Contracting with endogenously incomplete commitment: Escape clauses*

Thomas P. Tangerås† and Wolfgang H. Gick‡

December 19, 2023

Abstract

We establish circumstances when a principal benefits from limiting contract commitment in the mechanism proposed to the agent. Such situations occur under constrained contracting where the maximum number of admissible ex-ante contract offers is below the number of potential agent types. Enabling the agent to cancel a contract under predefined circumstances in return for a future offer by the principal, improves contract fit at the cost of dynamic inefficiency. We study consequences of including such an escape clause in an agreement and identify trade-offs involved in its design. The paper so develops a theory of contracting with endogenously incomplete commitment.

JEL classification: *D82, D84, D86.*
Keywords: *Constrained contracting, endogenously incomplete commitment, escape clauses, ratchet effect, stochastic revelation principle.*

.

---

†Corresponding author. Research Institute of Industrial Economics (IFN) P.O. Box 55665, 10215 Stockholm, Sweden, e-mail: thomas.tangeras@ifn.se

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

1   Introduction

A common assumption of contract theory is that the agent has private information about the cost of producing output demanded by the principal. The typical solution to the problem of contracting under asymmetric information contains a menu of contracts specified by the principal upfront. This ex-ante menu is constructed to be incentive compatible so that the agent receives the intended contract by revealing its cost. Under a standard assumption on the probability distribution of costs, the menu specifies one unique contract for each of the agent's possible cost realizations, or types.

In reality, agents seldom receive such an extensive menu of contracts to select from as prescribed by theory. For instance, agreements between regulatory authorities and regulated entities usually are one-size-fits-all. We explore consequences for mechanism design of an assumption that the maximal number of different contracts the principal can include in the ex-ante menu is smaller than the number of the agent's potential cost types. The main insight emerging from this analysis is that the principal in such circumstances may benefit from reserving the possibility to contract after the agent reports its cost. Allowing such ex-post contracting adds valuable adaptability to an environment where contracting is constrained ex ante. The mechanism features *incomplete commitment* in the sense that instances arise with positive probability when the agent first communicates with the principal and then receives a contract offer. Incompleteness is *endogenous* because situations when ex-post contracting may occur are described already in the mechanism presented by the principal to the agent.

We model incomplete commitment as an *escape clause*. This is "a statement in a contract that allows you to break all or part of the contract under particular conditions."[1] In our context, an escape clause is defined in terms of a subset of cost reports that nullify the ex-ante menu of contracts and entitle the agent to a new contract offer by the principal. An example of precisely such a stipulation is found in the regulation of water companies in England and Wales. This regulation permits companies to request an interim determination if their operating conditions change substantially between price reviews. The regulator may adjust the price limit if, for instance, the changes in costs amount to at least 10 percent of the company's turnover.[2] An escape clause is contained in the Swedish Electricity Act establishing the general framework for regulating electricity distribution networks in Sweden. According to this act, the regulatory authority may change the revenue cap during the regulatory period by request of the regulated firm if 1) circumstances warrant a substantial increase in the revenue cap; or 2) other valid reasons

---

[1]This definition is from the Cambridge Dictionary; see dictionary.cambridge.org/dictionary/english/escape-clause

[2]See ofwat.gov.uk/regulated-companies/price-review/interim-determinations/.

apply.[3] The regulatory framework does not place any limitations on this new contract other than it being a revenue cap. Regulation of electricity distribution networks in the United Kingdom allows adjustment of network tariffs to account for uncertainty. A re-opener mechanism decides "within a price control period, on additional allowances to deliver a project or activity once there is more certainty" (Ofgem, 2022, Section 6).[4]

Escape clauses occur also in real estate and venture capital agreements. A related stipulation is a break clause, typically featured in tenancy agreements, by which a party can end a contract prematurely. Common law in most countries specify rights to terminate a contract. Termination clauses apply if one of the contracting parties cannot fulfil its contractual obligations. However, contracts sometimes allow so-called termination for convenience which "gives the parties the right to terminate the contract without any reason or penalty".[5] These types of clauses usually do not require the principal to make a subsequent contract offer. But the principal generally cannot lose from proposing a new contract after a previous agreement has ended, so activation of these clauses likely leads to ex-post contracting.

Escape clauses are commonly found in trade agreements (Bagwell and Staiger, 2005; Beshkar and Bond, 2017) and in fiscal policy frameworks (Halac and Yared, 2014). In those contexts, escape clauses are safeguards that allow temporary deviations from an agreement in particular circumstances. However, our paper analyzes escape clauses that permanently terminate an initial agreement.

To understand why incomplete commitment can be valuable under asymmetric information, recall that the principal benefits from decreasing the contracted output of an agent with high marginal costs. Doing so reduces the informational rent an agent can earn by exaggerating its cost, thereby enabling the principal to diminish contracted transfer payments. From this second-best perspective, production of an agent with high marginal cost is excessive both in a constrained ex-ante environment and under ex-post contracting. In the first case, the distortion arises because contracts bundle high-cost with lower-cost types. In the second case, output is too high because the agent's informational rent is sunk when the principal makes its ex-post contract offer. Output can be more distorted under a constrained than an ex-post contract. The principal then increases expected surplus by entering the possibility for ex-post contracting into the mechanism. The following example illustrates this point.

Assume that the principal pays an agent a transfer t to supply output in quantity q

---

[3]Ellag (1997:857), 5 kap. 20§, our translation.

[4]The regulation also contains specified uncertainty mechanisms such as automatic indexation and pass-through of costs. Those types of corrections are sometimes referred to as Z-factor adjustments (Sappington and Weisman, 2010).

[5]See summize.com/resources/contract-termination.

for which the principal has positive, but diminishing marginal valuation $S'(q)$. The agent produces output at constant marginal cost which is either low, $\theta_1 > 0$, or high, $\theta_2 > \theta_1$. The principal cannot observe this marginal cost, but knows that it is low with probability $\nu > 0$ and high with probability $1 - \nu > 0$. The principal would like to minimize the transfer t for any quantity q produced by the agent, but the agent accepts the contract proposal only if the transfer is sufficient to cover its production cost.
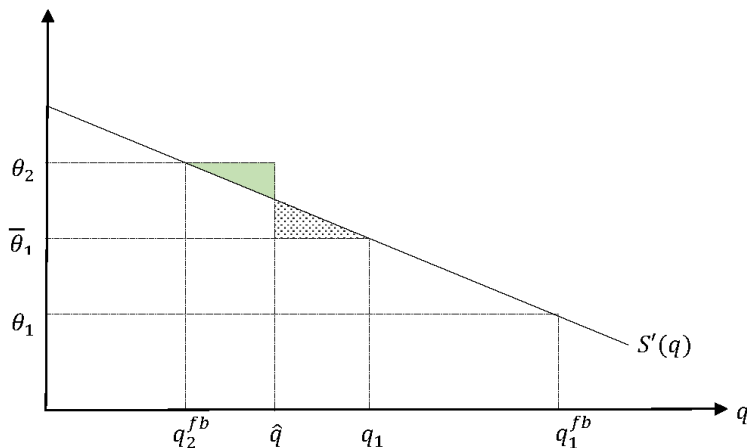

Figure 1: Efficiency benefits of ex-post contracting

Under complete information, the principal would instruct the agent to produce output at the point where the marginal benefit of the principal was equal to the marginal cost of the agent. This first-best quantity is identified by q_1^{fb} in Figure 1 for an agent with low marginal cost and by q_2^{fb} if the agent produces at high marginal cost. Under incomplete information, the principal offers a menu of two contracts. The agent produces the first-best output q_1^{fb} under the low-cost contract. The second-best output q_2^{sb} (not identified in the figure) under the high-cost contract is below q_2^{fb}. The informational rent paid to an agent with low marginal cost to maintain incentive compatibility increases the virtual marginal production cost above θ_2 for an agent with high marginal cost.

The assumption explored in this paper is that of constrained contracting. Its implication for the above example is that the principal cannot offer two contracts upfront, only one. Let the probability $1 - \nu$ of the agent having a high marginal cost be so large that the principal wants the agent to produce regardless of its marginal cost. The principal's most-preferred output $\hat{q}$ under constrained ex-ante contracting identified in Figure 1 balances the expected marginal distortion of reducing output below q_1^{fb} and increasing it above q_2^{sb}. This output satisfies $\hat{q} > q_2^{fb}$ because the agent's expected virtual marginal production cost under constrained contracting is strictly below θ_2 by the probability that the agent produces at low marginal cost.[6]

---

[6]This expected virtual marginal cost is given by $\nu(\theta_1 + (1 - \alpha)(\theta_2 - \theta_1)) + (1 - \nu)\theta_2 < \theta_2$, where

3

Assume now that the agent first reports its marginal cost and the principal then makes a contract offer. The expectation of future opportunism by the principal causes the agent to manipulate its cost report. Assume that the agent understates a high marginal cost with probability $1 - \sigma \in (0, 1)$, but truthfully reports a low marginal cost. Then, the principal cannot tell for sure whether the agent has a low or a high marginal cost following cost report θ_1. Based on its posterior belief about the distribution of cost types, the principal's sequentially rational output after observing a low cost report is identified by q_1 in Figure 1. The agent's expected virtual marginal cost $\bar{\theta}_1$ following cost report θ_1 is smaller than the expected virtual marginal cost under constrained ex-ante contracting because the principal attaches a larger probability to the event that the agent has low marginal cost under ex-post than ex-ante contracting.[7] Ex-ante contracting therefore generates too little output $\hat{q}$ compared to q_1 contingent on cost report θ_1. Measured relative to $\bar{\theta}_1$, this downward distortion yields an efficiency loss equal to the dotted area in Figure 1. The principal deduces that the agent has high marginal cost following cost report θ_2 because only the high-cost agent reports a marginal cost of this magnitude. The corresponding sequentially rational output equals q_2^{fb}. The ex-ante quantity $\hat{q}$ generates too much output in this case. Measured relative to marginal cost θ_2, this upward distortion generates an efficiency loss equal to the dark area in Figure 1. The increased adaptability of contracts to the reported circumstances of the agent generates an expected net benefit of choosing ex-post contracting over constrained ex-ante contracting measured by the dotted area multiplied by the ex-ante probability $\nu + (1 - \nu)(1 - \sigma)$ of a low cost report plus the dark area multiplied by the ex-ante probability $(1 - \nu)\sigma$ of a high cost report.

Strategic manipulation of cost reports inflicts a loss on the principal although the principal prefers ex-post to constrained ex-ante contracting. The benefit of a more truthful agent stems from the increased likelihood that the agent produces the better suited output q_2^{fb} instead of q_1 if it has a high marginal cost. However, complete truthfulness cannot be sustained as a Perfect Bayesian Equilibrium (PBE) because the principal would then infer from the cost report θ_1 that the agent had in fact a low marginal cost. The sequentially rational response would be to require the agent to produce q_1^{fb} and extract all rent through the transfer. The agent's anticipation of this *ratchet effect* (Weitzman, 1980; Freixas et al., 1985) places an upper bound on σ in equilibrium. The principal can mitigate its own commitment problem by offering an ex-ante contract designed for the low-cost agent, and reserve ex-post contracting for the high-cost agent through an escape clause activated by a reported marginal cost of θ_2. Hence, the principal's optimal mechanism features incomplete commitment by design, where an appropriate ex-ante contract is combined

---

$\alpha \in (0, 1)$ is the weight attached to the agent's rent by the principal.

[7]The posterior probability that the agent has marginal cost θ_1 [θ_2] contingent on the cost report θ_1 under ex-post contracting equals $\frac{\nu}{\nu + (1 - \nu)(1 - \sigma)} > \nu$, $[\frac{(1 - \nu)(1 - \sigma)}{\nu + (1 - \nu)(1 - \sigma)} < 1 - \nu]$.

with an escape clause to maximize expected surplus. The rest of the paper extends the example of constrained contracting to investigate endogenously incomplete commitment sustained by an escape clause in a more general model.

We assume in Section 2 that the agent can be one of a finite number $I \geq 2$ of cost types. The principal offers a menu of $K < I$ different contracts upfront. The mechanism may also specify circumstances that entitle the agent to an ex-post contract offer that renders the initial menu of contracts void. An application of the stochastic revelation principle by Bester and Strausz (2001) results in an optimization program where the principal maximizes expected surplus subject to standard incentive compatibility and individual rationality constraints, plus additional stochastic and sequential rationality conditions. These incorporate strategic misrepresentation of marginal cost by the agent and opportunistic behavior by the principal under ex-post contracting.

Section 3 derives a complete commitment benchmark against which to compare incomplete commitment mechanisms with ex-post contracting. This mechanism entails bundling of cost types into K cost groups because of constrained contracting.

Section 4 establishes fundamental properties of mechanisms with ex-post contracting. An escape clause is defined as the subset of cost reports that entitle the agent to an ex-post contract offer by the principal. We show that the escape clause so defined applies to high marginal cost realizations of the agent. This result vindicates the view of an escape clause as a stipulation that applies to unfavourable agent circumstances. As in the example, the agent understates its marginal cost with positive probability to mitigate the ratchet effect. Simultaneously binding upward and downward incentive compatibility constraints severely limit the degree of flexibility in ex-post contracting: The principal offers at most two different ex-post contracts no matter the size of the escape clause. Nearly all cost reports that activate the escape clause are equally (un)informative to the principal because of uniform randomization by the agent.[8] Consequently, a vague escape clause that does not involve any detailed communication of costs is near optimal when the likelihood of any individual cost realization is small.

We establish circumstances under which it is indeed optimal for the principal to include an escape clause in the mechanism. Section 5 generalizes the example when $K = 1$, so that the principal only can offer a one-size-fits-all contract upfront, to $I \geq 2$ cost types. Section 6 identifies a sufficient condition for when an escape clause is optimal if $K \geq 1$. This condition is fulfilled, for instance, if the principal places a sufficiently

---

[8] A classical model of ex-post contracting is the analysis of strategic information transmission by Crawford and Sobel (1982) where an informed agent sends a signal to the principal who then takes an action. The equilibrium features uniform randomization by the agent with partitions that are more or less informative. Transfer payments and an ex-post participation constraint by the agent reduce the informativeness of the agent's cost reports that trigger the escape clause in our model.

strong weight on efficiency relative to rent extraction. Section 7 identifies the fundamental trade-off involved in determining the size of the escape clause. Broadening it to include a marginally more efficient cost type improves the contract fit for that marginal cost type, but exacerbates dynamic inefficiency by increasing the already excessive ex-post quantity produced by the agent.

Section 8 argues that one can interpret constrained contracting in terms of the maximal number of binding incentive compatibility and individual rationality constraints. We also show how contracting costs can justify ex-post contracting absent any limit to the number of contracts the principal can propose ex ante. The section also discusses differences between an escape clause and another type of safeguard, the renegotiation clause.

Section 9 concludes the paper. All proofs are in the appendix.

Related literature   Our paper contributes to mechanism design with incomplete commitment. Commitment issues arise in a multitude of contracting problems. The seminal contributions by Freixas et al. (1985) and Laffont and Tirole (1988) study short-term contracting in a multi-period framework. Other applications are repeated sales (e.g. Tirole, 2016; Beccuti and Möller, 2018; Breig, 2022), organization design (e.g. Shin and Strausz, 2014) or auction design (e.g. Vartiainen, 2013; Skreta, 2015; Akbarpour and Li, 2020). A fundamental problem is the breakdown of the revelation principle. Bester and Strausz (2001, 2007), Skreta (2006) and Doval and Skreta (2022) develop methodologies for analyzing such mechanisms. All these papers treat incomplete commitment as exogenous. Ours is one of a few to consider incomplete commitment as a *mechanism design variable*, specifically in the form of an escape clause.[9] A string of papers (e.g. Ben-Porath et al., 2019; Hancart, 2022) establish conditions for when commitment does not benefit the principal, that is, ex-post yields the same expected surplus as ex-ante contracting. In our setting with constrained ex-ante contracting, the principal may strictly prefer incomplete commitment through a menu of ex-ante contracts augmented by an escape clause.

Escape clauses have been studied in models of optimal delegation where contracting is constrained in the sense that transfers between the principal and the agent cannot be state-contingent (e.g. Bagwell and Staiger, 2005; Beshkar and Bond, 2017; Coate and Milton, 2019).[10] Activation of an escape clause in this context usually implements a different predefined rule than the default rule, so the mechanism features complete commitment.

---

[9]Fudenberg and Tirole (1983) analyze sequential bargaining under incomplete information. Contracting is constrained by an assumption that the seller provides a single price offer in the first stage. They observe that the seller may strictly benefit from proposing a revised price if the buyer declines the seller's initial offer. Adding this second stage amounts to introducing ex-post contracting.

[10]Since the principal cannot use transfers to accomplish incentive compatibility, the remaining design issue is how much discretion to leave to the agent regarding which actions to choose. Hence, the term optimal delegation. First analyzed by Holmström (1984), Amador and Bagwell (2013) represents the most general treatment.

In our model, triggering the escape clause nullifies the initial menu of contracts and generates an ex-post contract offer. Halac and Yared (2020) analyze delegation under incomplete commitment. Activation of the escape clause by the agent implies that the principal pays a fixed cost to verify the agent's type and thereafter implements the efficient allocation. In our framework, the principal draws inferences based on the agent's observed behavior, but is unable to verify the agent's type directly.

