B^*}(x_{B^*}^*) \geq 0$. Hence, x_{B^*} is sequentially rational for all $\epsilon > 0$.

The key question is how ϵ affects the principal's ex-ante expected surplus. If $U_{A^*}(x_{A^*}^*) = U_{A^*}(x_{B^*}^*)$, then

$$\frac{\partial}{\partial \epsilon} W(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B}^*) = \nu_{B^*} W_{B^*}(x_{B^*}) - (G_{A^*-1} + \nu_{A^*} \sigma_{A^*}^*) (1 - \alpha) (\theta_{B^*} - \theta_{A^*}) \frac{\partial q_{B^*}}{\partial \epsilon} > 0.$$

The case with $U_{B^*}(x_{A^*}^*) = U_{B^*}(x_{B^*}^*)$ is qualitatively similar. We conclude that $\sigma_{jB^*}^* = 0$ for all $j \in \mathcal{C}^*$ also when $q_{B^*}^* = q_{B^*}^{fb}$. ■

Claim 16 shows that an agent with marginal cost θ_{B^*} truthfully reveals its cost in the special case where the escape clause only contains that one type, $\mathcal{B}^* = B^*$.

Claim 16 *A mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features ex-ante contracting ($\mathcal{A}^* \neq \emptyset$) and incomplete commitment ($\mathcal{B}^* = B^*$) is incentive optimal only if $\sigma_{B^*}^* = 1$.*

Proof. The method of proof is to show that an incentive feasible mechanism which strictly outperforms $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ exists if $\sigma_{B^*}^* < 1$. Item 1 of Claim 14 implies $\sum_{j=1}^{A^*-1} \sigma_{jB^*}^* = 0$ if $A^* \geq 2$. Claim 15 implies $\sum_{j=B^*+1}^I \sigma_{jB^*}^* = 0$ if $B^* \leq I - 1$. Hence, $\sigma_{A^*B^*}^* + \sigma_{B^*}^* = 1$ if $\underline{B}^* = B^*$. $\sigma_{A^*B^*}^* > 0$ only if $U_{B^*}(x_{B^*}^*) = U_{B^*}(x_{A^*}^*)$. Then, $U_{A^*}(x_{A^*}^*) > U_{A^*}(x_{B^*}^*)$ by strict monotonicity $q_{A^*}^* > q_{B^*}^*$. As we have previously verified, $\sigma_{B^*i} = 0$ for all $i \neq B^*$ in those conditions, which establishes $x_{B^*}^* = x_{B^*}^{fb}$. The expected surplus then equals

$$W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i \sigma_{ji}^* W_i(x_j^*) + (1 - \sigma_{B^*}^*) W_{B^*}(x_{A^*}^*) + \sigma_{B^*}^* w_{B^*}^{fb}.$$

Consider a modified mechanism $(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B}^*)$ that differs from the original mechanism by a reduced transfer $t_{A_k^*} = t_{A_k^*}^* - (\theta_{B^*} - \theta_{A^*})(q_{A^*}^* - q_{B^*}^*)$ to all cost groups A_k^* , $k \in \{1, \dots, K\}$,

and by $\sigma_{B^*} = 1$. Everything else is the same as in the original mechanism. This mechanism is incentive feasible and has expected surplus

$$W(\mathbf{x}, \Sigma | \mathcal{A}^*, \mathcal{B}^*) = W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) + G_{A^*}(\theta_{B^*} - \theta_{A^*})(q_{A^*}^* - q_{B^*}^*) + (1 - \sigma_{B^*}^*)(w_{B^*}^{fb} - W_{B^*}(x_{A^*}^*)),$$

which is strictly larger than $W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$. Hence, $\mathcal{B}^* = B^*$ implies $\sigma_{B^*}^* = 1$. ■

Claim 17 applies the revelation principle to the menu of *ex-ante* contracts and invokes the three previous claims.

Claim 17 *For any incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$), there exists an incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ where the reporting strategy Σ^* has the following properties:*

1. $\sigma_i^* = 1 \ \forall i \in \{1, \dots, A^* - 1\}$ if $A^* \geq 2$.
2. $\sum_{j=A^*}^{B^*-1} \sigma_{ji}^* = 1$, $i \in \{A^*, \underline{B}^*\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$.
3. $\sum_{j=A^*}^{B^*} \sigma_{ji}^* = 1$, $i \in \{A^*, \underline{B}^*\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$.
4. $\sigma_{A^* \underline{B}^*}^*(1 - \sigma_{A^*}^*) = 0$.
5. $\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji}^* = 1$, $i \in \{\underline{B}^* + 1, \dots, B^* - 1\}$ if $\underline{B}^* \leq B^* - 2$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$.
6. $\sum_{j \in \mathcal{B}^*} \sigma_{ji}^* = 1$, $i \in \{\underline{B}^* + 1, \dots, B^* - 1\}$ if $\underline{B}^* \leq B^* - 2$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$.
7. $\sum_{j \in \mathcal{B}^*} \sigma_{jB^*}^* = 1$.
8. $\sigma_{jB^*}^* > 0 \ \forall j \in \mathcal{B}^*$.
9. $\sigma_i^* = 1 \ \forall i \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$.

Proof. Construct Σ^* as follows: If $A^* \geq 2$, then $\sigma_i^* = 1 \ \forall i \in \{1, \dots, A^* - 1\}$. For $i \in \{A^*, \underline{B}^*\}$, $\sigma_{A^*i}^* = \sum_{j \in \mathcal{A}^*} \sigma_{ji}^{**}$ and $\sigma_{ji}^* = \sigma_{ji}^{**} \ \forall j \in \{\underline{B}^*, \dots, I\}$. Moreover, $\sigma_{ji}^* = \sigma_{ji}^{**} \ \forall (i, j) \in \{\underline{B}^* + 1, \dots, B^*\} \times \mathcal{I}$ if $\underline{B}^* \leq B^* - 1$, and finally $\sigma_i^* = 1 \ \forall i \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$. The modification from Σ^{**} to Σ^* does not affect posterior beliefs for any reported $j \in \mathcal{B}^*$ in Stage 4 of the game. Hence, $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ is incentive feasible. To derive incentive optimality, observe that

$$\nu_i(\sigma_{ji}^* - \sigma_{ji}^{**})W_i(x_j^*) = 0 \ \forall (i, j) \in \mathcal{I} \times \{\underline{B}^*, \dots, I\} \text{ and } \forall (i, j) \in \{\underline{B}^* + 1, \dots, I\} \times \mathcal{A}^* \text{ if } \underline{B}^* \leq I - 1$$

because either $\sigma_{ji}^* = \sigma_{ji}^{**}$ or $W_i(x_j^*) = 0$ in all those cases. This result explains the second row below:

$$\begin{aligned}
W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}^*, \Sigma^{**} | \mathcal{A}^*, \mathcal{B}^*) &= \sum_{i=1}^{B^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**}) W_i(x_j^*) \\
&= \sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**}) W_i(x_j^*) \\
&= \sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**}) [W_i(x_j^*) - W_i(x_i^*) + W_i(x_i^*)] \\
&= \sum_{i=1}^{A^*} \nu_i [\sum_{j=1}^{A^*} \sigma_{ji}^* W_i(x_j^*) - \sum_{j=1}^{A^*} \sigma_{ji}^{**} W_i(x_i^*)] \\
&= \sum_{i=1}^{A^*} \nu_i [\sigma_i^* - \sum_{j=1}^{A^*} \sigma_{ji}^{**}] W_i(x_i^*) = 0.
\end{aligned} \tag{34}$$

The third row follows from

$$\begin{aligned}
\sum_{j=1}^{A^*} \nu_{B^*}(\sigma_{jB^*}^* - \sigma_{jB^*}^{**}) W_{B^*}(x_j^*) \\
= \sum_{j=1}^{A^*} \nu_{B^*}(\sigma_{jB^*}^* - \sigma_{jB^*}^{**}) W_{B^*}(x_{A^*}^*) = 0
\end{aligned}$$

where

$$\sum_{j=1}^{A^*} \sigma_{jB^*}^* = \sigma_{A^*B^*}^* = \sum_{j=1}^{A^*} \sigma_{jB^*}^{**}$$

by construction of Σ^* . In the fourth row, we have added and subtracted $W_i(x_i^*)$ inside the square brackets. The fifth row follows from:

$$\sigma_{ji}^{**} (W_i(x_i^*) - W_i(x_j^*)) = 0 \quad \forall (i, j) \in \mathcal{A}^* \times \mathcal{A}^*.$$

This property holds trivially for $\sigma_{ji}^{**} = 0$. It hold also for $\sigma_{ji}^{**} > 0$ because then $W_i(x_i^*) = W_i(x_j^*)$. If $W_i(x_i^*) > W_i(x_j^*)$, then it would have been better to set $\sigma_{ji}^{**} = 0$. For $W_i(x_i^*) < W_i(x_j^*)$, it would have been better to set $\sigma_i^{**} = 0$, which would violate incentive feasibility. The first equality in the last row of (34) follows from $\sigma_{ji}^* = 0 \quad \forall (i, j) \in \mathcal{A}^* \times \mathcal{A}^*, i \neq j$. The second equality is implied by $\sigma_i^* = 1 = \sum_{j=1}^{A^*} \sigma_{ji}^{**} \quad \forall i \in \{1, \dots, A^* - 1\}$ if $A^* \geq 2$, and $\sigma_{A^*}^* = \sum_{j=1}^{A^*} \sigma_{jA^*}^{**}$ by construction of Σ^* .

