IFN Working Paper No. 1390, 2021

Contracting with Endogenously Incomplete Commitment: Escape Clauses

Thomas Tangerås and Wolfgang Gick

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Contracting with endogenously incomplete commitment: Escape clauses[*]

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

May 27, 2021

Abstract

We study mechanism design under endogenously incomplete commitment as it arises in contracting with *escape clauses*. An escape clause permits the agent to end a contractual relationship under specified circumstances, after which the principal can offer an ex-post contract. Escape clauses are valuable when the maximal number of initial contracts is smaller than the number of agent types. We identify a sufficient condition for incentive optimality of ex-post contracting. Escape clauses are always incentive optimal under severely constrained contracting. On the margin, the optimal escape clause balances the benefit of a better adapted contract against an increase in dynamic inefficiency.

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

.

[*]We thank Eberhard Feess, Roland Strausz, Menghan Xu and Renkun Yang for helpful discussion, as well as audiences at the 2020 Conference on Mechanism and Institution Design, the 2020 Econometric Society World Conference and the Research Seminar at the Free University of Bozen/Bolzano for additional comments. Financial support from Jan Wallanders och Tom Hedelius stiftelse (P15-0305) is gratefully acknowledged.

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

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

1 Introduction

Escape clauses are contract stipulations that specify contingencies under which one or more parties to a contract can terminate the agreement.[1] In effect, an escape clause is a safety valve that enables contracting parties to avoid satisfying the conditions of the agreement, for instance if circumstances render fulfillment of the contractual terms too costly. Public service contracts are a key domain for the use of escape clauses. The Swedish Electricity Act, for instance, states that the following provisions apply to the regulation of electricity distribution networks:

> "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. for other valid reasons."[2]

If a network owner activates the escape clause in accordance with this act, then the regulatory authority is under legal obligation to provide a modified regulatory contract. However, the legal framework places no restrictions on this new contract other than the requirement that it must be a revenue cap. This arrangement has two interesting properties that sets it apart from the mechanisms typically analyzed in contract theory. First, the agreement between the principal and the agent contains the possibility of *ex-post contracting* in the sense that the principal in certain situations will propose a contract after the agent has communicated with the principal. Second, the possibility for ex-post contracting is included in the agreement by *contractual design*, and occurs at the agent's initiative.[3]

The situation above cannot be described within the two most common paradigms of contract theory. In a *complete commitment* framework, the principal engages in *pure ex-ante contracting*. This means that the principal commits to all contractual offers prior to any communication from the agent to the principal. The second paradigm is one of *exogenously incomplete commitment*. In that setting, the principal by assumption is unable to commit to parts of the initial agreement, an inability which leads to instances of ex-post contracting.[4]

Motivated by escape clauses, we develop a framework that deviates from the two previous paradigms by exploring contracting under *endogenously incomplete commitment*. The principal has the power to commit to pure ex-ante contracting, but can also deviate from this premise. Specifically, the mechanism offered to the agent may include an escape clause that the agent can invoke under stated conditions. Triggering the escape clause renders the initial agreement void and leads to subsequent ex-post contracting.

[1]The Cambridge Dictionary defines an escape clause as "a statement in a contract that allows you to break all or part of the contract under particular conditions." dictionary.cambridge.org/dictionary/english/escape-clause

[2]Ellag (1997:857), 5 kap. 20 §; riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/ellag-1997857_sfs-1997-857. Our translation.

[3]Escape clauses are also common in real estate and venture capital agreements. A similar stipulation is a *break clause*, typically featured in tenancy agreements, by which a party can end a contract prematurely. These clauses usually do not require the principal to make a subsequent contract offer, but they do not prevent the principal from doing so, either. As the principal generally cannot lose from proposing a new contract after a previous agreement has ended, such agreements are also likely to feature ex-post contracting.

[4]Other terminology for describing such problems is mechanism design with limited or imperfect commitment.

We ask the following fundamental questions:

- What are the contractual implications of introducing an escape clause?
- When does the principal's optimal mechanism include an escape clause?
- Which trade-offs do the principal face when designing an optimal escape clause?

The principal cannot gain anything from including an escape clause under complete commitment power if contracting is *unconstrained.* This implies that the principal can offer and enforce an unlimited number of contracts up front without any cost. The agreement then specifies one (possibly unique) contract for every possible future pay-off relevant state of the world. The principal chooses this mechanism to maximize its expected surplus subject to all relevant incentive compatibility (IC) and individual rationality (IR) constraints by the agent. We assume instead that contracting is *constrained* in the sense that the number of different contracts the principal can offer the agent up front is smaller than the number of pay-off relevant states of the world. This realistic deviation from the standard framework creates a *raison d'être* for an escape clause by increasing contractual flexibility in an agreement between a principal and an agent.