Contracting is constrained by assumption in our model. Rogerson (2003) shows that in certain circumstances the loss in expected surplus of offering a cost-reimbursement contract combined with a fixed-price contract is no more than 25 percent relative to implementation of the second-best menu of contacts. Chu and Sappington (2007) show that a two-item menu in which the fixed-price contract is replaced by a linear cost-sharing rule does better in a broader range of circumstances. Our results suggest that adding an escape clause would increase the expected surplus of simple menus.

2 The contracting problem

The agent (here a monopoly firm) can be one of a finite number $I \geq 2$ of types. An agent of type $i \in \{1, 2, ..., I\} = \mathcal{I}$ has constant marginal production cost of $0 < \theta_i < \infty$. Types are ranked in order of increasing production cost: $\theta_{i+1} > \theta_i$ for all $i \in \{1, ..., I-1\}$. Let $\boldsymbol{\nu} = (\nu_1, ..., \nu_i, ..., \nu_I)$ be the probability distribution over the set of possible types $\boldsymbol{\theta} = (\theta_1, ..., \theta_i, ..., \theta_I)$, 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. Note that $G_0 = \nu_0 = 0$.

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 (transfers are non-negative because the agent cannot be forced to produce at a loss). An agent with marginal cost θ_i operating under contract x obtains the rent

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

The principal (here a regulatory authority) 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 continuous, 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

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

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 marginal costs, the principal therefore sets $U_i(x) = 0$ and maximizes

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

over $q \geq 0$. We assume that $\lim_{q\to 0} S'(q) > \theta_1$ 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 payments:

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

Let the first-best contract be strictly better from the principal's point of view than the outside option, $w_i^{fb} = W_i^{fb}(q_i^{fb}) > 0 \ \forall i \in \mathcal{I}$. The menu $\mathbf{x}^{fb} = (x_1^{fb}, ..., x_i^{fb}, ..., x_I^{fb})$ 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.

Constrained contracting   We study a contracting problem with 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 deviate from the canonical setup by limiting the total number K of different contracts the principal can offer the agent up front. K measures the degree to which the environment constrains contracting between the principal and the agent, with a smaller K meaning a more constrained environment. Contracting is unconstrained for $K \geq I$ because the principal then can offer one contract for each possible realization of the agent's marginal cost. Contracting is ex-ante constrained if $K < I$. 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.

The mechanism features *pure ex-ante contracting* if $\mathcal{B} = \emptyset$. This is the standard *complete commitment* setting of mechanism design, adapted here to the context of constrained contracting. It has *incomplete commitment* if $\mathcal{B} \neq \emptyset$. We interpret this property as the inclusion of the following *escape clause* in the mechanism:

> *All initial contract offers by the principal are void if the agent reports marginal cost θ_j, $j \in \mathcal{B}$. The agent will receive a new contract offer from the principal subsequent to invoking this clause.*

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 menu of contracts $\mathbf{x} = (\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{B}})$, $\mathbf{x}_{\mathcal{B}} = \{x_j\}_{j \in \mathcal{B}}$, is direct by assumption. Bester and Strausz (2001) show for the class of games we consider here that the principal cannot gain anything by extending communication to more general message spaces. 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 $\boldsymbol{\Sigma}$ (see below) maximize the expected surplus of the principal, and the agent communicates its type with the principal only once.

The information that forms the basis of the principal's contract offer in Stage 4 differs from the information underlying contract offers in Stage 1 because the later-stage contract

offer builds on information the principal has obtained from communicating with the agent, namely the cost report θ_j, $j \in \mathcal{B}$. The menu $\mathbf{x}_\mathcal{B}$ contains all elements of $\mathcal{B}$, but at most one of them will ever be proposed in equilibrium. Hence, the maximal number of contracts with positive output offered along the equilibrium path is $K + 1$. Observe also that the principal always can offer the null contract x_0 regardless of K. This is not unreasonable given the simplicity of this particular contract. The null contract is a simple way to handle *partial participation*, where some types do not produce a positive quantity in equilibrium.

A mechanism with incomplete commitment may involve the agent misrepresenting its type with positive probability in equilibrium. The reporting strategy of an agent of type $i \in \mathcal{I}$ is a probability distribution $\boldsymbol{\sigma}_i = (\sigma_{1i}, ..., \sigma_{ji}, ..., \sigma_{Ii})^T \in \Delta^{I-1}$, 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 $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}_1, ..., \boldsymbol{\sigma}_i, ..., \boldsymbol{\sigma}_I) \in \Delta^{2(I-1)}$ 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 introduced 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:

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

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

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

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

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

$$|\mathbf{x}_\mathcal{A}| \leq K \tag{7}$$

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 $\boldsymbol{\Sigma}$ must jointly form a 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 $\boldsymbol{\Sigma}$ 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 $(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}} | \mathcal{A}, \mathcal{B})$ is *incentive efficient* given $(\mathcal{A}, \mathcal{B})$ if it maximizes the principal's expected surplus

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

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 $\mathbf{\Sigma}$.

Complete commitment represents the default mode in mechanism design analysis where $\mathcal{B} = \emptyset$ so that no additional contracting occurs after the agent has reported its marginal cost. 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 $W(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}} | \mathcal{A}, \mathcal{B})$. An *incentive optimal* mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ solves this problem.

Instances can occur when an incomplete commitment mechanism can do as well as one with complete commitment, but no better. We stack the deck against escape clauses by letting the principal choose pure ex-ante contracting in this case. 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* $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ *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,*

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

and the escape clause is minimal in the following sense:

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

The incentive optimal mechanism entails *endogenously incomplete commitment* if $\mathcal{B}^* \neq \emptyset$. Condition (9) simply states that the proposed mechanism with the escape clause maximizes the principal's expected surplus across all possible incentive efficient mechanisms. Condition (10) requires in addition that the principal must strictly prefer the proposed mechanism with the escape clause to any incentive efficient mechanism without any es-

cape clause, but also that the principal cannot find another mechanism with a narrower escape clause $\mathcal{B} \subset \mathcal{B}^*$, $\mathcal{B} \neq \mathcal{B}^*$, that yields the same expected surplus as the proposed mechanism. The incentive optimal escape clause is minimal in this sense.

3 Contracting with complete commitment

This section analyzes properties of incentive efficient mechanisms under the assumption that the principal commits to an ex-ante menu of contracts $(\mathcal{A} \neq \emptyset)$, but does not engage in ex-post contracting $(\mathcal{B} = \emptyset)$, so that the mechanism features complete commitment. Contracting is (weakly) constrained, $K \leq I$. Such mechanisms establish the appropriate benchmark against which to evaluate incomplete commitment mechanisms.

By the revelation principle, we can restrict attention to truthful mechanisms, i.e. incentive feasible mechanisms for which $\boldsymbol{\Sigma} = \mathbf{I}$, where $\mathbf{I}$ is the I-dimensional identity matrix. The principal maximizes (8) over $\mathbf{x}$ subject to (2), (3) and (7).

To minimize transfer payments, the principal ensures that an agent with marginal cost θ_i, $i \in \{1, ..., I-1\}$, is indifferent between its designated contract x_i and the local downward contract x_{i+1}, and that an agent with maximal production cost θ_I is indifferent between producing and not. Formally,

$$U_i(x_i) = t_{i+1} - \theta_i q_{i+1} = U_{i+1}(x_{i+1}) + (\theta_{i+1} - \theta_i)q_{i+1} \ \forall i \in \{1, ..., I-1\}, \ U_I(x_I) = 0.$$

The rent of an agent with cost θ_i is found by adding up rents for more inefficient types,

$$U_i(x_i) = \sum_{j=i}^{I-1}(\theta_{j+1} - \theta_j)q_{j+1} \ \forall i \in \{1, ..., I-1\}, \ U_I(x_I) = 0, \tag{11}$$

loosely the discrete type version of the well-known integral in the continuous type case.

After a summation by parts, the expected surplus of the principal becomes

$$\sum_{i=1}^{I} \nu_i W_i^{sb}(q_i), \ W_i^{sb}(q) = S(q) - (\theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1}))q, \tag{12}$$

which is entirely a function of output $\mathbf{q} = \{q_1, ..., q_i, ..., q_I\}$.

Under constrained ex-ante contracting, set $\mathcal{A}$ is partitioned into K non-empty *cost groups*, indexed by $k \in \{1, ..., K\}$. Each group k defines a set $\mathcal{A}_k \subset \mathcal{A}$ consisting of all cost types that are pooled into the same contract. Each cost group is a convex set because incentive compatibility implies that output is non-increasing in the agent's marginal cost. Monotonicity of output implies that the set of cost types for which there is ex-ante contracting is convex and contains the most efficient cost types: $\mathcal{A} = \{1, ..., A\}$, where θ_A is the marginal cost of the least efficient agent that produces positive output in the mecha-

nism. The mechanism features full participation if $A = I$. Otherwise, $\mathcal{C} = \{A+1,...,I\}$. Let $\theta_{\underline{A}_k}$ be the marginal cost of the most efficient agent contained in $\mathcal{A}_k$, and let $\theta_{\overline{A}_k}$ be the marginal cost of the least efficient agent contained the same cost group. We identify the contract designed for cost group k by $x_{A_k} = (q_{A_k}, t_{A_k})$. 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 $\mathcal{A}_K$.

Maximization of the expected welfare function

$$W_{\mathcal{A}_k}(q) = \sum_{i \in \mathcal{A}_k} \nu_i W_i^{sb}(q) \tag{13}$$

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

$$S'(\hat{q}_{A_k}) = \sum_{i \in \mathcal{A}_k} \frac{\nu_i}{\nu_{\mathcal{A}_k}} \left(\theta_i + \frac{G_{-i}}{\nu_i}(1-\alpha)(\theta_i - \theta_{-i})\right). \tag{14}$$

In this expression, $\nu_{\mathcal{A}_k} = \sum_{i \in \mathcal{A}_k} \nu_i$ measures the ex-ante probability that the agent's marginal cost is contained in $\mathcal{A}_k$. Importantly, the output of an agent with marginal cost $\theta_{\overline{A}_k}$ is larger in a pooled cost group than the corresponding second-best output, $\hat{q}_{A_k} > q_{\overline{A}_k}^{sb}$, where the second-best quantity q_i^{sb} of an agent with marginal cost θ_i solves

$$S'(q_i^{sb}) = \theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1}). \tag{15}$$

This upward inefficiency of constrained contracting occurs as a result of the contract x_{A_k} pooling types with a virtual marginal production costs equal to or below that of the upper boundary type $\overline{A}_k$. Conversely, the output of an agent with marginal cost $\theta_{\underline{A}_k}$ is smaller in a pooled cost group than the second-best quantity, $\hat{q}_{\underline{A}_k} = \hat{q}_{A_k} < q_{\underline{A}_k}^{sb}$, because the contract pools types with a virtual marginal production costs equal to or above that of the lower boundary type $\underline{A}_k$ under constrained contracting.[11]

Assume that the principal strictly prefers the second-best contract to the outside option for all possible cost realizations, $w_i^{sb} = W_i^{sb}(q_i^{sb}) > 0$ for all $i \in \mathcal{I}$. The menu $\mathbf{x}^{sb} = (x_1^{sb}, ..., x_i^{sb}, ..., x_I^{sb})$ of second-best contracts then features full participation. By assumption (1), the second-best contract for each type is unique, $|\mathbf{x}^{sb}| = I$. Therefore $\mathbf{x}^{sb}$ is incentive infeasible under constrained contracting, $K < I$. The principal might then consider reducing contract commitment by allowing ex-post contracting.

---

[11]The appendix contains a fuller treatment, where we also analyze other important aspects of constrained contracting, namely the optimal scope A of the ex-ante mechanism and the optimal partitioning of $\mathcal{A}$ into the K specific cost groups.

4   Contracting with incomplete commitment

This section establishes fundamental properties of the contract menus and the agent's reporting strategies in incentive feasible and incentive optimal mechanisms with incomplete commitment. For any incentive feasible mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ that features ex-post contracting, $\mathcal{B} \neq \emptyset$, we let $\underline{B}$ and B be the minimal and maximal types, respectively, contained in $\mathcal{B}$. $\underline{C}$ 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 $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ in question is incentive optimal.

Lemma 1 (Fundamental properties of contracts) *Any incentive feasible mechanism* $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ *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, $j \in \{\underline{B}, ..., B-1\}$, yield the same contract $x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}})$.*

   (b) *Cost report θ_B yields contract x_B^{fb}.*

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

3. *The mechanism exploits all available ex-ante contractual flexibility, $|\mathbf{x}_{\mathcal{A}^*}^*| = K$.*

4. *All cost types contained in $\mathcal{A}^*$ are more efficient that those contained in $\mathcal{B}^*$, and all ex-ante contracts have strictly higher output than all ex-post contracts, $\theta_{A^*} < \theta_{\underline{B}^*}$ and $q_{A^*}^* > q_{\underline{B}^*}^*$.*

5. *All cost types contained in $\mathcal{B}^*$ are more efficient that those contained in $\mathcal{C}^*$ if the mechanism features partial participation, $\theta_{B^*} < \theta_{\underline{C}^*}$ if $\mathcal{C}^* \neq \emptyset$.*

Proof. See Appendix A.2. ∎

Item 1 of Lemma 1 shows that incomplete commitment increases the degrees of freedom in the mechanism by at most 2 compared to a mechanism with complete commitment. The number of different contracts contained in the menu $\mathbf{x}$ of incentive feasible contracts equals $|\mathbf{x}| = |\mathbf{x}_{\mathcal{A}}| + |\mathbf{x}_{\mathcal{B}}| \leq K + 2$ under incomplete commitment. This limited additional flexibility is not an artifact of our choice to define incentive optimal mechanisms in terms of those with minimal escape clauses, as the result applies to all incentive feasible mechanisms. Instead, flexibility is limited under ex-post contracting by incentive compatibility and the

14

ratchet effect that render incentive compatibility constraints simultaneously downward- and upward-binding. We return to this issue shortly.

By Item 2 of the lemma, a "no-distortion-at-the-bottom" result applies to the upper boundary cost type B that invokes the escape clause, $x_B = x_B^{fb}$ in case $|\mathbf{x}_\mathcal{B}| = 2$. This property follows from the discretion of mechanisms with incomplete commitment. The contract x_B leaves an informational rent to any agent with a smaller marginal cost $\theta_i < \theta_B$. Unlike in the complete commitment setting, the transfers necessary to reach incentive compatibility are sunk after the agent has announced marginal cost θ_B at Stage 4 of the game. Consequently, there is no ex-post trade-off between efficiency and rent extraction. If $|\mathbf{x}_\mathcal{B}| = 2$, the agent reports θ_B only if it indeed represents the agent's true marginal cost; see below. Upon observing cost report θ_B, the principal's sequentially rational choice therefore is to offer the first-best efficient contract for that specific cost type.

Item 3 is evident because the principal always can increase expected surplus by adding a new contract to the mechanism as long as not all ex-ante contractual flexibility has been fully exploited. Items 4 and 5 imply that ex-ante contracts are designed for more efficient types, whereas the escape clause is intended for less favorable agent circumstances. Sufficiently inefficient cost types may not produce at all in equilibrium.

Figure 2 illustrates a partitioning of cost types that is consistent with Lemma 1 for an agent with 16 possible cost types and under the assumption that the principal only can offer one single contract ex ante, $K = 1$. The mechanism is designed such that an agent
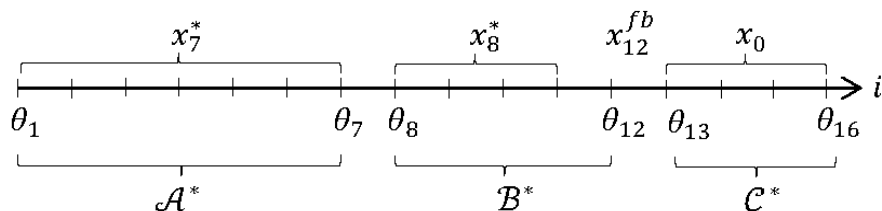

Figure 2: Partitioning of cost types in an incentive optimal mechanism.

who reports marginal cost in the span between θ_1 and θ_7 receives the ex-ante contract x_7^*, where $A^* = 7$ marks the upper boundary cost type for the ex-ante contract. The agent invokes the escape clause for any marginal cost report between θ_8 and θ_{12}, where $\underline{B}^* = 8$ is the lower boundary and $B^* = 12$ the upper boundary cost type for the escape clause. The agent receives the same ex-post contract offer x_8^* for nearly all cost reports that activate the escape clause. The exception is for the upper boundary cost report θ_{12}, subsequent to which the principal offers the agent the associated first-best contract x_{12}^{fb}. The agent is not allowed to produce anything for reported marginal cost equal to or above θ_{13}, so that $\underline{C}^* = 13$ marks the lower boundary cost type for non-production.

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 $x^*_{\underline{\mathcal{B}}^*}$.[12] For instance, $\underline{\mathcal{B}}^* = \{8, 9, 10, 11\}$ in Figure 2.