Item 1 Follows directly from the construction of Σ^* .

Item 2 By construction of Σ^* , $\sum_{j=1}^{A^*-1} \sigma_{ji}^* = 0, i \in \{A^*, B^*\}$, if $A^* \geq 2$. If $|\mathbf{x}_{B^*}^*| = 2$, then $q_{B^*}^* > q_{B^*}^{**}$, and we can apply Item 2 of Claim 14 to get $\sum_{j=B^*}^I \sigma_{ji}^* = 0, i \in \{A^*, B^*\}$.

Item 3 From the proof of the previous item, we have $\sum_{j=A^*}^I \sigma_{ji}^* = 1, i \in \{A^*, B^*\}$. The result then trivially follows if $B^* = I$. Let $B^* \leq I - 1$, so that $\mathcal{C} = \{B^* + 1, \dots, I\}$. We can then apply Item 3 of Claim 14 to obtain $\sum_{j=B^*+1}^I \sigma_{jA^*}^* = 0$ and also $\sum_{j=B^*+1}^I \sigma_{jB^*}^* = 0$ if $B^* \leq B^* - 1$. We can finally apply Claim 15 to obtain $\sum_{j=B^*+1}^I \sigma_{jB^*}^* = 0$ if $B^* = B^*$.

Item 4 Observe that $\sigma_{A^*B^*}^* > 0$ only if $U_{B^*}(x_{B^*}^*) = U_{B^*}(x_{A^*}^*)$. But then $U_{A^*}(x_{A^*}^*) > U_{A^*}(x_{B^*}^*)$ by $q_{A^*}^* > q_{B^*}^{**}$. From Claim 2, we then obtain $U_{A^*}(x_{A^*}^*) > U_{A^*}(x_j^*)$, and therefore

$\sigma_{jA^*}^{**} = 0$, for all $j \in \{\underline{B}^*, \dots, I\}$. Hence, $1 = \sum_{j=1}^I \sigma_{jA^*}^{**} = \sum_{j=1}^{A^*} \sigma_{jA^*}^{**} = \sigma_{A^*}^*$.

Item 5 Assume that $\underline{B}^* \leq B^* - 2$. From Item 4 of Claim 14, we get $\sum_{j=1}^{A^*} \sigma_{ji}^* = 0$ for all $i \in \{\underline{B}^* + 1, \dots, B^* - 1\}$. $|\mathbf{x}_{\underline{B}^*}^*| = 2$ implies $q_{\underline{B}^*}^* > q_{B^*}^*$, and we can invoke Item 2 of Claim 14 to get $\sum_{j=B^*}^I \sigma_{ji}^* = 0$ for all $i \in \{\underline{B}^* + 1, \dots, B^* - 1\}$.

Item 6 Follows directly from Item 3 and Item 4 of Claim 14.

Item 7 If $\underline{B}^* \leq B^* - 1$, then Item 4 of Claim 14 implies $\sum_{j \in \mathcal{A}^*} \sigma_{jB^*}^* = 0$, whereas $\sum_{j \in \mathcal{C}^*} \sigma_{jB^*}^* = 0$ if $\mathcal{C}^* \neq \emptyset$ from Claim 15. If $\underline{B}^* = B^*$, then the result follows directly from Claim 16.

Item 8 Follows directly from $z^* = B^*$.

Item 9 Follows directly from the construction of Σ^* . ■

Claim 17 is close to Lemma 3, but still differs in multiple aspects. A major difference is that the lemma is specific about the randomization strategies an agent with marginal cost θ_i , $i \in \{A^*, \dots, B^*\}$, uses for cost reports θ_j , $j \in \underline{\mathcal{B}}^*$. We next establish incentive optimality of uniform randomization strategies.

Claim 18 *For any incentive optimal mechanism $(\mathbf{x}^*, \Sigma^{**} | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$), there exists an incentive optimal mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ in which the reporting strategy Σ^* has the following properties:*

1. $\sigma_{ji}^* = \frac{1 - \sigma_{A^*i}^{**}}{|\underline{\mathcal{B}}^*|} \forall (i, j) \in \{A^*, \dots, B^* - 1\} \times \underline{\mathcal{B}}^*$ if $|\mathbf{x}_{\underline{\mathcal{B}}^*}^*| = 2$,
2. $\sigma_{jB^*}^* = \frac{1 - \sigma_{B^*}^{**}}{|\underline{\mathcal{B}}^*|} \forall j \in \underline{\mathcal{B}}^*$ if $|\mathbf{x}_{\underline{\mathcal{B}}^*}^*| = 2$,
3. $\sigma_{ji}^* = \frac{1 - \sigma_{A^*i}^{**}}{|\underline{\mathcal{B}}^*|} \forall (i, j) \in \{A^*, \dots, B^*\} \times \underline{\mathcal{B}}^*$ if $|\mathbf{x}_{\underline{\mathcal{B}}^*}^*| = 1$.

Proof. Using the results in Claim 17, $q_{\underline{B}^*}^*$ solves

$$S'(q_{\underline{B}^*}^*) = \frac{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^{**} (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i))}{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^{**}} \quad \forall j \in \underline{\mathcal{B}}^*.$$

Multiply both sides of the equation by the denominator on the right-hand side, and sum up across all $j \in \underline{\mathcal{B}}^*$ to solve for

$$S'(q_{\underline{B}^*}^*) = \frac{\sum_{i=A^*}^{B^*-1} \nu_i (1 - \sigma_{A^*i}^{**}) (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) + \nu_{B^*} (1 - \sigma_{B^*}^{**}) \theta_{B^*}}{\sum_{i=A^*}^{B^*-1} \nu_i (1 - \sigma_{A^*i}^{**}) + \nu_{B^*} (1 - \sigma_{B^*}^{**})} \quad \text{if } |\mathbf{x}_{\underline{\mathcal{B}}^*}^*| = 2,$$

and

$$S'(q_{\underline{B}^*}^*) = \frac{\sum_{i=A^*}^{B^*} \nu_i (1 - \sigma_{A^*i}^{**}) (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i))}{\sum_{i=A^*}^{B^*} \nu_i (1 - \sigma_{A^*i}^{**})} \quad \text{if } |\mathbf{x}_{\underline{\mathcal{B}}^*}^*| = 1,$$