We introduce constrained contracting in the Baron and Myerson (1982) framework of how to regulate a monopoly with unknown costs. The principal (a regulatory authority) contracts with one agent (a monopoly firm) to produce quantity q of a good in exchange for transfer t. The agent has private information about its constant marginal cost θ_i. The principal knows the domain and probability distribution over the agent's cost types $i \in \{1....,I\} = \mathcal{I}$. A larger i means that the agent has higher marginal cost. Under the standard regularity condition of increasing virtual marginal cost, the second-best optimal mechanism consists of a menu of I different ex ante contracts, one contract $x_i^{sb} = (q_i^{sb}, t_i^{sb})$ for every cost type. However, the principal cannot implement this menu because of constrained contracting: the maximum number of unique contracts in any ex-ante menu is restricted to at most $K < I$ by assumption.

The principal can adapt to an environment with constrained contracting in several ways. The first is to offer pooling contracts to a subset $\mathcal{A} \subseteq \mathcal{I}$ of cost types. The principal then commits up front to a menu containing no more than K different contracts. An agent that subsequently reports marginal cost θ_j, $j \in \mathcal{A}$, receives the stipulated contract x_j. Second, the principal can exclude a subset $\mathcal{C} \subseteq \mathcal{I}$ of cost types. An agent that reports marginal cost θ_j, $j \in \mathcal{C}$, receives the null contract $q_0 = t_0 = 0$. Third, the principal can insert an escape clause into the mechanism offered to the agent. This escape clause is characterized in terms of a subset $\mathcal{B} \subseteq \mathcal{I}$ that does not involve any stipulated contract offer. Instead, an agent that reports marginal cost θ_j, $j \in \mathcal{B}$, will receive a subsequent contract offer x_j from the principal. The escape clause increases flexibility by allowing contracts to adapt to the circumstance of the agent ex post. The principal achieves this flexibility without any substantial increase in the number of contracts offered to the agent. The principal makes at most $K + 1$ contract offers on the equilibrium path regardless of the size of $\mathcal{B}$ because ex-post contracts are offered after the agent has reported marginal cost. For instance, the agent never receives more than a single contract offer in the polar extreme case of pure ex-post contracting, where $\mathcal{A} = \emptyset$.

The contractual implications of an escape clause Introducing an escape clause is associated with two fundamental incentive problems. The principal must prevent abuse of the escape clause since activating it can otherwise be a profitable way for the agent to improve contract conditions. Incentive compatibility of the ex-ante menu is maintained by increasing the transfer payment for any reported marginal cost θ_j, $j \in \mathcal{A}$. An escape clause therefore increases the expected informational rent of the agent. The second fundamental problem is a *ratchet effect* that arises because of contractual opportunism (Weitzman, 1980; Freixas et al., 1985). Consider an agent with marginal cost $\theta_i = \theta_j$, $j \in \mathcal{B}$, so that the agent triggers the escape clause by a truthful cost report. Suppose the principal believes that the agent will only ever exaggerate its marginal cost. The principal then infers from the cost report θ_j that the agent has marginal cost $\theta_i \leq \theta_j$. The escape clause renders the initial agreement void, so the agent's outside option to an ex-post contract is no contract at all. The principal prevents unnecessary transfer payments by offering an ex-post contract that gives zero rent to an agent with marginal cost $\theta_i = \theta_j$. The agent foresees this contractual opportunism by the principal and therefore lies about its cost. Manipulation of cost reports to maintain informational rent under ex-post contracting constitutes the ratchet effect.

Ex-post contracting causes the revelation principle to break down (Laffont and Tirole, 1988). Bester and Strausz (2001) demonstrate that a stochastic revelation principle applies instead in situations such as the above. There exists an incentive efficient direct mechanism in which the agent reports its true cost with positive probability. The agent's reporting strategy and the ex-post contracts by the principal form a Perfect Bayesian Equilibrium (PBE). The cost reports by the agent are optimal given the agent's correct anticipation of how these reports affect future contract offers; the ex-post contract offers are sequentially rational subject to the principal's posterior beliefs about the agent's cost being based on the agent's actual reporting strategies and derived using Bayes' rule.