Lemma 2 (Fundamental properties of reporting strategies) *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 least efficient type truthfully report their cost, $\sigma^*_i = 1 \ \forall i \in \{1, ..., A^* - 1\}$ 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 $\underline{\mathcal{B}}^*$, $\sigma^*_{jA^*} = \frac{1 - \sigma^*_{A^*}}{|\underline{\mathcal{B}}^*|} \ \forall j \in \underline{\mathcal{B}}^*$.*

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, $\sigma^*_{A^*\underline{B}^*} = 1 - \sigma^*_{\underline{B}^*} \geq 0$ 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 $\underline{\mathcal{B}}^*$, $\sigma^*_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|} \ \forall(i, j) \in \{\underline{B}^* + 1, B^* - 1\} \times \underline{\mathcal{B}}^*$ if $|\mathcal{B}^*| \geq 3$.*

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

       i. *$\sigma^*_{B^*} < 1$ 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. See Appendix A.3. ∎

---

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

Lemma 2 contains a near-complete characterization of the incentive optimal reporting strategy under endogenously incomplete commitment, despite the potentially large set of cost types and feasible randomization strategies. It is lengthy because the incentive optimal reporting strategies under ex-post contracting depend on the number $|\mathcal{B}^*|$ of types in the escape clause. The reporting strategies of the individual cost types are mostly simple, which we illustrate based on Figure 2. By Item 1(a) of Lemma 2, any agent with marginal cost between θ_1 and θ_6 truthfully reports its marginal cost and produces under the ex-ante contract x_7^*. This result follows from an application of the revelation principle to the menu of ex-ante contracts. By Items 2(a)iii and 2(b) of the lemma, any agent with marginal cost between θ_8 and θ_{11} uniformly randomizes between cost reports θ_8 and θ_{11} and thereby receives the ex-post contract x_8^* with probability 1. By Item 3 of the lemma, any agent with marginal cost equal to or above θ_{13} truthfully reports its marginal cost and receives the null contract. In Figure 2, remaining uncertainty relates to the probability $1 - \sigma_7^* \in [0,1)$ with which an agent with marginal cost θ_7 exaggerates its marginal cost to activate the escape clause and thereby receive the ex-post contract offer x_8^*. Additional uncertainty is associated with the probability $1 - \sigma_{12}^* \in (0,1)$ with which an agent with marginal cost θ_{12} understates its marginal cost to receive the ex-post contract x_8^*.[13]

Understatement of marginal costs by some relatively inefficient types is fundamental to ensure incentive compatibility of ex-post contracts for cost types $i < B^*$ in equilibrium. Suppose, for instance, that no agent with marginal cost equal to or above θ_9 ever pretends to have marginal cost θ_8 in Figure 2. The principal would then infer from a cost report equal to θ_8 that the agent had marginal cost of no more than θ_8. The sequentially rational ex-post contract offer x_8 would then involve a transfer $t_8 \leq \theta_8 q_8$ by the principal in an effort to minimize agent rent. Anticipating this ratchet effect, an agent with marginal cost θ_8 would expect to earn non-positive rent by a truthful cost report. It would be better for the agent to exaggerate its marginal cost, for instance to θ_{12}, and earn strictly positive rent. Consequently, the mechanism would be incentive incompatible. The requirement that incentive compatibility constraints must be locally upward-binding underlies the finding in Lemma 1 that ex-post contract offers for all cost reports θ_j, $j \in \{\underline{B}, ..., B-1\}$ are the same in any incentive feasible mechanism. We summarize the most important qualitative features of incentive optimal mechanisms with incomplete commitment as:

Observation 1 *The incentive optimal mechanism includes an escape clause to accommodate situations in which the agent has high marginal costs. Ex-post contracts are distorted from an ex-ante perspective mainly because (i) the principal treats informational rent as a sunk cost when making ex-post contract offers; (ii) the agent may trigger the escape clause*

---

[13]An agent with marginal cost $\theta_{\underline{B}^*}$ may understate marginal cost to θ_{A^*} under very specific circumstances; see Item 2(a)ii of Lemma 2.

by exaggerating its cost.

Implementation through a vague escape clause In the mechanisms described above, transactions between the principal and the agent build on highly detailed communication. The mechanism describes for each possible cost report θ_j, $j \in \mathcal{A}$, which contract x_j of K options the agent shall receive; it defines a subset $\mathcal{B}$ of cost types such that the agent triggers the escape clause for all marginal cost reports θ_j, $j \in \mathcal{B}$. Finally, the mechanism may also specify a non-empty subset $\mathcal{C}$ such that the agent does not produce anything for any cost report θ_j, $j \in \mathcal{C}$. Such high level of contractual detail can be costly to implement in practice, and delineating the exact boundaries of the escape clause seems particularly challenging. An interesting question for mechanism design then relates to the extent to which detailed communication adds economic value to the principal.

To gauge the value of communication between the principal and the agent in our context, consider first the posterior beliefs generated by the agent's cost reports under the assumption that the escape clause encompasses three or more types $|\mathcal{B}^*| \geq 3$, and that ex-post contracting yields two different contract offers, $|\mathbf{x}_{B^*}^*| = 2$. On the basis of the reporting strategies in Lemma 2, the principal attaches posterior probability equal to 1 of a truthful report for all cost reports θ_j such that $j \in \mathcal{A}^* \cup \mathcal{C}^*$ and for the specific cost report θ_{B^*}. Because of uniform randomization of cost reports, the principal forms the same set of posterior beliefs,

$$
\begin{aligned}
\mu_{jA^*}^* &= \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \\
\mu_{ji}^* &= \frac{\nu_i}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \quad \forall i \in \underline{\mathcal{B}}^* \\
\mu_{jB^*}^* &= \frac{\nu_{B^*}(1 - \sigma_{B^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)},
\end{aligned}
\tag{16}
$$

subsequent to any cost report θ_j such that $j \in \underline{\mathcal{B}}^*$. Based on these posterior beliefs, the sequentially rational ex-post contract $x_{\underline{B}^*}^*$ chosen by the principal features the transfer payment $t_{\underline{B}^*}^* = \theta_{B^*} q_{\underline{B}^*}^*$ and output requirement $q_{\underline{B}^*}^*$ characterized by

$$
\begin{aligned}
S'(q_{\underline{B}^*}^*) = &\frac{\nu_{A^*}(1 - \sigma_{A^*}^*)(\theta_{A^*} + (1 - \alpha)(\theta_{B^*} - \theta_{A^*}))}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)} \\
&+ \frac{\sum_{i \in \underline{\mathcal{B}}^*} \nu_i(\theta_i + (1 - \alpha)(\theta_{B^*} - \theta_i)) + \nu_{B^*}(1 - \sigma_{B^*}^*)\theta_{B^*}}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \underline{\mathcal{B}}^*} \nu_i + \nu_{B^*}(1 - \sigma_{B^*}^*)}.
\end{aligned}
\tag{17}
$$

The right-hand side of (17) measures the expected virtual marginal cost of an agent that invoked the escape clause by reporting a marginal cost between $\theta_{\underline{B}^*}$ and θ_{B^*-1}, where the

expectation is taken over the principal's posterior belief distribution (16).

Consider now an alternative mechanism and sequence of events:

Stage 1: The principal commits to the VC mechanism consisting of the menu $\mathbf{x}_{\mathcal{A}^*}^{VC}$ of contracts, where

$$q_j^{VC} = q_j^*, \ t_j^{VC} = t_j^* - (\theta_{B^*} - \theta_{A^*})(q_{\underline{B}}^* - q_{\underline{B}}^{VC}) \ \forall j \in \mathcal{A}^*,$$

augmented by the vague escape clause (VC):

> *The agent has the right to terminate the initial agreement and request a new contract offer from the principal.*

Stage 2: The agent

- selects x_i^{VC} if $|\mathcal{A}^*| \geq 2$ and the marginal cost is θ_i, $i \in \{1, ..., A^* - 1\}$.

- selects $x_{A^*}^{VC}$ with probability $\sigma_{A^*}^*$ and activates the escape clause VC with probability $1 - \sigma_{A^*}^*$ if the agent has marginal cost θ_{A^*}.

- activates the escape clause VC if the marginal cost is θ_i, $i \in \mathcal{B}^*$.

- rejects the contract 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 $x_{\underline{B}^*}^{VC}$ featuring transfer payment $t_{\underline{B}^*}^{VC} = \theta_{B^*} q_{\underline{B}^*}^{VC}$ and output requirement $q_{\underline{B}^*}^{VC}$ characterized by

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

$$(18)$$

The above mechanism features restricted communication in the sense that the agent never directly reports its cost to the principal, only implicitly through its choices. The agent either self-selects one of the ex-ante contracts, activates the escape clause, or completely rejects the offer after which the game ends. The escape clause VC essentially is a termination clause. Contrary to the escape clause that forms the foundation of the incentive optimal direct mechanism, the above clause applies generally. In the above game, the principal cannot challenge the agent's decision to activate the clause. We now show the consequences when the principal introduces a mechanism with simpler communication.

The principal forms the posterior beliefs

$$\mu_{A^*}^{VC} = \frac{\nu_{A^*}(1 - \sigma_{A^*}^*)}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \mathcal{B}^*} \nu_i}, \ \mu_i^{VC} = \frac{\nu_i}{\nu_{A^*}(1 - \sigma_{A^*}^*) + \sum_{i \in \mathcal{B}^*} \nu_i} \ \forall i \in \mathcal{B}^* \qquad (19)$$

about the agent's distribution of marginal costs subsequent to the activation of the escape clause by the agent. The right-hand side of (18) measures the expected virtual marginal cost of an agent that invoked the escape clause VC, where the expectation is taken over the principal's posterior belief distribution (19). Those beliefs differ from the initial mechanism only by the factor $v_{B^*}\sigma^*_{B^*}$ in (16). The belief system (19) places higher posterior probability than (16) on the agent having the high marginal cost realization θ_{B^*}. This property increases the expected virtual marginal cost relative to the initial mechanism, which implies that the equilibrium output requirement satisfies $q^{VC}_{\underline{B}^*} < q^*_{\underline{B}^*}$. This property of the VC mechanism reduces the transfer payments necessary to maintain incentive compatibility, which tends to increase the expected surplus of the principal under the VC compared to the initial mechanism. However, the initial mechanism provides more flexibility than the VC mechanism, in particular because the former mechanism implements $x^{fb}_{B^*}$ subsequent to the marginal cost report θ_{B^*}. These pros and cons are both negligible for small $v_{B^*}\sigma^*_{B^*}$ because then the contracts are almost identical in both mechanisms.

Proposition 1 *Assume that the incentive optimal mechanism features incomplete commitment. Then there exists a restricted communication mechanism augmented by a vague escape clause that can be sustained as a PBE. This mechanism generates in the limit $\nu_{B^*}\sigma^*_{B^*} \to 0$ the same expected surplus as the incentive optimal mechanism.*

Proof. See Appendix A.4. ∎

Proposition 1 holds because all cost reports θ_j, $j \in \underline{\mathcal{B}}^*$, provide exactly the same information to the principal about the cost distribution of the agent. Only the report θ_{B^*} potentially produces different information than the others. This additional information is negligible in expectation in a large type space (so that ν_i is small for all $i \in \mathcal{I}$). Hence, the proposition shows that the value of direct communication is small under plausible circumstances. A policy implication is that the principal has little to gain from specifying a detailed escape clause. A vague escape clause does nearly as well in equilibrium.

This section has characterized properties of incentive optimal mechanisms under incomplete commitment. However, we have not yet established if there are circumstances under which the principal strictly prefers incomplete commitment over pure ex-ante contracting. The next sections establish sufficient conditions for this to be the case.

5 Maximally constrained ex-ante contracting

Consider the polar extreme case of maximally constrained ex-ante contracting where the principal can offer at most one contract upfront to the agent, that is, $K = 1$. Many

real-life contracts have this one-size-fits-all property. We first compare pure ex-ante with pure ex-post contracting. This analysis is interesting in its own right as it provides insight into the relative merits of offering contracts ex ante relative to ex post. In the first case, the principal commits to one single contract $x = (q, t)$. The agent accepts this contract if it has marginal cost $\theta_i \leq \frac{t}{q}$, but rejects it otherwise. Under pure ex-post contracting, the principal does not propose any contract up front. Instead, the principal 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 based on the cost report. The ex-post contracting situation is particularly simple under full participation. The agent then reports its cost, after which the principal offers a contract. We will demonstrate that the principal strictly prefers pure ex-post over pure ex-ante contracting under the assumption that $K = 1$.

The surplus-maximizing ex-ante contract   Suppose the principal implements an 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 $t_A = \theta_A q$ for arbitrary quantity q. Maximization of the principal's expected surplus

$$\tilde{W}_A(q) = \sum_{i=1}^{A} \nu_i [S(q) - (\theta_i + (1 - \alpha)(\theta_A - \theta_i))q] \tag{20}$$

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

$$S'(\hat{q}_A) = \sum_{i=1}^{A} \frac{\nu_i}{G_A}(\theta_i + (1 - \alpha)(\theta_A - \theta_i)). \tag{21}$$

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 $\hat{A} \geq 2$.[14] Let the maximum be strict:

$$\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}}) > \tilde{W}_A(\hat{q}_A) \ \forall A \in \mathcal{I} \cup \emptyset, \ A \neq \hat{A}. \tag{22}$$

Incentive feasible ex-post contracts   Consider pure ex-post contracting, and assume that the principal allows ex-post contracting if and only if the agent reports marginal cost equal to or below $\theta_{\hat{A}}$. Assume that the agent randomizes uniformly across all cost reports θ_j, $j \in \{1, ..., \hat{A} - 1\}$ if it has marginal cost $\theta_i < \theta_{\hat{A}}$. Let $\sigma_{\hat{A}} \in (0, 1)$ be the

---

[14]A sufficient condition for $\hat{A} \geq 2$ is $W_2^{sb}(q_1^{fb}) > 0$. This condition is satisfied for instance if $\theta_2 - \theta_1$ is small, since then $W_2^{sb}(q_1^{fb}) \approx w_1^{fb} > 0$.

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 $\hat{A} \leq I - 1$, and the agent has marginal cost $\theta_i > \theta_{\hat{A}}$. The reporting strategies generate the same probability distribution

$$\mu_{ji} = \frac{\nu_i}{G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})} \; \forall i \in \{1, ..., \hat{A} - 1\}, \; \mu_{j\hat{A}} = \frac{\nu_{\hat{A}}(1 - \sigma_{\hat{A}})}{G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})} \quad (23)$$

of the agent's true marginal cost θ_i for any cost report θ_j, $j \in \{1, ..., \hat{A} - 1\}$.

The sequentially rational contract offer by the principal equals $x_{\hat{A}}^{fb}$ subsequent to receiving the cost report $\theta_{\hat{A}}$ as the principal attaches probability 1 to the event that the agent was truthful. After receiving a cost report θ_j, $j \in \{1, ..., \hat{A} - 1\}$, the principal offers the transfer $t = \theta_{\hat{A}} q$ for any arbitrary q, under the assumption that the principal wants the agent to produce for all marginal cost realizations equal to or below $\theta_{\hat{A}}$. The expected ex-post surplus of the principal then equals

$$\tilde{\Omega}_{\hat{A}}(q, \sigma_{\hat{A}}) = \sum_{i=1}^{\hat{A}-1} \nu_i[S(q) - (\theta_i + (1 - \alpha)(\theta_{\hat{A}} - \theta_i))q] + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})[S(q) - \theta_{\hat{A}}q] \quad (24)$$

divided by $G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})$. Maximization yields q_1 as the solution to

$$S'(q_1) = \frac{\sum_{i=1}^{\hat{A}-1} \nu_i(\theta_i + (1 - \alpha)(\theta_{\hat{A}} - \theta_i)) + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})\theta_{\hat{A}}}{G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})} < \theta_{\hat{A}} = S'(q_{\hat{A}}^{fb}). \quad (25)$$

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

$$\frac{\partial q_1}{\partial \sigma_{\hat{A}}} = \frac{-\alpha \nu_{\hat{A}}}{S''(q_1)} \frac{\sum_{i=1}^{\hat{A}-1} \nu_i(\theta_{\hat{A}} - \theta_i)}{(G_{\hat{A}-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}}))^2} > 0,$$

since the expected virtual marginal cost of an agent reporting $\theta_j < \theta_{\hat{A}}$ is smaller when $\sigma_{\hat{A}}$ is larger.

The ex ante expected surplus of the principal equals

$$\Omega_{\hat{A}}(q_1, \sigma_{\hat{A}}) = \tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}}) + \nu_{\hat{A}}\sigma_{\hat{A}}w_{\hat{A}}^{fb} \quad (26)$$

under pure ex-post contracting. The principal benefits from a more truthful agent,

$$\frac{\partial \Omega_{\hat{A}}}{\partial \sigma_{\hat{A}}} = \nu_{\hat{A}}[w_{\hat{A}}^{fb} - W_{\hat{A}}^{fb}(q_1)] > 0,$$

22

because an agent with marginal cost $\theta_{\hat{A}}$ is more likely to receive a contract better suited (from the principal's perspective) to the agent's particular circumstances if $\sigma_{\hat{A}}$ is larger. The marginal effect on q_1 of an increase in $\sigma_{\hat{A}}$ has only a second-order effect on the principal's expected surplus.