In the above expressions, $\sigma_{A^*i}^{**} = 0$ for all $i \geq \underline{B}^* + 1$. We can write the principals' expected surplus as

$$\begin{aligned} & \sum_{i=1}^{B^*} \nu_i W_i(x_i^*) + \nu_{A^*} (1 - \sigma_{A^*}^{**}) [W_{A^*}(x_{\underline{B}^*}^*) - W_{A^*}(x_{A^*}^*)] + \nu_{\underline{B}^*} \sigma_{A^* \underline{B}^*}^{**} \\ & \times [W_{\underline{B}^*}(x_{A^*}^*) - W_{\underline{B}^*}(x_{\underline{B}^*}^*)] + \nu_{B^*} (1 - \sigma_{B^*}^{**}) [W_{B^*}(x_{\underline{B}^*}^*) - W_{B^*}(x_{B^*}^*)] \end{aligned} \quad (35)$$

in the incentive optimal mechanism $(\mathbf{x}^*, \Sigma^{**} | \mathcal{A}^*, \mathcal{B}^*)$. The expected surplus depends on Σ^{**} only through $(\sigma_{A^*}^{**}, \sigma_{A^* \underline{B}^*}^{**}, \sigma_{B^*}^{**})$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $(\sigma_{A^*}^{**}, \sigma_{A^* \underline{B}^*}^{**})$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, recalling $x_{\underline{B}^*}^* = x_{B^*}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$. Let $(\sigma_{A^*}^*, \sigma_{A^* \underline{B}^*}^*, \sigma_{B^*}^*) = (\sigma_{A^*}^{**}, \sigma_{A^* \underline{B}^*}^{**}, \sigma_{B^*}^{**})$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $(\sigma_{A^*}^*, \sigma_{A^* \underline{B}^*}^*) = (\sigma_{A^*}^{**}, \sigma_{A^* \underline{B}^*}^{**})$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$. Then Σ^* only modifies cost reports θ_j , $j \in \underline{\mathcal{B}}^*$. Everything else is the same as in the original mechanism. Therefore, the mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ also yields expected surplus (35) if it is incentive feasible. To close the proof, we demonstrate sequential rationality of $x_{\underline{B}^*}^*$ in the modified mechanism.

By way of the uniform distribution of reporting strategies in Σ^* , the posterior beliefs regarding agent costs are identical for all $j \in \underline{\mathcal{B}}^*$. The maximal surplus the principal can achieve by offering a deviation contract $x_h^d = (q_h^d, \theta_h q_h^d)$ in Stage 4 that leaves an agent of cost type $h \in \{A^*, \dots, B^* - 1\}$ indifferent between accepting or rejecting the ex-post contract is proportional to

$$\Omega_h = \sum_{i=A^*}^h \nu_i (1 - \sigma_{A^*i}^{**}) [S(q_h^d) - (\theta_i + (1 - \alpha)(\theta_h - \theta_i)) q_h^d]$$

where

$$S'(q_h^d) = \frac{\sum_{i=A^*}^h \nu_i (1 - \sigma_{A^*i}^{**}) (\theta_i + (1 - \alpha)(\theta_h - \theta_i))}{\sum_{i=A^*}^h \nu_i (1 - \sigma_{A^*i}^{**})}.$$

If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then the principal's expected surplus of offering $x_{\underline{B}^*}^*$ at Stage 4 subsequent to a cost report θ_j , $j \in \underline{\mathcal{B}}^*$, is proportional to

$$\Omega_{B^*}^* = \sum_{i=A^*}^{B^*-1} \nu_i (1 - \sigma_{A^*i}^{**}) [S(q_{\underline{B}^*}^*) - (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) q_{\underline{B}^*}^*] + \nu_{B^*} (1 - \sigma_{B^*}^{**}) [S(q_{\underline{B}^*}^*) - \theta_{B^*} q_{\underline{B}^*}^*].$$

If $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, then the principal's expected surplus of offering $x_{\underline{B}^*}^*$ at Stage 4 subsequent to a cost report θ_j , $j \in \underline{\mathcal{B}}^*$, is proportional to

$$\Omega_{B^*}^* = \sum_{i=A^*}^{B^*-1} \nu_i (1 - \sigma_{A^*i}^{**}) [S(q_{\underline{B}^*}^*) - (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) q_{\underline{B}^*}^*] + \nu_{B^*} [S(q_{\underline{B}^*}^*) - \theta_{B^*} q_{\underline{B}^*}^*].$$

The mechanism $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ is incentive feasible if and only if the following sequential rationality constraint is met:

$$\Omega_{B^*}^* \geq \Omega_h \quad \forall h \in \{A^*, \dots, B^* - 1\}. \quad (36)$$

We now show that sequential rationality (32) of $x_{\underline{B}^*}^*$ in the original mechanism implies sequential rationality (36) of $x_{\underline{B}^*}^*$ in the modified mechanism. Summing up (32) over all $j \in \underline{\mathcal{B}}^*$ and using $\sum_{j \in \underline{\mathcal{B}}^*} \Omega_{jh}^* = \Omega_{B^*}^*$ returns the aggregate sequential rationality condition

$$\Omega_{B^*}^* \geq \sum_{j \in \underline{\mathcal{B}}^*} \Omega_{jh}(\sigma_{jh}^{**}) = \bar{\Omega}_h(\Sigma_h^{**}) \quad \forall h \in \{A^*, \dots, B^* - 1\}. \quad (37)$$

of the initial mechanism. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then Σ_h^{**} is a $(B^* - \underline{B}^*) \times (h + 1 - A^*)$ matrix that identifies how each of the cost types $i \in \{A^*, \dots, h\}$ randomizes across cost reports θ_j , $j \in \{\underline{B}^*, \dots, B^* - 1\}$. Instead, Σ_h^{**} has dimension $(B^* - A^*) \times (h + 1 - A^*)$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, because then the agent may optimally randomize across all cost types $j \in \mathcal{B}^*$. The final step is to show that $\bar{\Omega}_h(\Sigma_h^{**}) \geq \Omega_h$.

Consider the problem of minimizing $\bar{\Omega}_h(\Sigma_h)$ over Σ_h subject to $0 \leq \sigma_{ji} \leq 1$ for all $\sigma_{ji} \in \Sigma_h$, $\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji} = 1 - \sigma_{A^*i}^{**}$ for all $i \in \{A^*, \dots, h\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $\sum_{j=\underline{B}^*}^{B^*} \sigma_{ji} = 1 - \sigma_{A^*i}^{**}$ for all $i \in \{A^*, \dots, h\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$. By way of the envelope theorem we obtain:

$$\frac{\partial \Omega_{jh}}{\partial \sigma_{ji}} = \nu_i W_i(x_{jh}^d) = \nu_i [S(q_{jh}^d) - (\theta_i + (1 - \alpha)(\theta_h - \theta_i))q_{jh}^d],$$

with the cross-partial derivative of

$$\frac{\partial^2 \Omega_{jh}}{\partial \sigma_{ji} \partial \sigma_{jl}} = -\nu_i \nu_l \frac{[S'(q_{jh}^d) - \theta_i - (1 - \alpha)(\theta_h - \theta_i)][S'(q_{jh}^d) - \theta_l - (1 - \alpha)(\theta_h - \theta_l)]}{\sum_{i=A^*}^h \nu_i \sigma_{ji} S''(q_{jh}^d)}.$$

If we define

$$y_{jih} = \nu_i [S'(q_{jh}^d) - \theta_i - (1 - \alpha)(\theta_h - \theta_i)],$$

and $\mathbf{y}_{jh} = (y_{jA^*h}, \dots, y_{jh h})^T$, then we can write the Hessian matrix $\mathbf{H}_{jh}$ of $\Omega_{jh}(\sigma_{jh})$ as:

$$\mathbf{H}_{jh} = \frac{-\mathbf{y}_{jh} \mathbf{y}_{jh}^T}{\sum_{i=A^*}^h \nu_i \sigma_{ji} S''(q_{jh}^d)}.$$

By implication:

$$\sigma_{jh}^T \mathbf{H}_{jh} \sigma_{jh} = \frac{-\sigma_{jh}^T \mathbf{y}_{jh} \mathbf{y}_{jh}^T \sigma_{jh}}{\sum_{i=A^*}^h \nu_i \sigma_{ji} S''(q_{jh}^d)} = \frac{-(\sigma_{jh}^T \mathbf{y}_{jh})^2}{\sum_{i=A^*}^h \nu_i \sigma_{ji} S''(q_{jh}^d)} \geq 0.$$

Positive definiteness of $\mathbf{H}_{jh}$ implies that $\Omega_{jh}(\boldsymbol{\sigma}_{jh})$ is a convex function. Consequently, $\bar{\Omega}_h(\boldsymbol{\Sigma}_h)$ is convex because it is a sum of (additively separable) convex functions. Because all constraints are linear, all solutions to the $(B^* - \underline{B}^*) \times (h+1 - A^*) [(B^* - A^*) \times (h+1 - A^*)$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1]$ first-order conditions

$$\nu_i W_i(x_{jh}^d) - \lambda_{ih} - \xi_{jih} + \xi_{jih} = 0, \quad (38)$$

the $h+1 - A^*$ equality constraints

$$\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji} = 1 - \sigma_{A^*i}^{**} \text{ if } |\mathbf{x}_{\mathcal{B}^*}^*| = 2; \quad \sum_{j=\underline{B}^*}^{B^*} \sigma_{ji} = 1 - \sigma_{A^*i}^{**} \text{ if } |\mathbf{x}_{\mathcal{B}^*}^*| = 1, \quad (39)$$

and the $(B^* - \underline{B}^*) \times (h+1 - A^*) [(B^* - A^*) \times (h+1 - A^*)$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1]$ complementary slackness conditions

$$\sigma_{ji} \in [0, 1], \quad \xi_{jih} \geq 0, \quad \xi_{jih} \geq 0, \quad \sigma_{ji} \xi_{jih} = (1 - \sigma_{ji}) \xi_{jih} = 0 \quad (40)$$

minimize $\bar{\Omega}_h(\boldsymbol{\Sigma}_h)$. In the first-order condition (38), λ_{ih} represents the Lagrangian multiplier on the equality constraint (39), ξ_{jih} is the Karush-Kuhn-Tucker (KKT) multiplier on $\sigma_{jih} \geq 0$, and ξ_{jih} is the KKT multiplier on $\sigma_{ji} \leq 1$.