A reporting strategy for which the agent possibly exaggerates marginal cost, but also delivers a truthful report with positive probability, cannot be sustained as an equilibrium under an escape clause. Consequently, the agent sometimes *understates* marginal cost in equilibrium. The IC constraints for such cost types are locally both downward- and upward-binding, and so the ratchet effect implies pooling within $\mathcal{B}$. If $\theta_{\underline{B}}$ is the smallest and θ_B the largest cost report that triggers the escape clause, all cost reports θ_j, $j \in \{\underline{B}, ..., B-1\}$, yield the same contract offer $x_{\underline{B}}$ in equilibrium. Introducing an escape clause therefore increases the contractual degree of freedom by at most 2 contract offers compared to a mechanism with complete commitment. Further, all cost types contained in $\mathcal{A}$ are more efficient than those in $\mathcal{B}$, which are more efficient than those in $\mathcal{C}$.[5] An agent that triggers the escape clause does so because the initial contract offers are insufficient to provide the required return given the agent's production cost. The purpose of an escape clause is to enable a relatively inefficient agent to obtain a contract better suited to that agent's particular circumstances.

The ratchet effect also implies that the informational content to the principal when observing any specific cost report θ_j, $j \in \mathcal{B}$, is limited. We show that uniform randomization across $\mathcal{B} \backslash B$

[5]Specifically, $\mathcal{A} = \{1, ..., A\}$, $\mathcal{B} = \{\underline{B}, ..., B\}$, where $\underline{B} = A + 1$, and $\mathcal{C} = \{B+1, ..., I\}$ if $B \leq I - 1$.

for an agent with marginal cost θ_i, $i \in \mathcal{B}$, is an optimal reporting strategy. All cost reports that trigger the escape clause, other than perhaps θ_B, then are equally (un)informative. Hence, the principal's loss in expected surplus from offering a *vague escape clause* is limited, relative to specifying the exact circumstances $\mathcal{B}$ under which the escape clause applies. A vague escape clause essentially allows the agent to opt out of an initial menu of contracts at free will.

When mechanisms contain an escape clause Entering an escape clause into a mechanism increases flexibility by increasing the number of contracts in equilibrium, but this comes at the cost of increased informational rent. Ex-post contracts are furthermore dynamically inefficient because they entail a trade-off between output and rent extraction that is inefficient from an ex ante perspective. This inefficiency occurs because the informational rent to the agent associated with the escape clause is sunk when the principal makes the ex-post contract offer. Because of these inefficiencies, it is not self-evident that the principal will always find it optimal to include an escape clause under constrained contracting.

We consider first the case of severely constrained contracting in which the principal at most can offer one single contract up front ($K = 1$). This assumption captures the "one-size-fits-all" property of many agreements. It is then better for the principal to engage in pure ex-post contracting rather than offer a single pooling contract up front. Doing so increases the flexibility of contracts without creating any additional informational rent. The model therefore predicts mechanisms with severely constrained contracting always to contain an escape clause.

Things are more complicated if the principal can offer multiple contracts up front ($K \geq 1$). We establish a sufficient condition on the menu of pure ex-ante contracts when it is optimal to include an escape clause in the mechanism. Let q_A be the output of the least efficient agent θ_A that operates under the ex-ante menu. Pooling a subset of cost types in an ex-ante contract distorts this output upward from a second-best perspective, $q_A > q_A^{sb}$. Such upward distortion occurs also under ex-post contracting because of dynamic inefficiency. Introducing an escape clause that only applies to an agent with marginal cost θ_A, increases the expected surplus of the principal if output exceeds the first-best level, $q_A > q_A^{fb}$. The ex-post contract x_A^{fb} under the escape clause then provides a better balance between efficiency and informational rent than the ex-ante contract, from a second-best perspective. The output condition is satisfied, for instance if the principal places sufficient weight on efficiency relative to rent extraction.

The trade-offs involved in designing an escape clause An escape clause is characterized in terms of its lower and upper boundaries $\theta_{\underline{B}}$ and θ_B. Broadening it by reducing the lower boundary to $\theta_{\underline{B}-1} = \theta_A$ has two effects on the expected surplus of the principal. An agent with marginal cost θ_A will now be on an ex-post instead of an ex-ante contract. This modification benefits the principal all else equal if it reduces the output distortion. However, this modification also increases ex-post output because an agent that triggers the escape clause on average now is more efficient than before. This output expansion has a first-order negative effect on expected informational rent. On the margin, an optimally designed escape clause balances the benefit of a better adapted contract against the cost of an increase in dynamic inefficiency.