Even the agent benefits in expectation from a more truthful reporting strategy, as

$$\frac{\partial}{\partial \sigma_{\hat{A}}} \sum_{i=1}^{\hat{A}} \nu_i U_i(x_1) = \sum_{i=1}^{\hat{A}-1} \nu_i(\theta_{\hat{A}} - \theta_i)\frac{\partial q_1}{\partial \sigma_{\hat{A}}} > 0.$$

The agent is indifferent between truthfully reporting its marginal cost $\theta_{\hat{A}}$ and understating it to $\theta_j < \theta_{\hat{A}}$, all else equal. However, the agent benefits from the indirect effect on q_1 because the higher output increases informational rent whenever the agent has marginal cost $\theta_i < \theta_{\hat{A}}$. Both the principal and the agent therefore agree ex ante that more truthful behavior would be better under pure ex-post contracting.

However, there is an upper bound to the agent's truthfulness about its marginal cost $\theta_{\hat{A}}$ that is consistent with sequential rationality of x_1. If $\sigma_{\hat{A}}$ is too large, then it becomes sequentially rational for the principal to exclude this cost type after receiving a cost report $\theta_j < \theta_{\hat{A}}$. Doing so would allow the principal to save on informational rent without sacrificing much efficiency. This is the ratchet effect. Suppose the principal, instead of x_1, implements an ex-post contract that only an agent with marginal cost equal to or below θ_A, $A \in \{1, ...\hat{A}-1\}$, would accept. The associated rent-minimizing transfer by the principal equals $t_A = \theta_A q$ for arbitrary quantity q. Based on the posterior beliefs (23), this alternative strategy yields expected surplus $\tilde{W}_A(q)$ divided by $G_{A^-1} + \nu_{\hat{A}}(1 - \sigma_{\hat{A}})$. The optimal deviation contract therefore equals $\hat{x}_A$. The expected deviation profit delivers a necessary and sufficient condition

$$\tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}}) \geq \tilde{W}_A(\hat{q}_A) \ \forall A \in \{1, ...\hat{A}-1\}$$

for sequential rationality of x_1. By comparison of (25) with (21), the ex-post quantity q_1 converges to $\hat{q}_{\hat{A}}$ when $\sigma_{\hat{A}} \to 0$. Then, $\tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}})$ converges to $\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}})$ as $\sigma_{\hat{A}} \to 0$. By way of (22), it follows that x_1 is sequentially rational if $\sigma_{\hat{A}} > 0$ is sufficiently close to zero. We conclude that the pair $(x_1, x_{\hat{A}}^{fb})$ of ex-post contracts can be sustained in equilibrium if the probability is sufficiently large that an agent with marginal cost $\theta_{\hat{A}}$ understates its cost to achieve the anticipated ex-post contract x_1 instead of the ex-post contract $x_{\hat{A}}^{fb}$.

Comparison of ex-post and ex-ante contracting A comparison of the incentive efficient ex-ante contract $\hat{x}_{\hat{A}}$ with the menu $(x_1, x_{\hat{A}}^{fb})$ of ex-post contracts delivers

$$\Omega_{\hat{A}}(q_1, \sigma_{\hat{A}}) - \tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}}) = \tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}}) - \tilde{\Omega}_{\hat{A}}(\hat{q}_{\hat{A}}, \sigma_{\hat{A}}) + \nu_{\hat{A}} \sigma_{\hat{A}}(w_{\hat{A}}^{fb} - W_{\hat{A}}^{fb}(\hat{q}_{\hat{A}})) > 0 \ \forall \sigma_{\hat{A}} > 0. \quad (27)$$

In this expression, $\tilde{\Omega}_{\hat{A}}(q_1, \sigma_{\hat{A}}) > \tilde{\Omega}_{\hat{A}}(\hat{q}_{\hat{A}}, \sigma_{\hat{A}})$ because q_1 represents a better ex-post trade-off between efficiency and rent extraction than $\hat{q}_{\hat{A}}$, given $\sigma_{\hat{A}}$. In addition, ex-post contracting enables the principal to supply a tailor-made ex-post contract to an agent with reported marginal cost $\theta_{\hat{A}}$. As $\hat{x}_{\hat{A}}$ maximizes the principal's expected surplus across all incentive feasible pure ex-ante contracts, there exist incentive feasible pure ex-post contracts $(x_1, x_{\hat{A}}^{fb})$ that strictly outperform all incentive feasible pure ex-ante contracts.

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

Under maximally constrained contracting, ex-post contracts generally offer superior fit to the economic environment compared to a pure ex-ante contract, despite strategic manipulation of cost reports by the agent. Since under ex-post contracting the principal can always add an ex-ante contract without reducing expected surplus, we establish the following result without additional proof:

Proposition 3 *The incentive optimal mechanism features an escape clause if contracting is maximally ex-ante constrained ($K = 1$), and the principal's surplus-maximizing ex-ante contract involves some pooling of cost types ($\hat{A} \geq 2$).*

6 Generally constrained ex-ante contracting

So far we have established the incentive optimality of introducing an escape clause when the principal only can offer a one-size-fits-all contract under ex-ante contracting. The value of ex-post contracting is smaller if the principal can offer more complex contracts ex ante, that is, when K is larger, because then the principal can account for more contingencies already beforehand. Still, the principal may have insufficient degrees of freedom to account for all potential contingencies ex ante. This occurs for any K if the type space is large. This plausible scenario leads to the question whether escape clauses are incentive optimal for more generally constrained mechanisms such that $1 \leq K < I$? The next result establishes a simple sufficient condition for this to be the case.

Lemma 3 *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. See Appendix A.5. ∎

By pooling a subset of cost types into a cost group K that contains multiple cost types, the least efficient cost type $\hat{A}$ in that group produces an inefficiently high output from the viewpoint of the second-best contract, $\hat{q}_{\hat{A}} > q_{\hat{A}}^{sb}$; see Section 3. The surplus-maximizing output could potentially be upward distorted even compared to the first-best solution, $\hat{q}_{\hat{A}} > q_{\hat{A}}^{fb}$. To see why, subtract (14) from $S'(q_{\hat{A}}^{fb}) = \theta_{\hat{A}}$ to get the difference

$$\nu_{\hat{A}_K}[S'(q_{\hat{A}}^{fb}) - S'(\hat{q}_{\hat{A}})] = \sum_{i \in \hat{A}_K} \nu_i(\theta_{\hat{A}} - \theta_i) - \sum_{i \in \hat{A}_K} G_{-i}(1-\alpha)(\theta_i - \theta_{-i})$$

in marginal expected surplus. The first term on the right-hand side measures the effect on the virtual marginal cost of pooling cost type $\hat{A}$ with more efficient cost types $i < \hat{A}$. The pooling effect tends to increase $\hat{q}_{\hat{A}}$ compared to the first-best output $q_{\hat{A}}^{fb}$. The second term on the right-hand side above measures the adjustment of the informational rent, which occurs under constrained contracting, but not under first-best contracting. This adjustment tends to reduce $\hat{q}_{\hat{A}}$ relative to $q_{\hat{A}}^{fb}$. By adding a small escape clause to the initial mechanism that only contains the $\hat{A}$ type, the principal can reduce an excessive distortion $\hat{q}_{\hat{A}} > q_{\hat{A}}^{sb}$ for the cost type $\hat{A}$ if $\hat{q}_{\hat{A}} > q_{\hat{A}}^{fb}$. The reason is that the sequentially optimal contract implements $x_{\hat{A}}^{fb}$.

The adjustment for informational rent vanishes in the limit as $\alpha \to 1$ since rent extraction has a negligible effect on output $\hat{q}_{\hat{A}}$ when α is close to one. We conclude:

Proposition 4 *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).*

Sometimes the principal can offer more complex contracts than one-size-fits-all and may place a lot of weight on rent extraction in the design of the mechanism. The above results do not apply when $K \geq 2$ and α is small. Still, there are plausible circumstances under which the incentive optimal mechanism features incomplete commitment:

Proposition 5 *Assume that the mechanism that maximizes the principal's expected surplus across all incentive feasible mechanisms with complete commitment features partial participation. Assume also that the incremental difference in marginal production costs is small for the boundary cost type ($\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$ is close to zero). The incentive optimal mechanism then contains an escape clause if $W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > 0$.*

Proof. See Appendix A.6. ∎

In the best mechanism with full commitment, the principal is nearly indifferent between including an agent with boundary marginal cost $\theta_{\hat{A}}$ or excluding it from the mechanism if the type space is large and the mechanism features partial participation. A better alternative could be to include the boundary type through an escape clause instead. This occurs if the expected surplus of doing so is sufficiently high in the sense that $W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > 0$. The second-best welfare function is the correct welfare metric because it conveys the optimal trade-off between efficiency and rent extraction from an ex-ante perspective.

The two previous sections have established circumstances under which the incentive optimal mechanism features incomplete commitment. However, we have not discussed the trade-offs faced by the principal in the design of the escape clause. This is the topic of our next section.

7   Trade-offs in the design of an escape clause

To examine the boundaries of the escape clause, consider a mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ featuring an escape clause $\mathcal{B}^*$ that contains at least two cost types. Assume that the agent truthfully reports its marginal cost θ_i if $i \in \mathcal{A}^* \cup \mathcal{C}^*$ and randomizes uniformly across all marginal cost reports θ_j, $j \in \mathcal{B}^*$, if $i \in \mathcal{B}^*$. This mechanism delivers the expected surplus

$$w^* = \sum_{i \in \mathcal{A}^*} \nu_i W_i(x_i^*) + \Omega_{\mathcal{B}^*}(q_{\underline{B}^*}^*).$$

In this expression,

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

defines the expected surplus of the ex-post contract $x_{\underline{B}^*}^* = (q_{\underline{B}^*}^*, \theta_{B^*} q_{\underline{B}^*}^*)$.

Compare now the initial mechanism to a modified mechanism $(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$ with an incrementally smaller escape clause. Specifically, the lower boundary type $\underline{B}^*$ is included in the least efficient cost group, so that $\mathcal{A}_K = \mathcal{A}_K^* \cup \underline{B}^*$, and the escape clause is correspondingly reduced to $\mathcal{B} = \mathcal{B}^* \backslash \underline{B}^*$. All other cost groups remain the same as before. The modification of the escape clause reduces ex-post output to $q_{\underline{B}} < q_{\underline{B}^*}^*$ in the sequentially rational ex-post contract $x_{\underline{B}} = (q_{\underline{B}}, \theta_{B^*} q_{\underline{B}})$ because the escape clause now consists of less efficient cost types than before. This output reduction changes the downward incentive compatibility constraints, which affects the transfer to the more efficient types. The initial

ex-ante contracts are modified as follows:

$$q_j = q_j^*, \ t_j = t_j^* + t, \ t = (\theta_{\underline{B}^*} - \theta_{A^*})(q_{A^*}^* - q_{\underline{B}^*}^*) - (\theta_{B^*} - \theta_{A^*})(q_{\underline{B}}^* - q_{\underline{B}}) \ \forall j \in \mathcal{A}^*,$$

whereas $x_{\underline{B}^*}^* = (q_{A^*}^*, t_{A^*}^* + t)$. Every ex-ante contract has the same output requirement as before but all transfer payments are adjusted by the same amount.[15] This particular mechanism generates the expected surplus

$$w = \sum_{i \in \mathcal{A}} \nu_i W_i(x_i) + \Omega_{\mathcal{B}}(q_{\underline{B}}) = \sum_{i \in \mathcal{A}^*} \nu_i W_i(x_i^*) - G_{A^*}(1 - \alpha)t + \nu_{\underline{B}^*} W_{\underline{B}^*}(x_{\underline{B}^*}) + \Omega_{\mathcal{B}}(q_{\underline{B}}).$$

The second term measures the expected economic effect of the change in transfer payments to any agent with marginal cost equal to or below θ_{A^*}. The third effect is the expected surplus of an agent with marginal cost $\theta_{\underline{B}^*}$ who is now on an ex-ante contract. The last term is the expected surplus of the escape clause.

We can decompose the net benefit to the principal of the incentive optimal mechanism over the modified one into three separate effects:

$$w^* - w = \nu_{\underline{B}}^*[W_{\underline{B}^*}^{sb}(q_{\underline{B}^*}^*) - W_{\underline{B}^*}^{sb}(q_{A^*}^*)]$$
$$- G_{\underline{B}^*}(1 - \alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q_{\underline{B}^*}^* - q_{\underline{B}}) - [\Omega_{\mathcal{B}}(q_{\underline{B}}) - \Omega_{\mathcal{B}}(q_{\underline{B}^*}^*)].$$

The first term on the right-hand side is the effect on the expected surplus of an agent with marginal cost $\theta_{\underline{B}^*}$ producing ex-post output $q_{\underline{B}^*}^*$ instead of the ex-ante output $q_{A^*}^*$, evaluated on the basis of the second-best welfare function. The ex-post contract provides a better ex-ante trade-off between efficiency and rent extraction than the ex-ante contract if the ex-post output is closer to the second-best output than the ex ante output so that $q_{\underline{B}}^{sb} < q_{\underline{B}^*}^* < q_{A^*}^*$. Output $q_{\underline{B}^*}^*$ is larger under escape clause $\mathcal{B}^*$ compared to $q_{\underline{B}}$ under the smaller escape clause $\mathcal{B}$ because an agent that has invoked the escape clause on average is more efficient under $\mathcal{B}^*$ than $\mathcal{B}$. The increase in output has a first-order effect on ex ante expected informational rent because the principal chooses $q_{\underline{B}^*}^*$ and $q_{\underline{B}}$ ex post after the information rent is sunk. The dynamic inefficiency of a broader escape clause is measured by the first negative term on the second line above. The final effect is the one on the inefficiency of output $q_{\underline{B}^*}^*$, relative to $q_{\underline{B}}$ under the ex-post welfare function $\Omega_{\mathcal{B}}(q)$.[16]

Based on these trade-offs, the initial mechanism is incentive optimal only if:

$$\nu_{\underline{B}}^*[W_{\underline{B}^*}^{sb}(q_{\underline{B}^*}^*) - W_{\underline{B}^*}^{sb}(q_{A^*}^*)] \geq G_{\underline{B}^*}(1 - \alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q_{\underline{B}^*}^* - q_{\underline{B}}).$$

---

[15] The proof that this mechanism is incentive feasible is available on request.

[16] This final effect is of second-order importance in a large type space in the sense that $\frac{\Omega_{\mathcal{B}}(q_{\underline{B}^*}^*) - \Omega_{\mathcal{B}}(q_{\underline{B}})}{\theta_{\underline{B}^*+1} - \theta_{\underline{B}^*}} \to$ 0 for $\theta_{\underline{B}^*+1} - \theta_{\underline{B}^*} \to 0$.

We interpret the left-hand side of this inequality as the *expected improvement in contract fit* associated with an agent of marginal cost $\theta_{\underline{B}^*}$ receiving an ex-post contract $x_{\underline{B}^*}^*$ that is better suited to that agent (from the principal's perspective) than the ex-ante contract $x_{A^*}^*$. This benefit must be sufficiently large to outweigh the *expected increase in dynamic inefficiency* associated with ex-post contracting under the escape clause on the right-hand side of the inequality. We summarize this fundamental trade-off as:

Remark 1 *The design of an incentive optimal escape clause balances the expected improvement in contract fit against the expected increase in dynamic inefficiency.*

8   Discussion

Contract complexity   We have interpreted K as the number of contracts contained in the ex-ante menu offered to the agent, a property of many real-life mechanisms. Regulatory mechanisms most often have only one single contract. Mobile subscription plans with different monthly download allowances, mortgage loans with different interest rate maturities, and electricity retail contracts with hourly, monthly or yearly average prices, are examples of menus of contracts with a finite number of offers. However, in our context the principal could as well specify a single ex-ante rule defined as a continuous function $x(\theta) = (q(\theta), t(\theta))$ with the property that all contracts x_{A_k}, $k \in \{1, ..., K\}$, lie somewhere on $x(\theta)$. With this formulation, it is not self-evident that one can discuss constrained contracting in terms of the number of contracts. However, K more generally specifies the maximal number of binding incentive compatibility and individual rationality constraints.

Let a local incentive compatibility constraint $U_i(x_i) = U_i(x_{i+1})$ be *non-trivially binding* if $x_i \neq x_{i+1}$ (trivially binding if $x_i = x_{i+1}$). Likewise, an individual rationality constraint $U_i(x_i) = 0$ is non-trivially binding if $x_i \neq x_0$ (trivially binding if $x_i = x_0$). An incentive efficient mechanism with pure ex-ante contracting features K non-trivially binding IC and IR constraints in our model with constrained contracting, $K < I$. These binding constraints pin down the expected surplus to the principal and the agent of the mechanism. The continuous mapping $x(\theta)$ will have these exact same properties. Therefore the parameter K in a more general sense represents a measure of *contract complexity*. The larger is K, the more complex is the mechanism.