Obviously, $\boldsymbol{\Sigma}_h^*$, $\xi_{ji} = \xi_{jih} = 0$ and $\lambda_{ih} = \nu_i W_i(x_h^d)$ jointly solve (38)-(40). Hence, $\Omega_{B^*}^* \geq \bar{\Omega}_h(\boldsymbol{\Sigma}_h^{**}) \geq \bar{\Omega}_h(\boldsymbol{\Sigma}_h^*) = \Omega_h$ for all $h \in \{A^*, \dots, B^* - 1\}$. ■

To close the proof of Lemma 3, we need a final result.

Claim 19 *A mechanism $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) and is characterized either by (i) $\underline{B}^* = B^* - 1$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, or (ii) $\underline{B}^* \leq B^* - 2$, is incentive optimal only if $\sigma_{ji}^* = 0$ for all $(i, j) \in \mathcal{B}^* \times \mathcal{A}^*$.*

Proof. Suppose $\sigma_{ji}^* > 0$ for some $(i, j) \in \mathcal{B}^* \times \mathcal{A}^*$. We will show that there exists an incentive feasible mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ that also features incomplete commitment ($\mathcal{B} \neq \emptyset$) and yields the same expected surplus as the original mechanism, but $\mathcal{B} \subset \mathcal{B}^*$ and $\mathcal{B} \neq \mathcal{B}^*$. Consequently, the original mechanism is not minimal and therefore cannot be incentive optimal.

By way of Item 4 in Claim 14, we know that $\sigma_{ji}^* > 0$, $(i, j) \in \mathcal{B}^* \times \mathcal{A}^*$, implies $i = \underline{B}^*$. From Item 2 and Item 3 of Claim 17, we can set $\sigma_{j\underline{B}^*}^* = 0$ for all $j \in \{1, A^* - 1\}$ if $A^* \geq 2$. Moreover, $\sigma_{A^*\underline{B}^*}^* > 0$ implies $\sigma_i^* = 1$ for all $i \in \mathcal{A}^* \cup \mathcal{C}^*$ by Item 1, Item 4 and Item 9 of Claim 17. From Claim 18, we apply uniform randomization to derive the posterior

probability distribution

$$\begin{aligned}\mu_{j\underline{B}^*}^* &= \frac{\nu_{\underline{B}^*}(1 - \sigma_{A^*\underline{B}^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* - \nu_{B^*}\sigma_{B^*}^*} \\ \mu_{ji}^* &= \frac{\nu_i}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* - \nu_{B^*}\sigma_{B^*}^*} \forall i \in \{\underline{B}^* + 1, \dots, B^* - 1\} \\ \mu_{jB^*}^* &= \frac{\nu_{B^*}(1 - \sigma_{B^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* - \nu_{B^*}\sigma_{B^*}^*}\end{aligned}$$

for all $j \in \underline{\mathcal{B}}^*$, and $\mu_{B^*}^* = 1$, if $|\mathbf{x}_{\mathcal{B}^{**}}^*| = 2$. Instead,

$$\begin{aligned}\mu_{j\underline{B}^*}^* &= \frac{\nu_{\underline{B}^*}(1 - \sigma_{A^*\underline{B}^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^*} \\ \mu_{ji}^* &= \frac{\nu_i}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^*} \forall i \in \{\underline{B}^* + 1, \dots, B^*\}\end{aligned}$$

for all $j \in \underline{\mathcal{B}}^*$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$. The mechanism has expected surplus

$$\begin{aligned}W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) &= \sum_{i=1}^{B^*} \nu_i W_i(x_i^*) + \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* [W_{\underline{B}^*}(x_{A^*}^*) - W_{\underline{B}^*}(x_{\underline{B}^*}^*)] \\ &\quad + \nu_{B^*}(1 - \sigma_{B^*}^*) [W_{B^*}(x_{\underline{B}^*}^*) - W_{B^*}(x_{B^*}^*)].\end{aligned}$$

Consider the modified mechanism $(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$ in which $\mathcal{A}_k = \mathcal{A}_k^*$ for all $k \in \{1, \dots, K-1\}$ if $K \geq 2$, $\mathcal{A}_K$ is extended to include the cost type $\underline{B}^*$, and $\mathcal{B}$ is correspondingly reduced. That is, $A = \underline{B}^*$ and $B = \underline{B}^* + 1$. Let $x_j = x_j^*$ for all $j \neq \underline{B}^*$ and $x_{\underline{B}^*} = x_{A^*}^*$. As for reporting strategies, let $\sigma_i = \sigma_i^* = 1$ for all $i \in \mathcal{A}^* \cup \mathcal{C}^*$ and $\sigma_{\underline{B}^*} = \sigma_{A^*\underline{B}^*}^*$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then $\sigma_{j\underline{B}^*} = \frac{1-\sigma_{B^*}^*}{|\underline{\mathcal{B}}^*|-1}$, $\sigma_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|-1}$, $i \in \{\underline{B}^* + 1, \dots, B^* - 1\}$, and $\sigma_{jB^*}^* = \frac{1-\sigma_{B^*}^*}{|\underline{\mathcal{B}}^*|-1}$ for all $j \in \{\underline{B}^* + 1, \dots, B^* - 1\}$, whereas $\sigma_{B^*} = \sigma_{B^*}^*$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 1$, then $\sigma_{j\underline{B}^*} = \frac{1-\sigma_{B^*}^*}{|\underline{\mathcal{B}}^*|-1}$, $\sigma_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|-1}$, $i \in \{\underline{B}^* + 1, \dots, B^*\}$, for all $j \in \{\underline{B}^* + 1, \dots, B^*\}$. Observe in particular that $\sigma_{ij} = 0$ for all $(i, j) \in \mathcal{B} \times \mathcal{A}$ by construction. Moreover, $W(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*) = W(\mathbf{x}, \Sigma | \mathcal{A}, \mathcal{B})$. The uniform distribution of cost reports Σ yields posterior beliefs $\mu_{ji} = \mu_{ji}^*$ for all $(i, j) \in \mathcal{I} \times \{\underline{B}^* + 1, \dots, B^*\}$. Hence, the modified mechanism is incentive feasible by sequential rationality of $\mathbf{x}_{\mathcal{B}^*}^*$. ■

Claims 16-19 map into the items of Lemma 3 as follows. Item 1 of Claim 17 implies Item 1(a). Item 1 and Item 3 of Claim 18 imply Item 1(b). Claim 16 implies Item 2(a)i. Item 2 and Item 4 of Claim 17 imply Item 2(a)ii. Item 1 and Item 3 of Claim 18 and Claim 19 imply Item 2(a)iii. Item 1 of Claim 18 and Claim 19 imply Item 2(b). Item 8 of Claim 17 and Item 2 of Claim 18 imply Item 2(c)i. Item 7 of Claim 17 and Item 3 of Claim 18 imply Item 2(c)ii. Item 9 of Claim 17 implies Item 3. ■

4.3 Incentive optimality of an escape clause

Assume that $1 \leq K < I$.