In our interpretation, K literally means the number of different contracts offered to the agent up front. More generally, K is equal to the maximal number of non-trivially binding IC and IR constraints in an ex-ante mechanism. In this sense, K is a measure of contract complexity. We treat K as a parameter. One could argue that contracting costs are what constrain the number of contracts. We illustrate in a simple example how even small contracting costs can be sufficient to render constrained contracting efficient from the viewpoint of the principal. One can also think of other types of clauses that trigger ex-post contracting, such as renegotiation clauses. Our general message is that there are circumstances under constrained contracting where mechanisms with ex-post contracting dominate all those with complete commitment. If renegotiation or other clauses are better than escape clauses from the viewpoint of the principal, such a result would reinforce the argument for mechanisms with endogenously incomplete commitment.

Related literature Our paper contributes to the literature on mechanism design with incomplete commitment. Seminal contributions were Freixas et al. (1985) and Laffont and Tirole (1988), who analyzed short-term contracting in a multi-period framework.[6] Bester and Strausz (2001, 2007), Skreta (2006), and, more recently, Doval and Skreta (2021) have developed a more general methodology for analyzing mechanisms with incomplete commitment.[7] Commitment issues arise naturally in a setting with repeated sales (e.g. Tirole, 2016; Beccuti and Möller, 2018; Breig, 2020), in organizations (e.g. Shin and Strausz, 2014) and in auctions (e.g. Vartiainen, 2013; Skreta, 2015; Akbarpour and Li, 2020). A common denominator of these papers is that limited commitment is imposed. Ours appears to be the first to consider commitment as a mechanism design variable, specifically in the form of an escape clause.[8] Applying the Bester and Strausz (2001) methodology enables us to characterize equilibrium contracts under incomplete commitment in much larger detail than what is usually the case.[9] In particular, we show that nearly all ex-post contracts will be the same and that uniform randomization strategies form part of an incentive optimal mechanism.[10]

We consider mechanisms in which the agent receives a transfer in exchange for completing an assigned task. Following Holmström (1984), a large literature has developed in which the principal and the agent contract on some action by the agent, but where no contingent transfers

[6] The literature on dynamic contracting in which the principal commits to a long-term contract, is surveyed in Bergemann and Välimäki (2019).

[7] A precursor is Kumar (1985), who characterizes a "noisy" revelation principle in a sequential incentive mechanism.

[8] Fudenberg and Tirole (1983) analyze bargaining under incomplete information. A seller proposes to trade an indivisible good at an initial price. If the buyer declines the offer, then the seller proposes a revised price. The buyer either accepts or rejects the new offer, after which the game ends. They make the interesting observation that the seller may benefit from adding the second stage, i.e. introduce ex-post contracting. Contracting is constrained in their setting by an assumption that the seller provides a single price offer in the first stage.

[9] Fiocco and Strausz (2015) fully characterize the equilibrium in a two-period model of optimal regulation without commitment where the agent can be one of two cost types.

[10] This randomization strategy is similar to that found in the classical analysis of strategic information transmission by Crawford and Sobel (1982). The informed agent (Sender) submits a signal containing pay-off relevant information to the principal (Receiver) who then takes an action that affects both. Sender randomizes across signals to increase expected rent. The model can sustain partition equilibria that are more or less informative. Our model features transfer payments and an ex-post participation constraint by the agent. These features reduce the informativeness of the agent's cost reports that trigger the escape clause in our setting.

are available for accomplishing incentive compatibility. Since the fundamental design issue is how much freedom to leave to the agent, these are often referred to as models of optimal delegation; see Amador and Bagwell (2013) for a general treatment. In our context, contracting is constrained by the number of different contracts the principal can offer the agent up front, whereas contracting is constrained by the lack of transfers in delegation problems. This constraint also leaves room for escape clauses to improve efficiency (e.g. Bagwell and Staiger, 2005; Beshkar and Bond, 2017; Coate and Milton, 2019). In particular, Halac and Yared (2020) show that the optimal mechanism features an escape clause if the principal can verify the agent's type by paying a fixed cost. Delegation models differ from the current framework by featuring complete commitment. The only issue for implementation is to ensure incentive compatibility by the agent. This property is plausible for typical applications such as design of international agreements and fiscal rules. The purpose of escape clauses in such contexts is to permit members to *temporarily* suspend their obligations under the agreement for a limited period of time to allow them to respond to temporary shocks, but there is no ex-post contracting.[11,12]

We organize our paper into the following sections. Section 2 describes the contracting problem. Section 3 characterizes mechanisms under complete commitment (pure ex-ante contracting). We establish fundamental properties of contracts and reporting strategies in mechanisms with incomplete commitment in Section 4, which results in a proposition on vague escape clauses. Section 5 demonstrates the dominance of pure ex-post over pure ex-ante contracting under severely constrained contracting. Section 6 identifies a sufficient condition for when ex-post contracting is incentive optimal under general constrained contracting. We establish the fundamental tradeoffs involved in designing an optimal escape clause in Section 7. Section 8 discusses key modeling assumptions, whereas Section 9 concludes the paper. Lengthy proofs are in the appendix.