Sources of constrained contracting   The maximal number K of contracts the principal can offer the agent ex ante is exogenous in the model. There can be several reasons why a principal would limit the number of contracts. For instance, the Swedish Regulatory Authority for the Electricity Market offers one single regulatory contract to avoid

discriminating across different electricity distribution networks ex post. We briefly explore a different avenue to explain $K < I$. Let K be endogenous, but assume that there is a fixed cost C of adding each additional contract to any given menu of contracts. This cost arises both for ex-ante and ex-post contracts. An important difference is that the cost of specifying an ex-ante contract x_j, $j \in \mathcal{A}$, arises regardless of whether this contract stipulation is actually used, whereas the cost of specifying ex-post contract x_j only arises after the agent has activated the escape clause by reporting marginal cost θ_j, $j \in \mathcal{B}$.

The contracting cost approach, introduced by Dye (1985), has suffered criticism for being too ad hoc, as it is difficult to relate the economic magnitude of such costs relative to other important economic effects of contracting. For instance, Segal (1999) argues that the economic value of a contract stipulation is likely to be large relatively to the cost of inserting this contract stipulation into the contract. If so, then contracts should be close to being complete (K is close to I in our setting). From that perspective, costs of writing contracts cannot explain the prevalence of incomplete contracting.

Adding an arbitrary contract to an initial menu of contracts is unlikely to be very costly. But not all contract additions generate economic value to the principal. Additional contracts must be incentive compatible and increase the principal's expected surplus relative to the initial menu of contracts. Characterizing an incentive compatible and surplus-increasing incremental contract is much more challenging in terms of time and resources than simply adding an arbitrary contract. The complexity of this task is probably larger and its value smaller as the number of initial contracts is larger. For these reasons we let the cost of adding a contract be non-negligible. Yet, we will characterize circumstances under which the principal would rather constrain the number of ex-ante contracts and rely on some ex-post contracting than increase the number of ex-ante contracts, even for small but positive C.

Revisit the example in the introduction. The principal has four options under complete commitment. Null contracting, yields zero expected surplus. Option two is to offer the first-best efficient contract to the low-cost type. This contract yields expected surplus $\nu w_1^{fb} - C$. The third option is a single contract offer $\hat{x} = (\hat{q}, \theta_2 \hat{q})$ that is acceptable to the agent regardless of its marginal cost. This contract yields the expected surplus $\hat{w} - C$. Option four is to implement the second-best mechanism at the expense of increased contracting costs, which yields expected surplus

$$w^{sb} = \nu w_1^{fb} + (1-\nu) w_2^{sb} - 2C.$$

Assume that participation is valuable in the sense of $\hat{w} > \nu w_1^{fb}$ and that contracting costs are small in the sense of $C < \hat{w}$. Under these assumptions, the principal prefers either $\hat{x}$ or the second-best menu under complete commitment.

We consider two possible mechanisms under incomplete commitment. The first option is pure ex-post contracting as described in the introduction. This mechanism yields expected surplus $\tilde{\omega} - C$. We showed in Section 5 that $\tilde{\omega} > \hat{w}$ for some incentive feasible mechanism with pure ex-post contracting, so $\hat{x}$ cannot be incentive optimal under costly contracting. The second mechanism features ex-ante contract offer $(q_1^{fb}, \theta_1 q_1^{fb} + (\theta_2 - \theta_1) q_2^{fb})$ which the agent receives by reporting marginal cost θ_1. The agent invokes an escape clause by reporting marginal cost θ_2, after which the principal offers the ex-post contract x_2^{fb}. This mechanism is incentive feasible and yields expected surplus

$$w_B = \nu w_1^{fb} - C + (1 - \nu)[W_2^{sb}(q_2^{fb}) - C].$$

The principal faces a trade-off relative to the second-best mechanism of

$$w_B - w^{sb} = \nu C - (1 - \nu)[w_2^{sb} - W_2^{sb}(q_2^{fb})].$$

The principal reduces expected contracting costs by including an escape clause in the mechanism. However, from an ex-ante perspective the second-best contract x_2^{sb} offers a better balance between efficiency and rent extraction than the discretionary contract x_2^{fb} when the agent is inefficient. The benefit of increasing the number of contracts from 1 to 2 is measured in terms of the expected incremental increase in surplus. This increase can be small, even if the absolute value of contracting is large. For instance, reduced contracting costs dominate increased contractual efficiency for arbitrary $C > 0$ if the high-cost event is unlikely, i.e. ν is large. Intuitively, an ex-post contract is better than an ex-ante contract to cover unlikely events. The cost effect dominates also if α is sufficiently close to one or if ν is sufficiently close to zero. The net benefit to the principal of second-best relative to first-best contracting is small if $\nu(1 - \alpha)$ is close to zero as the principal would mainly care about efficiency in the choice of q_2.[17] We summarize these findings as

Remark 2 *Constrained contracting with incomplete commitment can be justified on the basis of small contracting costs if the likelihood of inefficient outcomes is sufficiently small or the principal cares sufficiently about efficiency relative to minimizing agency rent.*

Other clauses The benchmark against which we evaluate mechanisms with incomplete commitment is the mechanism that maximizes the principal's expected surplus in the set of incentive feasible mechanisms with complete commitment. The general message is that incentive feasible mechanisms sometimes exist that strictly improve upon the complete

---

[17]The marginal efficiency effect $w_2^{sb} - W_2^{sb}(q_2^{fb})$ vanishes in the limit as $\alpha \to 1$ because then $q_2^{sb} \to q_2^{fb}$. To obtain the second result, use L'Hôpital's rule to get $\lim_{\nu \to 0} \frac{w_2^{sb} - W_2^{sb}(q_2^{fb})}{\nu} = \lim_{\nu \to 0} [S'(q_2^{sb}) - \theta_2] \frac{dq^{sb}}{d\nu} = 0$ and therefore $\lim_{\nu \to 0} \frac{w_B - w^{sb}}{\nu} = C > 0$.

commitment benchmark under constrained contracting. All such improvements must necessarily involve some form of incomplete commitment.

We have modeled incomplete commitment as an escape clause triggered by the agent through a marginal cost report θ_j, $j \in \mathcal{B}$, where $\mathcal{B}$ is specified in the mechanism upfront. All initial contract offers are then invalidated. This formulation of incomplete commitment is inspired by qualitative properties of real-life escape clauses. However, our results do not rule out the possibility that other mechanisms featuring incomplete commitment could outperform mechanisms with escape clauses, from the viewpoint of the principal.

A *renegotiation clause* is similar in spirit to an escape clause. Invoking a renegotiation clause also triggers ex-post contracting. A main difference is that the agent under a renegotiation clause will reject any ex-post contract offer that delivers lower rent than the best possible ex-ante contract.[18] To see the implications, assume that the agent receives one of K ex-ante contracts for cost reports θ_j, $j \in \mathcal{A} = \{1, ...A\}$. The agent triggers the renegotiation clause by reporting θ_j, $j \in \mathcal{B} = \{A+1, ...B\}$, $B \geq A+1$. An agent with marginal cost θ_i, $i \in \mathcal{B}$, that has invoked the renegotiation clause, accepts the ex-post contract x_j if and only if

$$U_i(x_j) \geq U_i(x_A).^{19}$$

The right-hand side of this ex-post individual rationality constraint depends on the agent's marginal cost θ_i, unlike in the case of the escape clause where the right-hand side is zero. This modification has an impact on the principal's sequentially rational choice of the ex-post contract. For instance, the principal is unable to extract all rent ex post even if the agent truthfully reports its marginal cost. This property should dampen the ratchet effect associated with ex-post contracting and will most likely also affect the extent to which the agent manipulates cost reports in equilibrium. As our paper has shown, such effects have implications for the incentive optimal mechanism that are far from obvious.

9 Conclusion

This paper has developed a theory of endogenously incomplete commitment in mechanism design, framed in the context of an escape clause. Triggering an escape clause terminates the initial agreement and generates a revised contract offer from the principal. The motive for an escape clause arises from an assumption of constrained contracting where

---

[18]A renegotiation clause means that the mechanism may feature partial renegotiation (i.e. only for a subset of cost reports) as opposed to full renegotiation as has previously been studied by Hart and Tirole (1988), Laffont and Tirole (1990), and more recently by Maestri (2017).

[19]Formally, the agent evaluates x_j against all x_h, $h \in \mathcal{A}$. However, incentive compatibility and monotonicity of output of the menu of ex-ante contracts implies $U_i(x_A) - U_i(x_h) = U_A(x_A) - U_A(x_h) + (\theta_i - \theta_A)(q_h - q_A) \geq 0$ for all $(i, h) \in \mathcal{B} \times \mathcal{A}$.

the maximal number of different contracts the principal can propose up front is smaller than the size of the agent's type space. The admissible number of ex-ante contracts represents a measure of contract complexity.

Our findings demonstrate that a principal may want to allow discretion when it comes to future contracting, even if the principal has access to a very general reward structure with unconstrained transfers by which to incite agent behavior. In a setting where the principal cannot cover every possible pay-off relevant contingency in an ex-ante contract, ex-post contracting can increase contract fit that is valuable enough to dominate the dynamic inefficiency associated with discretionary contracting.

Many agreements feature endogenously incomplete commitment, even if not always by means of an escape clause. Optimal contract length is a major design issue in regulation and service procurement agreements.[20] Defining appropriate market size thresholds for when to regulate firms directly (ex-ante contracting) or indirectly through competition policy (ex-post contracting) is another arrangement with endogenously incomplete commitment. Those issues require independent analysis which we leave for future research.

References

Akbarpour, Mohammad and Shengwu Li, "Credible auctions: A trilemma," *Econometrica*, 2020, *88* (2), 425–467.

Amador, Manuel and Kyle Bagwell, "The theory of optimal delegation with an application to tariff caps," *Econometrica*, 2013, *81* (4), 1541–1599.

Bagwell, Kyle and Robert W. Staiger, "Enforcement, private political pressure, and the GATT/WTO escape clause," *Journal of Legal Studies*, 2005, *34* (2), 471–513.

Beccuti, Juan and Marc Möller, "Dynamic adverse selection with a patient seller," *Journal of Economic Theory*, 2018, *173*, 95–117.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman, "Mechanisms with evidence: Commitment anf robustness," *Econometrica*, 2019, *87* (2), 529–566.

Bergemann, Dirk and Juuso Välimäki, "Dynamic mechanism design: An introduction," *Journal of Economic Literature*, 2019, *57* (2), 235–274.

Beshkar, Mostafa and Eric W. Bond, "Cap and escape in trade agreements," *American Economic Journal: Microeconomics*, 2017, *9* (4), 171–202.

---

[20]Mechanism design has so far focused on long-term contracts; see Bergemann and Välimäki (2019) for a survey.

Bester, Helmut and Roland Strausz, "Contracting with imperfect commitment and the revelation principle: The single agent case," *Econometrica*, 2001, *69* (4), 1077–1098.

_ **and** _ , "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, 2007, *136* (1), 236–259.

Breig, Zachary, "Repeated contracting without commitment," *Journal of Economic Theory*, 2022, *204*, 105514.

Chu, Leon Yang and David E. M. Sappington, "Simple cost-sharing contracts," *American Economic Review*, 2007, *97* (1), 419–428.

Coate, Stephen and Ross T. Milton, "Optimal fiscal limits with overrides," *Journal of Public Economics*, 2019, *174*, 76–92.

Crawford, Vincent P. and Joel Sobel, "Strategic information transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Doval, Laura and Vasiliki Skreta, "Mechanism design with limited commitment," *Econometrica*, 2022, *90* (4), 1436–1500.

Dye, Ronald A., "Costly contract contingencies," *International Economic Review*, 1985, *26* (1), 233–250.

Freixas, Xavier, Roger Guesnerie, and Jean Tirole, "Planning under incomplete information and the ratchet effect," *Review of Economic Studies*, 1985, *52* (2), 173–191.

Fudenberg, Drew and Jean Tirole, "Sequential bargaining with incomplete information," *Review of Economic Studies*, 1983, *50* (2), 221–247.

Halac, Marina and Pierre Yared, "Fiscal rules and discretion under persistent shocks," *Econometrica*, 2014, *82* (5), 1557–1614.

_ **and** _ , "Commitment versus flexibility with costly verification," *Journal of Political Economy*, 2020, *128* (12), 4523–4573.

Hancart, Nathan, "The (no) value of commitment," 2022. Mimeo, Department of Economics, University College London.

Hart, Oliver D. and Jean Tirole, "Contract renegotiation and Coasian dynamics," *Review of Economic Studies*, 1988, *55* (4), 509–540.

Holmström, Bengt, "On the theory of delegation," in Marcel Boyer and Richard Kihlstrom, eds., *Bayesian Models in Economic Theory*, North-Holland, New York, 1984, pp. 115–141.

Laffont, Jean-Jacques and Jean Tirole, "The dynamics of incentive contracts," *Econometrica*, 1988, *56* (5), 1153–1175.

_ **and** _ , "Adverse selection and renegotiation in procurement," *Review of Economic Studies*, 1990, *57* (4), 597–625.

Maestri, Lucas, "Dynamic contracting under adverse selection and renegotiation," *Journal of Economic Theory*, 2017, *171*, 136–173.

Ofgem, "RIIO-ED2 Final determinations overview document," Decision OFG1163, 2022.

Rogerson, William P., "Simple menus of contracts in cost-based procurement and regulation," *American Economic Review*, 2003, *93* (3), 919–926.

Sappington, David E. M. and Dennis L. Weisman, "Price cap regulation: what have we learned from 25 years of experience in the telecommunication industry," *Journal of Regulatory Economics*, 2010, *38*, 227–257.

Segal, Ilya, "Complexity and renegotiation: A foundation for incomplete contracts," *Review of Economic Studies*, 1999, *66* (1), 57–82.

Shin, Dongsoo and Roland Strausz, "Delegation and dynamic incentives," *RAND Journal of Economics*, 2014, *45* (3), 495–520.

Skreta, Vasiliki, "Sequentially optimal mechanisms," *Review of Economic Studies*, 2006, *73* (4), 1085–1111.

_ , "Optimal auction design under non-commitment," *Journal of Economic Theory*, 2015, *159* (Part B), 854–890.

Tirole, Jean, "From bottom of the barrel to cream of the crop: Sequential screening with positive selection," *Econometrica*, 2016, *84* (4), 1291–1343.

Vartiainen, Hannu, "Auction design without commitment," *Journal of the European Economic Association*, 2013, *11* (2), 316–342.

Weitzman, Martin L., "The "ratchet principle" and performance incentives," *Bell Journal of Economics*, 1980, *11* (1), 302–308.

Appendix

This appendix first establishes four claims about incentive feasible mechanisms in our specific context. It then proves that a standard property of incentive efficient mechanisms holds: The incentive compatibility constraints of ex-ante contracts are locally downward-binding. Appendix A.1 provides a full characterization of incentive efficient mechanisms with complete commitment (pure ex-ante contracting). The rest of this appendix proves the lemmas and propositions as they appear chronologically in the main text.

The first claim reiterates the standard results that local downward 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* $(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$ *satisfies individual rationality (2) and incentive compatibility (3) if and only if the following conditions are all met:*

$$U_I(x_I) \geq 0, \tag{28}$$

$$U_i(x_i) \geq U_i(x_{i+1}) \ \forall i \in \{1, ..., I-1\}, \tag{29}$$

$$U_i(x_i) \geq U_i(x_{i-1}) \ \forall i \in \{2, ..., I\}, \tag{30}$$

$$q_i \geq q_{i+1} \ \forall i \in \{1, ..., I-1\}, \tag{31}$$

Proof. Necessity of (28)-(30) is obvious. Local incentive compatibility implies

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

By rearranging expressions we get

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

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, $j \in \{i+1, ..., I\}$ can be written as

$$U_i(x_i) - U_i(x_j) = \sum_{h=i}^{j-1} [U_h(x_h) - U_h(x_{h+1}) + (\theta_{h+1} - \theta_h)(q_{h+1} - q_j)] \geq 0 \ \forall i \in \{1, ..., I-1\}, \tag{32}$$

where non-negativity follows from the assumptions of local downward incentive compatibility (29) and monotonicity (31). The net benefit of truthfully reporting cost θ_i relative

to understating it to θ_j, $j \in \{1, ..., i-1\}$, equals

$$U_i(x_i)-U_i(x_j) = \sum_{h=j+1}^{i} [U_h(x_h)-U_h(x_{h-1})+(\theta_h-\theta_{h-1})(q_j-q_{h-1})] \geq 0 \; \forall i \in \{2, ..., I\}, \quad (33)$$

where non-negativity follows from the assumptions of local upward incentive compatibility (30) and monotonicity (31).

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

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* $U_h(x_h) > U_h(x_j) \; \forall (h, j) \in \{1, ..., i\} \times \{i+1, ..., I\}$.