Lemma 4 *Let $\hat{A}$ be the least efficient cost type that produces positive output $\hat{q}_{\hat{A}} > 0$ in the mechanism that maximizes the principal's expected surplus across all incentive feasible mechanisms with complete commitment (pure ex-ante contracting). The incentive optimal mechanism features incomplete commitment if $\hat{q}_{\hat{A}} > \hat{q}_{\hat{A}}^{fb}$.*

Proof. Denote by $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ the mechanism that maximizes the principal's expected surplus in the set of all incentive feasible mechanism with pure ex-ante contracting. Let $\hat{A}$ be the least efficient cost type that produces positive quantity $\hat{q}_{\hat{A}} > 0$ in this mechanism. Assume that $\hat{q}_{\hat{A}} > \hat{q}_{\hat{A}}^{fb}$. From Section 4.1, we know that $\hat{q}_{\hat{A}} = q_{\hat{A}}^{sb} \leq q_{\hat{A}}^{fb}$ if the least efficient cost group K only contains one element. Hence, $\hat{q}_{\hat{A}} > \hat{q}_{\hat{A}}^{fb}$ implies $|\hat{\mathcal{A}}_K| \geq 2$. Consider a modified mechanism $(\mathbf{x}, \mathbf{I}|\mathcal{A}, \mathcal{B})$ in which $\mathcal{A}_k = \hat{\mathcal{A}}_k$ for all $k \in \{1, \dots, K-1\}$ if $K \geq 2$, $\mathcal{A}_K = \hat{\mathcal{A}}_K \setminus \hat{A}$ and $\mathcal{B} = \hat{A}$. The least efficient cost type in the pure ex-ante mechanism has been moved into an escape clause designed for that specific type. Let $\mathbf{x}$ have the properties that $x_j = (\hat{q}_j, \hat{t}_j - (\theta_{\hat{A}} - \theta_{\hat{A}-1})(\hat{q}_{\hat{A}} - q_{\hat{A}}^{fb}))$ for all $j \in \mathcal{A}$, and assume that $x_{\hat{A}} = x_{\hat{A}}^{fb}$. We first verify incentive feasibility of the modified mechanism. Individual rationality (2) and incentive compatibility (3) follow from

$$U_i(x_i) - U_i(x_j) = U_i(\hat{x}_i) - U_i(\hat{x}_j) \geq 0 \quad \forall (i, j) \in \mathcal{A} \times \mathcal{A}$$

$$U_i(x_i) - U_i(x_{\hat{A}}^{fb}) = U_i(\hat{x}_i) - U_i(\hat{x}_{\hat{A}}) + (\theta_{\hat{A}-1} - \theta_i)(\hat{q}_{\hat{A}} - q_{\hat{A}}^{fb}) \geq 0 \quad \forall i \in \mathcal{A}$$

$$U_i(x_i) - U_i(x_j) = U_i(\hat{x}_i) - U_i(\hat{x}_j) + (\theta_{\hat{A}} - \theta_{\hat{A}-1})(\hat{q}_{\hat{A}} - q_{\hat{A}}^{fb}) > 0 \quad \forall (i, j) \in \{\hat{A}, \dots, I\} \times \mathcal{A}$$

$$U_i(x_i) - U_i(x_{\hat{A}}^{fb}) = (\theta_i - \theta_{\hat{A}})q_{\hat{A}}^{fb} > 0 \quad \forall i \in \{\hat{A} + 1, \dots, I\}, \quad \hat{A} \leq I - 1$$

$$U_i(x_{\hat{A}}^{fb}) = (\theta_{\hat{A}} - \theta_i)q_{\hat{A}}^{fb} \geq 0 \quad \forall i \in \hat{\mathcal{A}}$$

This mechanism trivially satisfies (4) because all types truthfully report cost with probability 1. $x_{\hat{A}}^{fb}$ is sequentially rational (5) because the only type that reports $\theta_{\hat{A}}$ is an agent with marginal cost $\theta_{\hat{A}}$. The mechanism satisfies the contracting constraint (7) by $|\mathbf{x}_{\mathcal{A}}| = |\hat{\mathbf{x}}_{\hat{\mathcal{A}}}|$. The expected surplus to the principal of the modified mechanism equals

$$W(\mathbf{x}, \mathbf{I}|\mathcal{A}, \mathcal{B}) = \sum_{i=1}^{\hat{A}-1} \nu_i W_i(\hat{x}_i) + \nu_{\hat{A}} w_{\hat{A}}^{fb} + G_{\hat{A}-1}(1 - \alpha)(\theta_{\hat{A}} - \theta_{\hat{A}-1})(\hat{q}_{\hat{A}} - q_{\hat{A}}^{fb}).$$

The difference

$$W(\mathbf{x}, \mathbf{I}|\mathcal{A}, \mathcal{B}) - W(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset) = \nu_{\hat{A}} [W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) - W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}})]$$

in expected surplus is strictly positive by strict concavity of $W_{\hat{A}}^{sb}(q)$ and $q_{\hat{A}}^{sb} < q_{\hat{A}}^{fb} < \hat{q}_{\hat{A}}$. As we have found an incentive feasible mechanism with incomplete commitment that strictly outperforms all incentive feasible mechanisms with complete commitment, the incentive optimal mechanism must feature incomplete commitment. ■

The proof of the following result is in Tangerås and Gick (2026):

Proposition 3 *The incentive optimal mechanism contains an escape clause if the principal attaches sufficient weight to efficiency relative to rent extraction (α is sufficiently close to 1).*

The next result applies to the general case of $K \geq 2$ and $\alpha \in (0, 1)$:

Proposition 4 *Assume that the mechanism that maximizes the principal's expected surplus across all incentive feasible mechanisms with complete commitment features partial participation. The incentive optimal mechanism then contains an escape clause if $\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$ is small and $W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > 0$.*

Proof. The chain of inequalities

$$0 < W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) - W_{\hat{A}+1}^{sb}(\hat{q}_{\hat{A}}) < (1 + \frac{G_{\hat{A}}}{v_{\hat{A}+1}}(1 - \alpha))(\theta_{\hat{A}+1} - \theta_{\hat{A}-1})\hat{q}_{\hat{A}}$$

implies $W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) \approx W_{\hat{A}+1}^{sb}(\hat{q}_{\hat{A}})$ for sufficiently small $\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$. Furthermore, $W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) \approx 0$ by (24) if $\hat{A} \leq I - 1$. If also $W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > 0$, then $W_{\hat{A}}^{sb}(q_{\hat{A}}^{sb}) > W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}})$. Strict concavity of $W_{\hat{A}}^{sb}(q)$, $q_{\hat{A}}^{fb} > q_{\hat{A}}^{sb}$ and $\hat{q}_{\hat{A}} > q_{\hat{A}}^{sb}$, see Lemma 10, then imply $\hat{q}_{\hat{A}} > q_{\hat{A}}^{fb}$. ■

4.4 Trade-offs in the design of an escape clause

This section in Tangerås and Gick (2026) discusses trade-offs in the determination of the escape clause, but does not contain any formal results.

5 Discussion

Tangerås and Gick (2026) discuss four model extensions. Two of those contain formal results. In Section 5.1, we provide a foundation for endogenously constrained contracting by introducing direct contracting costs. In Section 5.3, we establish a sufficient condition for when a mechanism in which the agent self-selects from a menu of ex-ante contract complemented by a general escape clause is close to incentive optimal.

5.1 Contracting costs as a source of incomplete commitment

Let K be endogenous, but assume that there is a fixed cost $c > 0$ of adding each additional contract to any given menu of contracts. This menu cost arises both for ex-ante and ex-post contracts. Tangerås and Gick (2026) prove the following result:

Proposition 5 *The incentive optimal mechanism features constrained contracting, $K \leq I - 1$, if $\nu_I w_I^{sb} < c$.*

All else equal, the principal optimally saves on contracting costs by reducing the number of ex-ante contracts below I if the likelihood ν_I of a very adverse marginal cost realization θ_I is sufficiently small.

Assume that the constrained full commitment mechanism $(\hat{\mathbf{x}}^K, \mathbf{I}|\hat{\mathcal{A}}^K, \emptyset)$ features incomplete participation, $\hat{A}_K \leq I - 1$. Modify it by an escape clause targeted to an agent with marginal production cost $\theta_{\hat{A}_K+1}$. Let the agent truthfully report its marginal cost in this modified mechanism. The first-best efficient contract $x_{\hat{A}_K+1}^{fb}$ then is sequentially rational following the marginal cost report $\theta_{\hat{A}_K+1}$. To maintain incentive compatibility, the principal must increase the transfer by $(\theta_{\hat{A}_K+1} - \theta_{\hat{A}_K})q_{\hat{A}_K+1}^{fb}$ for all marginal cost reports θ_j , $j \in \hat{\mathcal{A}}^K$. This mechanism is incentive feasible if and only if $\hat{q}_{\hat{A}_K} \geq q_{\hat{A}_K+1}^{fb}$, and delivers expected surplus

$$W(\hat{\mathbf{x}}^K, \mathbf{I}|\hat{\mathcal{A}}^K, \emptyset) - cK + \nu_{\hat{A}_K+1}[W_{\hat{A}_K+1}^{sb}(q_{\hat{A}_K+1}^{fb}) - c].$$

Proposition 6 *Assume that the mechanism $(\hat{\mathbf{x}}^K, \mathbf{I}|\hat{\mathcal{A}}^K, \emptyset)$ that maximizes the principal's expected surplus across all incentive feasible mechanisms with complete commitment features constrained contracting and partial participation, $\hat{A}_K \leq I - 1$. The incentive optimal mechanism contains an escape clause if $\theta_{\hat{A}_K+1} - \theta_{\hat{A}_K-1}$ is small and $W_{\hat{A}_K+1}^{sb}(q_{\hat{A}_K+1}^{fb}) > c$.*

Proof. To complete the proof we need to verify incentive feasibility, $\hat{q}_{\hat{A}_K} \geq q_{\hat{A}_K+1}^{fb}$. By an argument similar to the one in the proof of Proposition 4, the following approximations and chain of inequalities

$$w_{\hat{A}_K}^{sb} > W_{\hat{A}_K}^{sb}(q_{\hat{A}_K}^{fb}) \approx W_{\hat{A}_K}^{sb}(q_{\hat{A}_K+1}^{fb}) > W_{\hat{A}_K+1}^{sb}(q_{\hat{A}_K+1}^{fb}) > c > 0 \approx W_{\hat{A}_K}^{sb}(\hat{q}_{\hat{A}_K})$$

hold if $\theta_{\hat{A}_K+1} - \theta_{\hat{A}_K-1}$ is small, $\hat{A}_K \leq I - 1$, and $W_{\hat{A}_K+1}^{sb}(q_{\hat{A}_K+1}^{fb}) > c$. Strict concavity of the second-best welfare function, $q_{\hat{A}_K}^{fb} > q_{\hat{A}_K}^{sb}$ and $\hat{q}_{\hat{A}_K} > q_{\hat{A}_K}^{sb}$, and monotonicity then imply $\hat{q}_{\hat{A}_K} > q_{\hat{A}_K}^{fb} > q_{\hat{A}_K+1}^{fb}$. ■

5.2 Contract complexity

This section in Tangerås and Gick (2026) interprets K in terms of contract complexity, but does not contain any formal statements.

5.3 Implementation through a general escape clause

Let $(\mathbf{x}^*, \Sigma^* | \mathcal{A}^*, \mathcal{B}^*)$ be an incentive optimal mechanism, and assume that $\mathcal{B}^* \neq \emptyset$. To gauge the value of communication between the principal and the agent, consider the following mechanism and sequence of events:

Stage 1: The principal commits to a GC mechanism consisting of a menu $\mathbf{x}^{GC}$ of K contracts $x_{A_k^*}^{GC} = (q_{A_k^*}^{GC}, t_{A_k^*}^{GC})$, where

$$q_{A_k^*}^{GC} = q_{A_k^*}^*, \quad t_{A_k^*}^{GC} = t_{A_k^*}^* - (\theta_{B^*} - \theta_{A^*})(q_{B^*}^* - q_{B^*}^{GC}) \quad \forall k \in \{1, \dots, K\},$$

augmented by the general escape clause (GC):

The agent has the right to terminate the initial agreement and request a new contract offer from the principal regardless of the circumstances.

Stage 2: The agent

- selects $x_{A_k^*}^{GC}$ if $\theta_i \in \mathcal{A}_k^*$ and $\theta_i < \theta_{A^*}$.
- selects $x_{A^*}^{GC} = x_{A_K^*}^{GC}$ with probability $\sigma_{A^*}^*$ and activates the escape clause GC with probability $1 - \sigma_{A^*}^*$ if the agent has marginal cost θ_{A^*} .
- activates the escape clause GC if the marginal cost is θ_i , $i \in \mathcal{B}^*$.
- rejects the offer if $\mathcal{C}^* \neq \emptyset$ and the marginal cost is θ_i , $i \in \mathcal{C}^*$.

Stage 3: If the agent has invoked the escape clause in stage 2, then the principal offers the ex-post contract $x_{B^*}^{GC}$ featuring transfer payment $t_{B^*}^{GC} = \theta_{B^*} q_{B^*}^{GC}$ and output $q_{B^*}^{GC}$ solved by

$$S'(q_{B^*}^{GC}) = \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)(\theta_{A^*} + (1 - \alpha)(\theta_{B^*} - \theta_{A^*})) + \sum_{i \in \mathcal{B}^*} \nu_i(\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i))}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \mathcal{B}^*} \nu_i}. \quad (41)$$

The contract offered by the principal in Stage 3 builds on the posterior belief distribution

$$\mu_{A^*}^{GC} = \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \mathcal{B}^*} \nu_i}, \quad \mu_i^{GC} = \frac{\nu_i}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \mathcal{B}^*} \nu_i} \quad \forall i \in \mathcal{B}^* \quad (42)$$

about the agent's marginal cost subsequent to the activation of the escape clause.

In the incentive optimal mechanism, the posterior belief distribution equals

$$\begin{aligned}\mu_{A^*}^* &= \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \\ \mu_i^* &= \frac{\nu_i}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \quad \forall i \in \underline{\mathcal{B}}^* \\ \mu_{B^*}^* &= \frac{\nu_{B^*}(1 - \sigma_{B^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)},\end{aligned}\tag{43}$$

for any marginal cost report θ_j , $j \in \underline{\mathcal{B}}^*$. It differs only by the factor $\nu_{B^*}\sigma_{B^*}^*$ compared to the GC belief distribution. The corresponding output requirement $q_{\underline{B}^*}^* > q_{\underline{B}^*}^{GC}$ solves

$$\begin{aligned}S'(q_{\underline{B}^*}^*) &= \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)(\theta_{A^*} + (1 - \alpha)(\theta_{B^*} - \theta_{A^*}))}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \\ &\quad + \frac{\sum_{i \in \underline{\mathcal{B}}^*} \nu_i(\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) + \nu_{B^*}(1 - \sigma_{B^*}^*)\theta_{B^*}}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)}\end{aligned}\tag{44}$$

for all marginal cost reports θ_j , $j \in \underline{\mathcal{B}}^*$.

Proposition 7 *Assume that the incentive optimal mechanism features incomplete commitment. Then there exists a restricted communication mechanism augmented by a general escape clause that can be sustained as a PBE. This mechanism generates in the limit $\nu_{B^*}\sigma_{B^*}^* \rightarrow 0$ the same expected surplus as the incentive optimal mechanism.*

Proof. Consider first the simpler case with $|\mathbf{x}_{B^*}^*| = 1$. Here, ex-post contracting only ever results in one single contract proposal $x_{\underline{B}^*}^*$ regardless of the specific cost report θ_j , $j \in \mathcal{B}^*$. Incentive optimality of uniform randomization established in Lemma 3 then yields exactly the same distribution of posterior beliefs over the agent's marginal cost as in (42) for any cost report θ_j , $j \in \mathcal{B}^*$. Hence, the principal does not derive any additional information from the specific cost report θ_j than it can infer from the activation of the clause itself. The sequence of events specified above can therefore be sustained as a PBE. The restricted communication mechanism augmented by a GC clause generates precisely the same expected welfare as the incentive optimal mechanism.

Consider next the more complicated case with $|\mathbf{x}_{B^*}^*| = 2$. We start the proof by constructing a particular mechanism $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ and demonstrating its incentive feasibility. We then show that this mechanism can be implemented as a PBE of a restricted communication mechanism supported by a GC. We finally demonstrate that $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ converges in expected surplus to the incentive optimal mechanism $(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*)$ as $\nu_{B^*}\sigma_{B^*}^* \rightarrow 0$.