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

Proof. By (32), the net benefit of truthfully reporting cost θ_i, relative to exaggerating it to θ_j, $j \in \{i+1, ..., I\}$, satisfies

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

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

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

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

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

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

Proof. By (33), the net benefit of truthfully reporting cost θ_{i+1}, relative to understating it to θ_j, $j \in \{1, ..., i\}$, satisfies

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

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

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

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

Proof. Incentive compatibility (3) implies

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

Rearranging the two expressions yields

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

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

$$U_i(\hat{x}_i) = U_i(\hat{x}_{i+1}) \ \forall i \in \{1, ..., A-1\}, \ A \geq 2. \tag{34}$$

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

$$[U_i(\hat{x}_i) - U_i(\hat{x}_{i+1})][U_{i+1}(\hat{x}_{i+1}) - U_{i+1}(\hat{x}_i)] = 0 \ \forall i \in \{1, ..., \min\{A, I-1\}\}. \tag{35}$$

Suppose both local IC constraints are slack for some $i \in \{1, ..., \min\{A, I-1\}\}$. 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 $(i, j) \in \{1, ..., \min\{A, I-1\}\} \times \{\min\{A+1, I\}, ..., I\}$. 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 $(i, j) \in \{\min\{A+1, I\}, ..., I\} \times \{1, ..., \min\{A, I-1\}\}$. A marginal reduction in the transfer payment $\hat{t}_j$ by a small amount ϵ for all types $j \in \{1, ..., \min\{A; I-1\}\}$ 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 (35) is binding, but the local downward IC constraint is slack for some $i \in \{1, ..., A-1\}$, $A \geq 2$. 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 $(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$ by setting $t_j = \hat{t}_j - \epsilon$, $\epsilon > 0$, for all $j \in \{1, ..., i\}$ 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:

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

If $W_{i+1}(\hat{x}_j) > W_{i+1}(\hat{x}_{i+1})$ for some $j \in \{1, ..., i\}$, 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 $j \in \{1, ..., i\}$. We conclude

that an incentive efficient mechanism satisfies (35) only if $U_i(\hat{x}_i) = U_i(\hat{x}_{i+1})$ for all $i \in \{1, ..., A-1\}$, $A \geq 2$.

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 (35). 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. ∎

A.1 Mechanisms with complete commitment

This appendix contains a full analysis of incentive efficient mechanisms with complete commitment discussed in Section 3. We let $K \geq 1$, but do not necessarily assume that contracting is constrained. Hence, we allow $K \geq I$. The incentive efficient mechanism under complete commitment consists of a partitioning $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_1, ..., \mathcal{A}_k, ..., \mathcal{A}_{\tilde{K}}\}$ of $\mathcal{A} = \{1, ..., A\}$ into $\tilde{K} \leq \min\{K; I\}$ non-empty and convex *cost groups*. Recall that $\theta_{\underline{A}_k}$ is 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, $j \in \mathcal{A}_k$.

Lemma 4 *Assume that the mechanism $(\hat{\mathbf{x}}, \mathbf{I} | \hat{\mathcal{A}}, \emptyset)$ 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 $\hat{\mathcal{A}}_k$ is characterized by (14). If cost group k features pooling, $|\hat{\mathcal{A}}_k| \geq 2$, then output in this cost group is upward- [downward-] distorted relative to the second-best efficient output of the least [most] efficient cost type in $\hat{\mathcal{A}}_k$, $q_{\hat{A}_k}^{sb} < \hat{q}_{\hat{A}_k} < q_{\underline{\hat{A}}_k}^{sb}$*

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

$$W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}}) \geq 0 \geq W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_{k+1}}). \qquad (36)$$

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

$$W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) \geq 0 \geq W_{\hat{A}+1}^{sb}(\hat{q}_{\hat{A}}). \qquad (37)$$

Notice the similarity between (36) and (37) 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.

We first establish Item 1 of Lemma 4. Using Claim 5, we can derive expression (11) for agency rent of any incentive efficient mechanism with complete commitment. The expected welfare function (12) follows from substituting (11) and $\boldsymbol{\Sigma} = \mathbf{I}$ into (8) and applying the following summation by parts

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

The welfare function (12) is additively separable across cost groups, and the objective function (13) is strictly concave in output. Hence, a mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ with transfer payments that yield (11) and where output is characterized by (14) 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 start by demonstrating monotonicity of output. Comparing $\hat{q}_{\hat{A}_k}$ characterized in (14) with $\hat{q}_{\underline{\hat{A}}_k}^{sb}$ characterized in (15) delivers

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

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

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

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

We establish Item 2 by verifying that any incentive efficient mechanism features $W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) \geq W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_{k+1}})$ for all $k \in \{1, ...\hat{K}-1\}$ if the number $\hat{K} \leq \min\{K; I\}$ of non-empty cost groups satisfy $\hat{K} \geq 2$. 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: $\mathcal{A}_k = \hat{\mathcal{A}}_k \backslash \hat{A}_k$ 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 $j \in \{1, ...\hat{A}_k - 1\}$, $x_{\hat{A}_k} = \hat{x}_{\hat{A}_{k+1}}$, and $x_j = \hat{x}_j$ for all $j \in \{\hat{A}_k + 1, ...I\}$. 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 $i \in \{1, ..., \hat{A}_k - 2\}$ if $\hat{A}_k \geq 3$ and for all $i \in \{\hat{A}_k, ..., I-1\}$. Moreover, $U_{\hat{A}_k-1}(x_{\hat{A}_k-1}) = U_{\hat{A}_k-1}(x_{\hat{A}_k})$ 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 $\theta_{\underline{\hat{A}}_{k+1}}$ optimally belongs in cost group $\hat{A}_{k+1}$ only if $W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_{k+1}}) \geq W_{\underline{\hat{A}}_{k+1}}^{sb}(\hat{q}_{\hat{A}_k})$.

One can use the same recipe as above to establish Item 3. $\square$

A.2 Proof of Lemma 1

We prove the Lemma through a sequence of 7 claims. Let $z_j \in \{j, ..., I\}$ be the maximal cost type that reports θ_j with positive probability in the mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$: $\sigma_{jz_j} > 0$ and $\sigma_{ji} = 0$ for all $i \in \{z_j + 1, ..., I\}$ if $z_j \leq I - 1$. The type z_j exists by $\sigma_j > 0$.

Claim 6 *A mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ 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

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

41

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

Recall that $\underline{B}$ is the minimal cost type and B is the maximal cost type in $\mathcal{B}$ in a mechanism with incomplete commitment. We do not assume that the escape clause $\mathcal{B}$ is convex.

Claim 7 *A mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ 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. $x_j = x_{\underline{B}} \ \forall j \in \{\underline{B}, ..., z - 1\}$ if either $\underline{B} \leq B - 1$ or $z \geq B + 1$.

2. $x_j = x_0 \ \forall j \in \{z + 1, ..., I\}$ if $z \leq I - 1$.

Proof. The property $z_j \geq B \ \forall j \in \mathcal{B}$ holds trivially if $B = 1$. Assume that $B \geq 2$ and suppose $z_j < B$ for some $j \in \mathcal{B}$. Then $U_{z_j}(x_B) = U_B(x_B) + (\theta_B - \theta_{z_j})q_B > 0$ by $U_B(x_B) \geq 0$, $\theta_B > \theta_{z_j}$ and $q_B > 0$. By $\sigma_{jz_j} > 0$ and (4), it follows that $U_{z_j}(x_{z_j}) = U_{z_j}(x_j) = t_j - \theta_{z_j}q_j = 0$. $U_{z_j}(x_B) > U_{z_j}(x_{z_j})$ then follows, which is a violation of (3). We conclude that $z_j \geq B \ \forall j \in \mathcal{B}$. Suppose $z_j < z_h$ for some $(j, h) \in \mathcal{B} \times \mathcal{B}$. In this case, $U_{z_j}(x_{z_j}) = 0 < (\theta_{z_h} - \theta_{z_j})q_h = U_{z_j}(x_h)$, which again violates incentive compatibility. Hence, $z_j = z \geq B$ for all $j \in \mathcal{B}$.

Consider Item 1 of the claim. $\underline{B} \leq z - 1$ by the assumption of the claim. By the incentive compatibility constraint (3),

$$U_z(x_{z-1}) = U_{z-1}(x_{z-1}) - (\theta_z - \theta_{z-1})q_{z-1} \leq U_z(x_z) = 0.$$

Invoking incentive compatibility (3) again, plus individual rationality (2) and Claim 6 yields

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

Combining these two inequalities delivers

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

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 $q_j = q_{z-1} = q_{\underline{B}}$ for all $j \in \{\underline{B}, ..., z - 1\}$. We can now invoke Claim 4 to obtain $x_j = x_{\underline{B}}$ for all $j \in \{\underline{B}, ..., z - 1\}$.

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 $j \in \{z + 1, ..., I\}$. 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 $j \in \{z + 1, ..., I\}$. ∎

Claim 8 *A mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}, \mathcal{B})$ with incomplete commitment $(\mathcal{B} \neq \emptyset)$ is incentive feasible only if $|\mathbf{x}_{\mathcal{B}}| \in \{1, 2\}$. Incentive feasibility implies $x_j = x_{\underline{B}} \; \forall j \in \{\underline{B}, ..., B - 1\}$ if $\underline{B} \leq B - 1$.*

Proof. We prove the claim in reverse order. Let $\underline{B} \leq B - 1$. By the previous claim, $\underline{B} \leq B - 1 \leq z - 1$ and then all contracts x_j, $j \in \{\underline{B}, ..., B - 1\}$ are identical and equal to $x_{\underline{B}}$. Seeing as $\mathcal{B} \subseteq \{\underline{B}, ..., B\}$, $|\mathbf{x}_{\mathcal{B}}| \in \{1, 2\}$ if $\underline{B} \leq B - 1$. Tivially, $|\mathbf{x}_{\mathcal{B}}| = 1$ if $\underline{B} = B$. ∎

Claim 8 establishes Item 1 of Lemma 1. Consider Item 2. $|\mathbf{x}_B| = 1$ if $z \geq B + 1$ by Claim 7. Hence, $|\mathbf{x}_B| = 2$ implies $z = B$. Claims 6, 7 and $z = B$ then imply $x_j = (q_j, \theta_B q_j)$ for all $j \in \mathcal{B}$. Invoking Claim 8 yields $x_j = x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}})$ for all $j \in \{\underline{B}, ..., B - 1\}$ if $|\mathbf{x}_{\mathcal{B}}| = 2$. Next:

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

The first (weak) inequality follows from incentive compatibility, the second (strict) inequality from $q_{\underline{B}} \neq q_B$ by $x_{\underline{B}} \neq x_B$ and monotonicity of output. Furthermore,

$$U_i(x_i) = U_i(x_0) = 0 > -(\theta_i - \theta_B) q_B = U_i(x_B) \; \forall i \in \{B + 1, ...I\}, \; B \leq I - 1.$$

The first string of equalities follow from $z = B$ for $|\mathbf{x}_{\mathcal{B}}| = 2$ and Claim 7. $U_i(x_i) > U_i(x_B)$ for all $i \neq B$ implies $\sigma_{Bi} = 0$ for all $i \neq B$ by (4). Hence, $\mu_{BB} = 1$ 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 $w_B^{fb} > 0$. This completes the proof of Item 2 of Lemma 1.

To prove items 3-5, we characterize additional properties of incentive optimal mechanisms with incomplete commitment.

Claim 9 *A mechanism $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ 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 $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}, \mathcal{B})$ as follows: $\mathcal{A}_l = \mathcal{A}_l^*$ for all $l \leq K^*$ if $K^* \geq 1$. If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then $\mathcal{A}_{K^*+1} = \mathcal{B}^* \backslash B^*$ 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 $\mathcal{B} \subset \mathcal{B}^* \cup \emptyset$, $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ 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 $j \in A^*_k$. Then $x^*_{\mathcal{A}^*_k} \in \{x^*_{\underline{B}^*}, x^*_{B^*}\}$ by Claim 4. Construct a modified mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}, \mathcal{B})$ 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 $\mathcal{A}_k = \mathcal{A}^*_k \cup \mathcal{B}^*\backslash B^*$ and $\mathcal{B} = B^*$. If $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ and $x^*_{\mathcal{A}^*_k} = x^*_{B^*}$, then $\mathcal{A}_k = \mathcal{A}^*_k \cup B^*$ and $\mathcal{B} = \mathcal{B}^*\backslash B^*$. If $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$, then $\mathcal{A}_k = \mathcal{A}^*_k \cup \mathcal{B}^*$ 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 9 is that $\mathcal{B}^* = \{\underline{B}^*, ..., B^*\}$. This property holds trivially if either $\underline{B}^* = B^* - 1$ or $\underline{B}^* = B^*$. If $\underline{B}^* \leq B^* - 2$ and $j \in \mathcal{A}^*$ for some $\underline{B}^* < j < B^*$, then $q^*_j = q^*_{\underline{B}^*}$ by Claim 7, which violates Claim 9.

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 $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment ($\mathcal{B}^* \neq \emptyset$).

Claim 10 *Assume that the incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ 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 $z^* \leq B^* + 1$. 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 $j \in \{\underline{B}^*, ..., z^* - 1\}$ by Claim 7. In particular, $q^*_{B^*+1} = q^*_{\underline{B}^*}$, which violates the necessary condition of incentive optimality established in Claim 9. 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. ■

Claim 11 *A mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ 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 $x^*_{B^*+1} = x^{fb}_{B^*+1}$ 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 10 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 9. If $q^*_{B^*+1} \in (0, q^*_{B^*})$, then $z^* = B^* + 1$ by Claim 10. 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 $j \in \{B^* + 2, ..., I\}$ 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 $(i,j) \in \{1,...,B^*\} \times \{B^*+1,...,I\}$. If $B^* \leq I-2$, then upward-binding IC and strict monotonicity also imply $\sigma^*_{ji} = 0$ for all $(i,j) \in \{B^*+2,...,I\} \times \{1,...,B^*\}$ 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 $i \in \{B^*+2,...,I\}$. In particular, $U_i(x^*_i) > U_i(x^*_{B^*+1})$ 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 $x^*_{B^*+1} = x^{fb}_{B^*+1}$. Finally, $\sigma^*_{j(B^*+1)} = 0$ for all $j \in \{1,...,\underline{B}^*-1\}$ if $\underline{B}^* \geq 2$ by $U_{\underline{B}^*}(x^*_{\underline{B}^*}) \geq U_{\underline{B}^*}(x^*_{\underline{B}^*-1})$, $q^*_{\underline{B}^*-1} > q^*_{\underline{B}^*}$ and Claim 3. Based on this information, we can write the principal's expected surplus of the proposed incentive optimal mechanism as:

$$W(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{j=1}^{B^*} \sum_{i=1}^{B^*} \nu_i \sigma^*_{ji} W_i(x^*_j) + \nu_{B^*+1} \Big[\sum_{j \in B^*} \sigma^*_{j(B^*+1)} W_{B^*+1}(x^*_{\underline{B}^*}) + \sigma^*_{B^*+1} w^{fb}_{B^*+1} \Big].$$

Consider the alternative mechanism $(\mathbf{x}^*, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$, where $\mathcal{A}_l = \mathcal{A}^*_l$ for all $l \in \{1,...,K-1\}$, 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 $x_{B^*+1} = x^{fb}_{B^*+1} = x^*_{B^*+1}$ 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

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

The difference

$$W(\mathbf{x}^*, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B}) - W(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*) = \nu_{B^*+1} \Big(1 - \sigma^*_{B^*+1} - \sum_{j \in \mathcal{B}^*} \sigma^*_{j(B^*+1)} \Big) w^{fb}_{B^*+1}$$
$$+ \nu_{B^*+1} \sum_{j \in \mathcal{B}^*} \sigma^*_{j(B^*+1)} [w^{fb}_{B^*+1} - W_{B^*+1}(x^*_{\underline{B}^*})]$$

in expected surplus between the two mechanisms is strictly positive by $x^*_{\underline{B}^*} \neq x^*_{B^*+1} = x^{fb}_{B^*+1}$ and because $z^* = B^*+1$ it follows that $\sum_{j \in \mathcal{B}^*} \sigma^*_{j(B^*+1)} > 0$. 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 $x^*_{B^*+1} = x^{fb}_{B^*+1}$. In particular, the expected surplus of the proposed incentive optimal

mechanism is:

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

Consider a modified mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B})$, where $\mathcal{B} = \{\underline{B}^*, ..., B^*+1\}$, $x_{B^*+1} = x_{B^*+1}^{fb}$ and $\sigma_i = \sigma_i^* + \sigma_{(B^*+1)i}^*$ for all $i \in \{B^*+2...., I\}$ 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 $(i,j) \in \{B^*+2,...I\} \times \{1,...,B^*\}$ 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

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

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

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

The difference in expected surplus between the two mechanisms is

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

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

$q^*_{B^*+1} = 0$ implies $z^* \neq B^* + 1$. $\blacksquare$

We can now draw conclusions about incentive optimal mechanisms with incomplete commitment. Claim 9 proves Item 3 of Lemma 1. By way of $z^* = B^*$ and Item 2 of Claim 7, $\mathcal{C}^* = \{B^* + 1, ..., I\}$ if $B^* \leq I - 1$. This proves Item 5. Moreover, $\mathcal{B}^* \cup \mathcal{C}^* = \{\underline{B}^*, ..., I\}$. 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 $\mathcal{A}^* = \{1, ..., A^*\}$, where $A^* = \underline{B}^* - 1$. This proves the first part of Item 4 of Lemma 1. Item 4(a) follows from Claim 9 and monotonicity. Item 4(b) follows from Claim 6, Claim 7 and $z^* = B$. Obviously, $U_{B^*}(x^*_{\underline{B}^*}) = U_{B^*}(x^*_{B^*}) = 0$. $\square$

A.3 Proof of Lemma 2

We first demonstrate some general properties of $\mathbf{\Sigma}^*$ in incentive optimal mechanisms $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ with incomplete commitment ($\mathcal{B}^* \neq \emptyset$). This is done in 6 claims. Claim 12 and Claim 13 establish combinations of cost types and cost reports that do not occur in an incentive optimal mechanism.