Subcase 1: An agent with marginal cost $\theta_{\underline{B}^*}$ reports marginal cost θ_{A^*} with zero probabil-

ity, $\sigma_{A^*B^*}^* = 0$. Let $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ be described by the ex-ante menu of contracts $x_i = x_{A_k}^{GC}$ for all $i \in \mathcal{A}_k^*$, $k \in \{1, \dots, K\}$. As for the reporting strategies, $\sigma_i = 1$ for all $i \in \{1, \dots, A^* - 1\}$ if $|\mathcal{A}^*| \geq 2$, $\sigma_{A^*} = \sigma_{A^*}^*$ and $\sigma_{jA^*} = \frac{1 - \sigma_{A^*}^*}{|\mathcal{B}^*|}$ for all $j \in \mathcal{B}^*$, $\sigma_{ji} = \frac{1}{|\mathcal{B}^*|}$ for all $(i, j) \in \mathcal{B}^* \times \mathcal{B}^*$, and $\sigma_i = 1$ for all $i \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$. By subtracting (44) from (41), we get

$$S'(q_{B^*}^{GC}) - S'(q_{B^*}^*) = \alpha \nu_{B^*} \frac{\sigma_{B^*}^* [\nu_{A^*}(1 - \sigma_{A^*}^*)(\theta_{B^*} - \theta_{A^*}) + \sum_{i \in \mathcal{B}^*} \nu_i(\theta_{B^*} - \theta_i)]}{[\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*}][\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*} - \nu_{B^*}\sigma_{B^*}^*]} > 0.$$

Hence, $q_{B^*}^* > q_{B^*}^{GC} = q_{B^*}$ by strict concavity of $S(q)$. By construction, $U_i(x_i) - U_i(x_j) = U_i(x_i^*) - U_i(x_j^*)$ for all $(i, j) \in \mathcal{A}^* \times \mathcal{A}^*$, $U_{A^*}(x_{A^*}) - U_{A^*}(x_{B^*}) = U_{A^*}(x_{A^*}^*) - U_{A^*}(x_{B^*}^*)$, and $q_{A^*}^* > q_{B^*}$, so we can apply Item 2 of Claim 2 and Item 1 of Claim 3 to verify individual rationality (2) and incentive compatibility (3) of $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$. The mechanism obviously satisfies the stochastic rationality constraint (4). The posterior beliefs (6) of the modified mechanism are calculated as $\mu_j = 1$ for all $j \in \mathcal{A}^* \cup \mathcal{C}^*$ and $\mu_{ji} = \mu_i^{GC}$ for all $j \in \mathcal{B}^*$ and $i \in A^* \cup \mathcal{B}^*$, where $\mu_{A^*}^{GC}$ and μ_i^{GC} for all $i \in \mathcal{B}^*$ were described in (42). Moreover, $|\mathbf{x}| = |\mathbf{x}_{\mathcal{A}^*}^*|$ implies (7). We finally need to verify sequential rationality (5) of x_{B^*} to establish incentive feasibility of $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$. To do so, we first define the function

$$\begin{aligned} \tilde{\Omega}(\sigma) &= \nu_{A^*}(1 - \sigma_{A^*}^*)[S(\tilde{q}(\sigma)) - (\theta_{A^*} + (1 - \alpha)(\theta_{B^*} - \theta_{A^*}))\tilde{q}(\sigma)] \\ &\quad + \sum_{i \in \mathcal{B}^*} \nu_i[S(\tilde{q}(\sigma)) - (\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i))\tilde{q}(\sigma)] - \nu_{B^*}\sigma[S(\tilde{q}(\sigma)) - \theta_{B^*}\tilde{q}(\sigma)], \end{aligned}$$

where $\tilde{q}(\sigma)$ is implicitly defined by

$$S'(\tilde{q}(\sigma)) = \frac{\sum_{i=A^*}^{B^*} \nu_i(\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) - \nu_{A^*}\sigma_{A^*}^*(\theta_{A^*} + (1 - \alpha)(\theta_{B^*} - \theta_{A^*})) - \nu_{B^*}\sigma\theta_{B^*}}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*} - \nu_{B^*}\sigma}.$$

By this construction, $\tilde{q}(\sigma_{B^*}^*) = q_{B^*}^*$, $\tilde{\Omega}(\sigma_{B^*}^*) = \Omega_{B^*}^*$, see the proof of Claim 18 for a definition, and $\tilde{q}(0) = q_{B^*}$. The expected surplus to the principal of offering the ex-post contract x_{B^*} after the agent has reported marginal cost θ_j , $j \in \mathcal{B}^*$, equals $\frac{\tilde{\Omega}(0)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*}}$, whereas the deviation profit of offering a deviation contract $x_h^d = (q_h^d, \theta_h q_h^d)$, $h \in \{A^*, \dots, B^* - 1\}$, equals $\frac{\Omega_h}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*}}$, see the proof of Claim 18 for the definition of Ω_h and discussion. Hence, x_{B^*} is sequentially rational if and only if $\tilde{\Omega}(0) \geq \Omega_h$ for all $h \in \{A^*, \dots, B^* - 1\}$. Sequential rationality of $x_{B^*}^*$ in the incentive optimal mechanism implies $\tilde{\Omega}(\sigma_{B^*}^*) \geq \Omega_h$ for all $h \in \{A^*, \dots, B^* - 1\}$. We close the proof of incentive feasibility by demonstrating $\tilde{\Omega}(0) > \tilde{\Omega}(\sigma_{B^*}^*)$. As

$$\tilde{\Omega}'(\sigma) = -\nu_{B^*}(S(\tilde{q}(\sigma)) - \theta_{B^*}\tilde{q}(\sigma)), \quad \tilde{\Omega}''(\sigma) = \frac{-1}{S''(\tilde{q}(\sigma))} \frac{\nu_{B^*}^2 (S'(\tilde{q}(\sigma)) - \theta_{B^*})^2}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \nu_{B^*} - \nu_{B^*}\sigma} > 0,$$

we have $\tilde{\Omega}'(\sigma) < \tilde{\Omega}'(\sigma_{B^*}^*) = -\nu_{B^*}W_{B^*}(x_{B^*}^*) \leq 0$ for all $\sigma < \sigma_{B^*}^*$, where we demonstrated $W_{B^*}(x_{B^*}^*) \geq 0$ in the proof of Claim 15. Hence, $\tilde{\Omega}(0) > \tilde{\Omega}(\sigma_{B^*}^*)$.

By incentive feasibility of $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ and $|\mathbf{x}_{B^*}| = 1$, the reduced communication mechanism augmented by the GC can be sustained as a PBE. This mechanism generates the same expected surplus to the principal as $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$, namely:

$$\begin{aligned} W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*) &= \sum_{i \in \mathcal{A}^*} \nu_i W_i(x_i^*) + \nu_{A^*}(1 - \sigma_{A^*}^*)[W_{A^*}^{fb}(q_{B^*}^*) - W_{A^*}^{fb}(q_{B^*}^*)] \\ &\quad + \sum_{i \in \mathcal{B}^*} \nu_i W_i(x_{B^*}^*) + \nu_{A^*}(1 - \alpha)(\theta_{B^*} - \theta_{A^*})(q_{B^*}^* - q_{B^*}^*), \end{aligned}$$

where $\nu_{A^*} = \sum_{i \in \mathcal{A}^*} \nu_i$. The net benefit of the incentive optimal mechanism equals

$$\begin{aligned} W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*) &= \nu_{A^*}(1 - \sigma_{A^*}^*)[W_{A^*}^{fb}(q_{B^*}^*) - W_{A^*}^{fb}(q_{B^*}^*)] + \sum_{i \in \mathcal{B}^*} \nu_i [W_i(x_{B^*}^*) - W_i(x_{B^*}^*)] \\ &\quad + \nu_{B^*}\sigma_{B^*}^*[W_{B^*}^{fb}(q_{B^*}^*) - W_{B^*}^{fb}(q_{B^*}^*)] - \nu_{A^*}(1 - \alpha)(\theta_{B^*} - \theta_{A^*})(q_{B^*}^* - q_{B^*}^*) \end{aligned}$$

$\lim_{\nu_{B^*}\sigma_{B^*} \rightarrow 0} [S'(q_{B^*}^*) - S'(q_{B^*}^*)] = 0$ implies $q_{B^*}^* \rightarrow q_{B^*}^*$ and $x_{B^*}^* \rightarrow x_{B^*}^*$ as $\nu_{B^*}\sigma_{B^*} \rightarrow 0$. Therefore, $\lim_{\nu_{B^*}\sigma_{B^*} \rightarrow 0} [W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)] = 0$.

Subcase 2: An agent with marginal cost θ_{B^*} reports marginal cost θ_{A^*} with positive probability, $\sigma_{A^*B^*}^* > 0$. This occurs only if $|\mathcal{B}^*| = 2$ and $|x_{B^*}^*| = 2$; see Lemma 3. As $\sigma_{A^*B^*}^* = 1 - \sigma_{B^*}^*$, the incentive optimal ex-post contract $x_{B^*}^* = (q_{B^*}^*, \theta_{B^*}q_{B^*}^*)$ has output

$$S'(q_{B^*}^*) = \frac{\nu_{B^*}\sigma_{B^*}^*(\theta_{B^*} + (1 - \alpha)(\theta_{B^*} - \theta_{B^*})) + \nu_{B^*}(1 - \sigma_{B^*}^*)\theta_{B^*}}{\nu_{B^*}\sigma_{B^*}^* + \nu_{B^*}(1 - \sigma_{B^*}^*)}.$$

Consider the modified mechanism $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ in which the menu $\mathbf{x}$ of contracts has the following properties: $x_j = (q_j^*, t_j^* - (\theta_{B^*} - \theta_{B^*})(q_{B^*}^* - q_{B^*}^*))$ for all $j \in \mathcal{A}^*$ and $x_{B^*} = x_{B^*} = (q_{B^*}^*, \theta_{B^*}q_{B^*}^*)$ where

$$S'(q_{B^*}^*) = \frac{\nu_{B^*}\sigma_{B^*}^*(\theta_{B^*} + (1 - \alpha)(\theta_{B^*} - \theta_{B^*})) + \nu_{B^*}\theta_{B^*}}{\nu_{B^*}\sigma_{B^*}^* + \nu_{B^*}}.$$

The reporting strategies are $\sigma_i = \sigma_i^* = 1$ for all $i \in \mathcal{A}^* \cup \mathcal{C}^*$. Moreover, $\sigma_{A^*B^*}^* = 1 - \sigma_{B^*}^*$, $\sigma_{B^*} = \sigma_{B^*B^*} = \sigma_{B^*}^*/2$ and $\sigma_{B^*B^*} = \sigma_{B^*} = 1/2$. In particular,

$$S'(q_{B^*}^*) - S'(q_{B^*}^*) = \frac{\alpha\nu_{B^*}\sigma_{B^*}^*\nu_{B^*}\sigma_{B^*}^*(\theta_{B^*} - \theta_{B^*})}{[\nu_{B^*}\sigma_{B^*}^*][\nu_{B^*}\sigma_{B^*}^* + \nu_{B^*}(1 - \sigma_{B^*}^*)]} > 0$$

implies $q_{B^*}^* > q_{B^*}^*$. It is straightforward to verify that the construction of the transfer

payments in $\mathbf{x}$, $U_i(x_i) - U_i(x_j) = U_i(x_i^*) - U_i(x_j^*)$ for all $(i, j) \in \mathcal{A}^* \times \mathcal{A}^*$, $U_{\underline{B}^*}(x_{\underline{B}^*}) = U_{\underline{B}^*}(x_{\mathcal{A}^*})$ and $q_{\mathcal{A}^*}^* > q_{\underline{B}^*}$ imply that the modified mechanism satisfies feasibility conditions (2)-(4) and (7). We calculate the posterior beliefs (6) of the modified mechanism as:

$$\mu_{\underline{B}^*} = \frac{\nu_{\underline{B}^*} \sigma_{\underline{B}^*}^*}{\nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* + \nu_{B^*}}, \quad \mu_{B^*} = \frac{\nu_{B^*}}{\nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* + \nu_{B^*}} \quad j \in \{\underline{B}^*, B^*\}.$$

We finally verify sequential rationality (5) of $x_{\underline{B}^*}$. In the incentive optimal mechanism, $x_{\underline{B}^*}^*$ is sequentially rational if and only if

$$\Omega_{B^*}^* = \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* [S(q_{\underline{B}^*}^*) - (\theta_{\underline{B}^*} + (1-\alpha)(\theta_{B^*} - \theta_{\underline{B}^*})) q_{\underline{B}^*}^*] + \nu_{B^*} (1 - \sigma_{B^*}^*) [S(q_{\underline{B}^*}^*) - \theta_{B^*} q_{\underline{B}^*}^*] \geq \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* w_{\underline{B}^*}^{fb}.$$

The right-hand side of this expression is the expected surplus of offering a deviation contract that is accepted only by an agent with marginal cost $\theta_{\underline{B}^*}$. The modified contract $x_{\underline{B}^*}$ is sequentially rational if and only if

$$\Omega_{B^*} = \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* [S(q_{\underline{B}^*}) - (\theta_{\underline{B}^*} + (1-\alpha)(\theta_{B^*} - \theta_{\underline{B}^*})) q_{\underline{B}^*}] + \nu_{B^*} [S(q_{\underline{B}^*}) - \theta_{B^*} q_{\underline{B}^*}] \geq \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* w_{\underline{B}^*}^{fb}.$$

One can then construct a similar function to $\tilde{\Omega}(\sigma)$ above to verify $\Omega_{B^*} > \Omega_{B^*}^*$.

By incentive feasibility of $(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)$ and $|\mathbf{x}_{B^*}| = 1$, the reduced communication mechanism augmented by the GC can be sustained as a PBE. This mechanism generates the expected surplus:

$$\begin{aligned} W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*) &= \sum_{i \in \mathcal{A}^*} \nu_i W_i(x_i^*) + \nu_{\underline{B}^*} (1 - \sigma_{\underline{B}^*}^*) W_{\underline{B}^*}(x_{\mathcal{A}^*}^*) + \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* W_{\underline{B}^*}(x_{\underline{B}^*}) + \nu_{B^*} W_{B^*}(x_{\underline{B}^*}) \\ &\quad + (\nu_{\mathcal{A}^*} + \nu_{\underline{B}^*} (1 - \sigma_{\underline{B}^*}^*)) (1 - \alpha) (\theta_{B^*} - \theta_{\underline{B}^*}) (q_{\underline{B}^*}^* - q_{\underline{B}^*}). \end{aligned}$$

The net benefit of choosing the incentive optimal over the reduced communication mechanism can be written as

$$\begin{aligned} W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*) &= \nu_{\underline{B}^*} \sigma_{\underline{B}^*}^* [W_{\underline{B}^*}^{fb}(q_{\underline{B}^*}^*) - W_{\underline{B}^*}^{fb}(q_{\underline{B}^*})] + \nu_{B^*} [W_{B^*}^{fb}(q_{\underline{B}^*}^*) - W_{B^*}^{fb}(q_{\underline{B}^*})] \\ &\quad + \nu_{B^*} \sigma_{B^*}^* [w_{B^*}^{fb} - W_{B^*}^{fb}(q_{\underline{B}^*}^*)] - (\nu_{\mathcal{A}^*} + \nu_{\underline{B}^*} (1 - \alpha)) (\theta_{B^*} - \theta_{\underline{B}^*}) (q_{\underline{B}^*}^* - q_{\underline{B}^*}) \end{aligned}$$

$$\lim_{\nu_{B^*} \sigma_{B^*}^* \rightarrow 0} [W(\mathbf{x}^*, \Sigma^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}, \Sigma|\mathcal{A}^*, \mathcal{B}^*)] = 0 \text{ since } q_{\underline{B}^*}^* \rightarrow q_{\underline{B}^*} \text{ for } \nu_{B^*} \sigma_{B^*}^* \rightarrow 0. \blacksquare$$

5.4 Other clauses

This section in Tangerås and Gick (2026) discusses a renegotiation clause relative to an escape clause, but does not contain any formal statements.

References

- BESTER, H. AND R. STRAUZ (2001): “Contracting with imperfect commitment and the revelation principle: The single agent case,” Econometrica, 69, 1077–1098.
- TANGERÅS, T. AND W. H. GICK (2026): “Contracting with endogenously incomplete commitment: Escape clauses,” <http://www.ifn.se/thomast>.