Claim 12 *Consider a mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ that features ex-ante contracting ($\mathcal{A}^* \neq \emptyset$) and incomplete commitment ($\mathcal{B}^* \neq \emptyset$). This mechanism is incentive optimal only if the following conditions are all met:*

1. $\sigma^*_{ji} = 0 \; \forall (i, j) \in \{1, ...A^* - 1\} \times \{\underline{B}^*, ..., I\}$ if $A^* \geq 2$.

2. $\sigma^*_{ji} = 0 \; \forall (i, j) \in \{1, ..., B^* - 1\} \times \{B^*, ...I\}$ if $q^*_{\underline{B}^*} > q^*_{B^*}$.

3. $\sigma^*_{ji} = 0 \; \forall (i, j) \in \{1, ..., B^* - 1\} \times \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$.

4. $\sigma^*_{ji} = 0 \; \forall (i, j) \in \{\underline{B}^* + 1, ..., I\} \times \mathcal{A}^*$ if $\underline{B}^* \leq I - 1$.

5. $\sigma^*_{j\underline{B}^*} = 0 \; \forall j \in \{1, ..., A^* - 1\}$ 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, we obtain:

$$U_i(x^*_i) - U_i(x^*_j) = U_i(x^*_i) - U_i(x^*_h) + U_h(x^*_h) - U_h(x^*_j) + (\theta_h - \theta_i)(q^*_h - q^*_j) \geq (\theta_h - \theta_i)(q^*_h - q^*_j). \tag{38}$$

Hence, $\sigma^*_{ji} = 0$ 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 (38) is strictly positive for all $(i, j) \in \{1, ..., A^* - 1\} \times \{\underline{B}^*, ..., I\}$ by $q^*_{A^*} > q^*_{\underline{B}^*} \geq q^*_j$ for all $j \in \{\underline{B}^*, ..., I\}$; see Lemma 1.

Item 2: If $h = B^* - 1$, then the rightmost expression in (38) is strictly positive for all $(i, j) \in \{1, ..., B^* - 2\} \times \{B^*, ...I\}$ by $q^*_{B^*-1} > q^*_{B^*} \geq q^*_j$ for all $j \in \{B^*, ...I\}$.

47

Item 3: If $h = B^*$, then the rightmost expression in (38) is strictly positive for all $(i,j) \in \{1, ..., B^* - 1\} \times C^*$ by $q_{B^*}^* > 0$.

Item 4: If $h = \underline{B}^*$, then the rightmost expression in (38) is strictly positive for all $(i,j) \in \{\underline{B}^* + 1, ..., I\} \times \mathcal{A}^*$ 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 (38) is strictly positive for all $j \in \{1, ..., A^* - 1\}$ that satisfy $q_j^* > q_{A^*}^*$.

Item 6: If $h = B^*$, then the rightmost expression in (38) is strictly positive for all $(i,j) \in C^* \times \mathcal{A}^*$ 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 $(i,j) \in C^* \times \mathcal{B}^*$ completes the proof. ∎

Claim 13 *A mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ that features ex-ante contracting ($\mathcal{A}^* \neq \emptyset$), incomplete commitment ($\mathcal{B}^* \neq \emptyset$) and partial participation ($C^* \neq \emptyset$), is incentive optimal only if $\sum_{j \in C^*} \sigma_{jB^*}^* = 0$.*

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

In case one, $q_{\underline{B}^*}^* > q_{B^*}^*$. By Claim 12, $\sigma_{I_{B^*}i}^* = 0$ for all $i \neq B^*$. Upon observing θ_{B^*}, the principal therefore deduces that the agent with probability one has 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

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

Let a modified mechanism $(\mathbf{x}^*, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ 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 $\mathbf{\Sigma}^*$ to $\mathbf{\Sigma}$ 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

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

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

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 from the main text the definition $\underline{\mathcal{B}}^* = \{\underline{B}^*, ..., B^* - 1\}$ 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 $h \in \{A^*, ..., B^* - 1\}$ indifferent between accepting or rejecting the ex-post contract, equals $\frac{\Omega_{jh}(\boldsymbol{\sigma}_{jh}^*)}{\sum_{i=A^*}^{B^*} \nu_i \sigma_{ji}^*}$, where

$$\Omega_{jh}(\boldsymbol{\sigma}_{jh}) = \sum_{i=A^*}^{h} \nu_i \sigma_{ji} W_i(x_{jh}^d) = \sum_{i=A^*}^{h} \nu_i \sigma_{ji}[S(q_{jh}^d) - (\theta_i + (1-\alpha)(\theta_h - \theta_i))q_{jh}^d], \qquad (39)$$

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

$$S'(q_{jh}^d) = \frac{\sum_{i=A^*}^{h} \nu_i \sigma_{ji}(\theta_i + (1-\alpha)(\theta_h - \theta_i))}{\sum_{i=A^*}^{h} \nu_i \sigma_{ji}}, \qquad (40)$$

characterizes the optimal output given the reporting strategy $\boldsymbol{\sigma}_{jh} = (\sigma_{jA^*}, ..., \sigma_{jh})$. The contract x_{jh}^{*d} results from replacing $\boldsymbol{\sigma}_{jh}$ by $\boldsymbol{\sigma}_{jh}^*$ in (39) and (40). The Stage 4 expected surplus of offering $x_{\underline{B}^*}^*$ subsequent to a cost report θ_j, $j \in \underline{\mathcal{B}}^*$, 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

$$\Omega_{jB^*}^* \geq \Omega_{jh}(\boldsymbol{\sigma}_{jh}^*) \; \forall (j, h) \in \underline{\mathcal{B}}^* \times \{A^*, ..., B^* - 1\}. \qquad (41)$$

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.

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

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

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

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

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

$$\Omega_{jB^*}(\boldsymbol{\sigma}_{jB^*}) \geq \Omega_{jh}(\boldsymbol{\sigma}_{jh}^*) \ \forall (j,h) \in \mathcal{B}^* \times \{A^*, ..., B^* - 1\}. \tag{42}$$

A marginal increase in ϵ has no effect on the right-hand side of (42). 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

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

Since $\frac{\partial \Omega_{jB^*}}{\partial \epsilon} > \nu_{B^*} \frac{\partial \sigma_{jB^*}}{\partial \epsilon} W_{B^*}(x_{\underline{B}^*}^*) \geq 0$ for all $\epsilon > 0$, 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

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

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

Claim 14 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 14 *A mechanism* $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ *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 $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ exists if $\sigma_{B^*}^* < 1$. Item 1 of Claim 12 implies $\sum_{j=1}^{A^*-1} \sigma_{jB^*}^* = 0$ if $A^* \geq 2$. Claim 13 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

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

50

Consider a modified mechanism $(\mathbf{x}, \boldsymbol{\Sigma} | \mathcal{A}^*, \mathcal{B}^*)$ 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^*, $k \in \{1, ..., K\}$, and by $\sigma_{B^*} = 1$. Everything else is the same as in the original mechanism. This mechanism is incentive feasible and has expected surplus

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

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

Claim 15 applies the revelation principle to the menu of *ex-ante* contracts and invokes the three previous claims.

Claim 15 *For any incentive optimal mechanism $(\mathbf{x}^*, \boldsymbol{\Sigma}^{**} | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment $(\mathcal{B}^* \neq \emptyset)$, there exists an incentive optimal mechanism $(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ where the reporting strategy $\boldsymbol{\Sigma}^*$ has the following properties:*

1. *$\sigma_i^* = 1 \ \forall i \in \{1, ..., A^* - 1\}$ 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. *$\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji}^* = 1$, $i \in \{\underline{B}^* + 1, ..., B^* - 1\}$ if $\underline{B}^* \leq B^* - 2$ and $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$.*

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

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

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

$$
\begin{aligned}
W(\mathbf{x}^*, \boldsymbol{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*) &- W(\mathbf{x}^*, \boldsymbol{\Sigma}^{**} | \mathcal{A}^*, \mathcal{B}^*) \\
&= \sum_{i=1}^{B^*} \sum_{j=1}^{A^*} \nu_i (\sigma^*_{ji} - \sigma^{**}_{ji}) W_i(x^*_j) \\
&= \sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i (\sigma^*_{ji} - \sigma^{**}_{ji}) W_i(x^*_j) \\
&= \sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i (\sigma^*_{ji} - \sigma^{**}_{ji}) [W_i(x^*_j) - W_i(x^*_i) + W_i(x^*_i)] \\
&= \sum_{i=1}^{A^*} \nu_i [\sum_{j=1}^{A^*} \sigma^*_{ji} W_i(x^*_j) - \sum_{j=1}^{A^*} \sigma^{**}_{ji} W_i(x^*_i)] \\
&= \sum_{i=1}^{A^*} \nu_i [\sigma^*_i - \sum_{j=1}^{A^*} \sigma^{**}_{ji}] W_i(x^*_i) = 0.
\end{aligned} \tag{43}
$$

The third row follows from

$$
\begin{aligned}
\sum_{j=1}^{A^*} \nu_{\underline{B}^*} (\sigma^*_{j\underline{B}^*} - \sigma^{**}_{j\underline{B}^*}) W_{\underline{B}^*}(x^*_j) \\
= \sum_{j=1}^{A^*} \nu_{\underline{B}^*} (\sigma^*_{j\underline{B}^*} - \sigma^{**}_{j\underline{B}^*}) W_{\underline{B}^*}(x^*_{A^*}) = 0
\end{aligned}
$$

where

$$
\sum_{j=1}^{A^*} \sigma^*_{j\underline{B}^*} = \sigma^*_{A^*\underline{B}^*} = \sum_{j=1}^{A^*} \sigma^{**}_{j\underline{B}^*}
$$

by construction of $\boldsymbol{\Sigma}^*$. In the fourth row, we have added and subtracted $W_i(x^*_i)$ inside the square brackets. The fifth row follows from:

$$
\sigma^{**}_{ji}(W_i(x^*_i) - W_i(x^*_j)) = 0 \ \forall (i, j) \in \mathcal{A}^* \times \mathcal{A}^*.
$$

This property holds for $\sigma^{**}_{ji} > 0$ because then $W_i(x^*_i) = W_i(x^*_j)$. For $W_i(x^*_i) > W_i(x^*_j)$ 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 (43) follows from $\sigma^*_{ji} = 0 \ \forall (i, j) \in \mathcal{A}^* \times \mathcal{A}^*, \ i \neq j$. The second equality is implied by $\sigma^*_i = 1 = \sum_{j=1}^{A^*} \sigma^{**}_{ji} \ \forall i \in \{1, ..., A^* - 1\}$ if $A^* \geq 2$, and $\sigma^*_{A^*} = \sum_{j=1}^{A^*} \sigma^{**}_{jA^*}$ by construction of $\boldsymbol{\Sigma}^*$.

Item 1 Follows directly from the construction of $\boldsymbol{\Sigma}^*$.

Item 2 By construction of $\boldsymbol{\Sigma}^*$, $\sum_{j=1}^{A^*-1} \sigma^*_{ji} = 0$, $i \in \{A^*, \underline{B}^*\}$, if $A^* \geq 2$. If $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$, then $q^*_{\underline{B}^*} > q^*_{B^*}$, and we can apply Item 2 of Claim 12 to get $\sum_{j=B^*}^{I} \sigma^*_{ji} = 0$, $i \in \{A^*, \underline{B}^*\}$.

Item 3 From the proof of the previous item, we have $\sum_{j=A^*}^{I} \sigma^*_{ji} = 1$, $i \in \{A^*, \underline{B}^*\}$. The result then trivially follows if $B^* = I$. Let $B^* \leq I - 1$, so that $\mathcal{C} = \{B^* + 1, ..., I\}$. We can then apply Item 3 of Claim 12 to obtain $\sum_{j=B^*+1}^{I} \sigma^*_{jA^*} = 0$ and also $\sum_{j=B^*+1}^{I} \sigma^*_{j\underline{B}^*} = 0$ if $\underline{B}^* \leq B^* - 1$. We can finally apply Claim 13 to obtain $\sum_{j=B^*+1}^{I} \sigma^*_{j\underline{B}^*} = 0$ if $\underline{B}^* = B^*$.

Item 4 Observe that $\sigma^*_{A^*\underline{B}^*} > 0$ only if $U_{\underline{B}^*}(x^*_{\underline{B}^*}) = U_{\underline{B}^*}(x^*_{A^*})$. But then $U_{A^*}(x^*_{A^*}) > U_{A^*}(x^*_{\underline{B}^*})$ by $q^*_{A^*} > q^*_{\underline{B}^*}$. From Claim 2, we then obtain $U_{A^*}(x^*_{A^*}) > U_{A^*}(x^*_j)$, and therefore

$\sigma^{**}_{jA^*} = 0$, for all $j \in \{\underline{B}^*, ..., I\}$. Hence, $1 = \sum^I_{j=1} \sigma^{**}_{jA^*} = \sum^{A^*}_{j=1} \sigma^{**}_{jA^*} = \sigma^*_{A^*}$.

Item 5 Assume that $\underline{B}^* \leq B^* - 2$. From Item 4 of Claim 12, we get $\sum^{A^*}_{j=1} \sigma^*_{ji} = 0$ for all $i \in \{\underline{B}^* + 1, ..., B^* - 1\}$. $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ implies $q^*_{\underline{B}^*} > q^*_{B^*}$, and we can invoke Item 2 of Claim 12 to get $\sum^I_{j=B^*} \sigma^*_{ji} = 0$ for all $i \in \{\underline{B}^* + 1, ..., B^* - 1\}$.

Item 6 Follows directly from Item 3 and Item 4 of Claim 12.

Item 7 If $\underline{B}^* \leq B^* - 1$, then Item 4 of Claim 12 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 13. If $\underline{B}^* = B^*$, then the result follows directly from Claim 14.

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

Item 9 Follows directly from the construction of $\mathbf{\Sigma}^*$. ∎

Claim 15 is close to Lemma 2, 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, $i \in \{A^*, ..., B^*\}$, uses for cost reports θ_j, $j \in \underline{B}^*$. We next establish incentive optimality of uniform randomization strategies.

Claim 16 *For any incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^{**} | \mathcal{A}^*, \mathcal{B}^*)$ that features incomplete commitment $(\mathcal{B}^* \neq \emptyset)$, there exists an incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ in which the reporting strategy $\mathbf{\Sigma}^*$ has the following properties:*

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

2. $\sigma^*_{jB^*} = \frac{1 - \sigma^*_{B^*}}{|\underline{\mathcal{B}}^*|} \; \forall j \in \underline{\mathcal{B}}^*$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$,

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

Proof. Using the results in Claim 15, we can write the principals' expected surplus as

$$\sum^{B^*}_{i=1} \nu_i W_i(x^*_i) + \nu_{A^*}(1 - \sigma^{**}_{A^*})[W_{A^*}(x^*_{\underline{B}^*}) - W_{A^*}(x^*_{A^*})] + \nu_{\underline{B}^*}\sigma^{**}_{A^*\underline{B}^*} \tag{44}$$
$$\times [W_{\underline{B}^*}(x^*_{A^*}) - W_{\underline{B}^*}(x^*_{\underline{B}^*})] + \nu_{B^*}(1 - \sigma^{**}_{B^*})[W_{B^*}(x^*_{\underline{B}^*}) - W_{B^*}(x^*_{B^*})]$$

in the incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^{**} | A^*, B^*)$. The expected surplus depends on $\mathbf{\Sigma}^{**}$ only through $(\sigma^{**}_{A^*}, \sigma^{**}_{A^*\underline{B}^*}, \sigma^{**}_{B^*})$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ and $(\sigma^{**}_{A^*}, \sigma^{**}_{A^*\underline{B}^*})$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$. Let $(\sigma^*_{A^*}, \sigma^*_{A^*\underline{B}^*}, \sigma^*_{B^*}) = (\sigma^{**}_{A^*}, \sigma^{**}_{A^*\underline{B}^*}, \sigma^{**}_{B^*})$. Then $\mathbf{\Sigma}^*$ only modifies cost reports θ_j, $j \in \underline{\mathcal{B}}^*$. Everything else is the same as in the original mechanism. Therefore, the mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ also yields expected surplus (44). To close the proof, we demonstrate sequential rationality of $x^*_{\underline{B}^*}$ in the modified mechanism. Recall $x^*_{\underline{B}^*} = x^*_{B^*}$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$.

By way of the uniform distribution of reporting strategies in $\mathbf{\Sigma}^*$, 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 $h \in \{A^*, ..., B^* - 1\}$ 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

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

In the above expressions, $\sigma_{A^*i}^{**} = 0$ for all $i \in \{\underline{B}^* + 1, ..., h\}$ if $h \geq \underline{B}^* + 1$. 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 $(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*)$ is incentive feasible if and only if the following sequential rationality constraint is met:

$$\Omega_{B^*}^* \geq \Omega_h \;\; \forall h \in \{A^*, ..., B^* - 1\}. \tag{45}$$

We now show that sequential rationality (41) of $x_{\underline{B}^*}^*$ in the original mechanism implies sequential rationality (45) of $x_{\underline{B}^*}^*$ in the modified mechanism. Summing up (41) over all $j \in \underline{\mathcal{B}}^*$ gives:

$$\Omega_{B^*}^* \geq \sum_{j \in \underline{\mathcal{B}}^*} \Omega_{jh}(\boldsymbol{\sigma}_{jh}^{**}) = \bar{\Omega}_h(\mathbf{\Sigma}_h^{**}) \;\; \forall h \in \{A^*, ..., B^* - 1\}. \tag{46}$$

If $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$, then $\mathbf{\Sigma}_h^{**}$ is a $(B^* - \underline{B}^*) \times (h + 1 - A^*)$ matrix that identifies how each of the cost types $i \in \{A^*, ..., h\}$ randomizes across cost reports θ_j, $j \in \{\underline{B}^*, ..., B^* - 1\}$. Instead, $\mathbf{\Sigma}_h^{**}$ 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 $\bar{\Omega}_h(\boldsymbol{\Sigma}_h^{**}) \geq \Omega_h$.

Consider the problem of minimizing $\bar{\Omega}_h(\boldsymbol{\Sigma}_h)$ over $\boldsymbol{\Sigma}_h$ subject to $0 \leq \sigma_{ji} \leq 1$ for all $\sigma_{ji} \in \boldsymbol{\Sigma}_h$, $\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji} = 1 - \sigma_{A^*i}^{**}$ for all $i \in \{A^*, ..., h\}$ if $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$ and $\sum_{j=\underline{B}^*}^{B^*} \sigma_{ji} = 1 - \sigma_{A^*i}^{**}$ for all $i \in \{A^*, ..., h\}$ 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 $\boldsymbol{y}_{jh} = (y_{jA^*h}....,y_{jhh})^T$, then we can write the Hessian matrix $\mathbf{H}_{jh}$ of $\Omega_{jh}(\boldsymbol{\sigma}_{jh})$ as:

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

By implication:

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

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

$$\nu_i W_i(x_{jh}^d) - \lambda_{ih} - \underline{\xi}_{jih} + \xi_{jih} = 0, \tag{47}$$

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

$$\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji} = 1 - \sigma_{A^*i}^{**} \text{ if } |\mathbf{x}_{\mathcal{B}^*}^*| = 2; \quad \sum_{j=\underline{B}^*}^{B^*} \sigma_{ji} = 1 - \sigma_{A^*i}^{**} \text{ if } |\mathbf{x}_{\mathcal{B}^*}^*| = 1, \tag{48}$$

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

slackness conditions

$$\sigma_{ji} \in [0,1], \ \underline{\xi}_{jih} \geq 0, \ \xi_{jih} \geq 0, \ \sigma_{ji}\underline{\xi}_{jih} = (1 - \sigma_{ji})\xi_{jih} = 0 \tag{49}$$

minimize $\bar{\Omega}_h(\boldsymbol{\Sigma}_h)$. In the first-order condition (47), λ_{ih} represents the Lagrangian multiplier on the equality constraint (48), $\underline{\xi}_{jih}$ 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, $\boldsymbol{\Sigma}_h^*$, $\underline{\xi}_{ji} = \xi_{ji} = 0$ and $\lambda_{ih} = \nu_i W_i(x_h^d)$ jointly solve (47)-(49). Hence, $\Omega_{B^*}^* \geq \bar{\Omega}_h(\boldsymbol{\Sigma}_h^{**}) \geq \bar{\Omega}_h(\boldsymbol{\Sigma}_h^*) = \Omega_h$ for all $h \in \{A^*, ..., B^* - 1\}$. ∎

To close the proof of Lemma 2, we need a final result.

Claim 17 *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 $\mathcal{B} \neq \mathcal{B}^*$. The original mechanism is not minimal and therefore cannot be incentive optimal.

By way of Item 4 in Claim 12, 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 15, 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 15. From Claim 16, we apply uniform randomization to derive the posterior probability distribution

$$\mu_{j\underline{B}^*}^* = \frac{\nu_{\underline{B}^*}(1 - \sigma_{A^*\underline{B}^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* - \nu_{B^*}\sigma_{B^*}^*}$$

$$\mu_{ji}^* = \frac{\nu_i}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^* - \nu_{B^*}\sigma_{B^*}^*} \forall i \in \{\underline{B}^* + 1, ..., B^* - 1\}$$

$$\mu_{jB^*}^* = \frac{\nu_{B^*}(1 - \sigma_{B^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^{**} - \nu_{B^*}\sigma_{B^*}^*}$$

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

$$\mu_{j\underline{B}^*}^* = \frac{\nu_{\underline{B}^*}(1 - \sigma_{A^*\underline{B}^*}^*)}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^*}$$

$$\mu_{ji}^* = \frac{\nu_i}{\sum_{i \in \mathcal{B}^*} \nu_i - \nu_{\underline{B}^*}\sigma_{A^*\underline{B}^*}^*} \forall i \in \{\underline{B}^* + 1, ..., B^*\}$$

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

$$W(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*) = \sum_{i=1}^{B^*} \nu_i W_i(x_i^*) + \nu_{\underline{B}^*} \sigma^*_{A^* \underline{B}_*}[W_{\underline{B}^*}(x^*_{A^*}) - W_{\underline{B}^*}(x^*_{\underline{B}^*})]$$

$$+ \nu_{B^*}(1 - \sigma^*_{B^*})[W_{B^*}(x^*_{\underline{B}^*}) - W_{B^*}(x^*_{B^*})].$$

Consider the modified mechanism $(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$ in which $\mathcal{A}_k = \mathcal{A}_k^*$ for all $k \in \{1, ...K-1\}$ 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 $\underline{B} = \underline{B}^* + 1$. 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 $|\mathbf{x}^*_{\underline{\mathcal{B}}^*}| = 2$, then $\sigma_{j\underline{B}^*} = \frac{1-\sigma_{\underline{B}^*}}{|\underline{\mathcal{B}}^*|-1}$, $\sigma_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|-1}$, $i \in \{\underline{B}^* + 1, ..., B^* - 1\}$, and $\sigma^*_{jB^*} = \frac{1-\sigma^*_{B^*}}{|\underline{\mathcal{B}}^*|-1}$ for all $j \in \{\underline{B}^* + 1, ..., B^* - 1\}$, whereas $\sigma_{B^*} = \sigma^*_{B^*}$. If $|\mathbf{x}^*_{\underline{\mathcal{B}}^*}| = 1$, then $\sigma_{j\underline{B}^*} = \frac{1-\sigma_{\underline{B}^*}}{|\underline{\mathcal{B}}^*|-1}$, $\sigma_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|-1}$, $i \in \{\underline{B}^* + 1, ..., B^*\}$, for all $j \in \{\underline{B}^* + 1, ..., B^*\}$. Observe in particular that $\sigma_{ij} = 0$ for all $(i,j) \in \mathcal{B} \times \mathcal{A}$ by construction. Moreover, $W(\mathbf{x}^*, \mathbf{\Sigma}^* | \mathcal{A}^*, \mathcal{B}^*) = W(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}, \mathcal{B})$. The uniform distribution of cost reports $\mathbf{\Sigma}$ yields posterior beliefs $\mu_{ji} = \mu^*_{ji}$ for all $(i,j) \in \mathcal{I} \times \{\underline{B}^* + 1, ..., B^*\}$. Hence, the modified mechanism is incentive feasible by sequential rationality of $\mathbf{x}^*_{\mathcal{B}^*}$. ∎

Claims 14-17 map into the items of Lemma 2 as follows. Item 1 of Claim 15 implies Item 1(a). Item 1 and Item 3 of Claim 16 imply Item 1(b). Claim 14 implies Item 2(a)i. Item 2 and Item 4 of Claim 15 imply Item 2(a)ii. Item 1 and Item 3 of Claim 16 and Claim 17 imply Item 2(a)iii. Item 1 of Claim 16 and Claim 17 imply Item 2(b). Item 8 of Claim 15 and Item 2 of Claim 16 imply Item 2(c)i. Item 7 of Claim 15 and Item 3 of Claim 16 imply Item 2(c)ii. Item 9 of Claim 15 implies Item 3. □

A.4 Proof of Proposition 1

Consider first the simpler case with $|\mathbf{x}^*_{\underline{\mathcal{B}}^*}| = 1$. 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 2 then yields exactly the same distribution of posterior beliefs over the agent's marginal cost as in (19) 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 following sequence of events can therefore be sustained as a PBE: The principal commits to a mechanism consisting of the menu $\mathbf{x}^*_{\mathcal{A}^*}$ of contracts, augmented by the vague escape clause (VC). The agent then selects x_i^* if $|\mathcal{A}^*| \geq 2$ and the agent has marginal cost θ_i, $i \in \{1, ..., A^* - 1\}$. The agent selects $x^*_{A^*}$ with probability $\sigma^*_{A^*}$ and activates the escape clause VC with probability $1 - \sigma^*_{A^*}$ if it has marginal cost θ_{A^*}. The agent activates the escape clause VC with probability 1 if the agent has marginal cost θ_i, $i \in \mathcal{B}^*$. The agent

rejects the contract offer if $\mathcal{C}^* \neq \emptyset$ and the agent has marginal cost θ_i, $i \in \mathcal{C}^*$. The principal offers the ex-post contract $x_{\underline{B}^*}^*$ in stage 3 if the agent has activated the escape clause in stage 2. This restricted communication mechanism augmented by a VC clause generates precisely the same expected welfare as the incentive optimal direct mechanism

Consider next the more complicated case with $|\mathbf{x}_{\mathcal{B}^*}^*| = 2$. We start the proof by constructing a particular mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ 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 VC. We finally demonstrate that $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ converges in expected surplus to the incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ as $\nu_{B^*}\sigma_{B^*} \to 0$.

Subcase 1: An agent with marginal cost $\theta_{\underline{B}^*}$ reports marginal cost θ_{A^*} with zero probability, $\sigma_{A^*\underline{B}^*}^* = 0$. This is the case discussed in Section 4. Let $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ be described by the menu of contracts $\mathbf{x}_{\mathcal{A}^*} = \mathbf{x}_{\mathcal{A}^*}^{VC}$, $x_j = x_{\underline{B}^*} = x_{\underline{B}^*}^{VC}$ for all $j \in \mathcal{B}^*$ and $x_j = x_0$ for all $j \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$. As for the reporting strategies, $\sigma_i = 1$ for all $i \in \{1, \ldots A^* - 1\}$ 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 (17) from (18), we get

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

Hence, $q_{\underline{B}^*}^* > q_{\underline{B}^*}^{VC} = q_{\underline{B}^*}$ 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 $(i,j) \in \mathcal{A}^* \times \mathcal{A}^*$, $U_{A^*}(x_{A^*}) - U_{A^*}(x_{\underline{B}^*}) = U_{A^*}(x_{A^*}^*) - U_{A^*}(x_{\underline{B}^*}^*)$, and $q_{A^*}^* > q_{\underline{B}^*}$, 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 $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$. 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 $\mu_{ji} = \mu_i^{VC}$ for all $j \in \mathcal{B}^*$ and $i \in A^* \cup \mathcal{B}^*$, where $\mu_{A^*}^{VC}$ and μ_i^{VC} for all $i \in \mathcal{B}^*$ were described in (19). Moreover, $|\mathbf{x}_{\mathcal{A}^*}| = |\mathbf{x}_{\mathcal{A}^*}^*|$ implies (7). We finally need to verify sequential rationality (5) of $x_{\underline{B}^*}$ to establish incentive feasibility of $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$. To do so, we fist define the function

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

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

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

By this construction, $\tilde{q}(\sigma_{B^*}^*) = q_{\underline{B}^*}^*$, $\tilde{\Omega}(\sigma_{B^*}^*) = \Omega_{B^*}^*$, see the proof of Claim 16 for a defini-

58

tion, and $\tilde{q}(0) = q_{B^*}$. The expected surplus to the principal of offering the ex-post contract $x_{\underline{B}^*}$ after the agent has reported marginal cost θ_j, $j \in \mathcal{B}^*$, equals $\frac{\tilde{\Omega}(0)}{\nu_{A^*}(1-\sigma_{A^*}^*)+\nu_{\mathcal{B}^*}}$, whereas the deviation profit of offering a deviation contract $x_h^d = (q_h^d, \theta_h q_h^d)$, $h \in \{A^*, ..., B^* - 1\}$, equals $\frac{\Omega_h}{\nu_{A^*}(1-\sigma_{A^*}^*)+\nu_{\mathcal{B}^*}}$, see the proof of Claim 16 for the definition of Ω_h and discussion. Hence, $x_{\underline{B}^*}$ is sequentially rational if and only if $\tilde{\Omega}(0) \geq \Omega_h$ for all $h \in \{A^*, ..., B^* - 1\}$. Sequential rationality of $x_{\underline{B}^*}^*$ in the incentive optimal mechanism implies $\tilde{\Omega}(\sigma_{B^*}^*) \geq \Omega_h$ for all $h \in \{A^*, ..., B^* - 1\}$. We close the proof of incentive feasibility by demonstrating $\tilde{\Omega}(0) > \tilde{\Omega}(\sigma_{B^*}^*)$. As

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

we have $\tilde{\Omega}'(\sigma) < \tilde{\Omega}'(\sigma_{B^*}^*) = -\nu_{B^*} W_{B^*}(x_{\underline{B}^*}^*) \leq 0$ for all $\sigma < \sigma_{B^*}^*$, where we demonstrated $W_{B^*}(x_{\underline{B}^*}^*) \geq 0$ in the proof of Claim 13. Hence, $\tilde{\Omega}(0) > \tilde{\Omega}(\sigma_{B^*}^*)$.

By way of incentive feasibility of $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ and $|\mathbf{x}_{\mathcal{B}^*}| = 1$, the reduced communication mechanism augmented by a vague escape clause (VC) described in Section 4 can be sustained as a PBE. This mechanism generates the same expected surplus to the principal as $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$, namely:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\Omega_{B^*} > \Omega_{\underline{B}^*}^*$.

By the properties of $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$, the following mechanism sustained by a vague escape clause (VC) can be sustained as a PBE: The principal offers $\mathbf{x}_{\mathcal{A}^*}$ in Stage 1. In Stage 2, any agent with marginal cost θ_i, $i \in \mathcal{A}^*$, selects the contract x_i. An agent with marginal cost $\theta_{\underline{B}^*}$ selects x_{A^*} with probability $1 - \sigma_{\underline{B}^*}^*$ and invokes the escape clause with probability $\sigma_{\underline{B}^*}^*$. If $\mathcal{C}^* \neq \emptyset$, the any agent with marginal cost θ_i, $i \in \mathcal{C}^*$ rejects the mechanism. In Stage 3, the principal offers the ex-post contract $x_{\underline{B}^*}$ if the agent has invoked the escape

clause. This mechanism generates expected surplus

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

$$+ (\nu_{\mathcal{A}^*} + \nu_{\underline{B}^*}(1 - \sigma_{\underline{B}^*}^*))(1 - \alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q_{\underline{B}^*}^* - q_{\underline{B}^*}).$$

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

$$W(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$$

$$= \nu_{\underline{B}^*}\sigma_{\underline{B}^*}^*[W_{\underline{B}^*}^{fb}(q_{\underline{B}^*}^*) - W_{\underline{B}^*}^{fb}(q_{\underline{B}^*})] + \nu_{B^*}[W_{B^*}^{fb}(q_{\underline{B}^*}^*) - W_{B^*}^{fb}(q_{\underline{B}^*})]$$

$$+ \nu_{B^*}\sigma_{\underline{B}^*}^*[w_{B^*}^{fb} - W_{B^*}^{fb}(q_{\underline{B}^*}^*)] - (\nu_{\mathcal{A}^*} + \nu_{\underline{B}^*})(1 - \alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q_{\underline{B}^*}^* - q_{\underline{B}^*})$$

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

A.5 Proof of Lemma 3

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 3, 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 $k \in \{1, ..., K-1\}$ if $K \geq 2$, $\mathcal{A}_K = \hat{\mathcal{A}}_K \backslash \hat{A}$ 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

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

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

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

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

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

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

$|\mathbf{x}_{\mathcal{A}}| = |\hat{\mathbf{x}}_{\hat{A}}|$. 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.$\square$

A.6 Proof of Proposition 5

If $\hat{C} \neq \emptyset$, then $\hat{A}+1 \leq I$. If $\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$ is small, then $W_{\hat{A}}^{sb}(\hat{q}_{\hat{A}}) \approx W_{\hat{A}+1}^{sb}(\hat{q}_{\hat{A}}) \approx 0$; see (37). 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 $\hat{q}_{\hat{A}} > q_{\hat{A}}^{sb}$ (Lemma 4) then imply $\hat{q}_{\hat{A}} > q_{\hat{A}}^{fb}.\square$