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.

A contract $x = (q, t)$ is a pair specifying an output requirement $q \geq 0$ that the agent has to satisfy 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

[11] Halac and Yared (2020) have an extension with limited commitment. Since the principal anyway always implements her most-preferred action subsequent to verification, limited commitment plays a role in their model only in the event the agent has *not* triggered the escape clause.

[12] Optimal contracting has been extended to a dynamic framework by way of the theory of optimal monetary discretion; see for instance, Athey et al. (2005), Halac and Yared (2014).

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. For any given output q, the principal wants to minimize the agent's rent by setting the transfer t as small as possible. 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.

Under complete information, the principal would set $U_i(x) = 0$ and maximize

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

over q. We assume that $S'(q) < \theta_I$ for some $q > 0$ and that $\lim_{q \to 0} S'(q) > 0$ is sufficiently large that the first-best contract $x_i^{fb} = (q_i^{fb}, t_i^{fb})$ entails strictly positive and bounded output and transfer payments:

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

We assume that the first-best contract is always 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 all cost types produce a strictly positive output in this mechanism.

We study a contracting problem with incomplete information. The setup is standard in the sense that everything is common knowledge except that the agent has private information about its marginal cost θ_i prior to contracting. The principal only knows the distribution characteristics $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$. The solution to this problem typically specifies a menu of contracts, one for each cost report of the agent. We deviate from this setup by limiting the total number of contracts K the principal can offer the agent up front. K is then a measure of the contractual constraints, with a smaller K meaning a more constrained environment. We refer to the polar case $K = 1$ as one of *severely constrained* contracting. For $K \geq I$, contracting is effectively unconstrained. We also deviate from the standard paradigm by assuming that the principal can reserve some flexibility to contract ex post.

Specifically, 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: $|\mathbf{x}_{\mathcal{A}}| \leq K$.[13]

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

- Rejection: The principal and the agent each receive their reservation utility 0. Game over.
- Acceptance: The game continues to the next stage.

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

- 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 each receive their reservation utility 0. 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. The mechanism features *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 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 $\mathbf{\Sigma}$ (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 that 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 can always offer the null contract x_0 regardless of K. This is not unreasonable given the simplicity of this particular contract. The null contract allows us to handle *partial participation*, where some types do not produce a positive quantity in equilibrium, in a simple manner.

[13] We denote the cardinality or rank of a set $\mathbf{\Upsilon}$ by $|\mathbf{\Upsilon}|$. The cardinality measures the number of *unique* elements in $\mathbf{\Upsilon}$, not the *total* number of elements.

A mechanism with incomplete commitment (ex post contracting) 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 of type $i \in \mathcal{I}$ claims to be of type $j \in \mathcal{I}$. We let $\sigma_i = \sigma_{ii}$ denote the probability that i truthfully reports its type. 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.

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

$$U_i(x_i) \geq 0 \ \forall i \in \mathcal{I} \tag{1}$$

$$U_i(x_i) \geq U_i(x_j) = U_j(x_j) + (\theta_j - \theta_i) q_j \ \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{2}$$

$$\sigma_i > 0, \ \sigma_{ji}(U_i(x_i) - U_i(x_j)) = 0 \ \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{3}$$

$$x_j \in \arg\max_{x' \in R_+^2} \textstyle\sum_{i=1}^{I} \mu_{ji} W_i(x') \ \forall j \in \mathcal{B} \text{ if } \mathcal{B} \neq \emptyset \tag{4}$$

$$\mu_{ji} = \frac{\nu_i \sigma_{ji}}{\sum_{h=1}^{I} \nu_h \sigma_{jh}} \ \forall (i,j) \in \mathcal{I} \times \mathcal{I} \tag{5}$$

$$|\mathbf{x}_{\mathcal{A}}| \leq K \tag{6}$$

Constraints (1) and (2) 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 $\mathcal{B} \neq \emptyset$. The three constraints (3)-(5) are the associated equilibrium conditions. First, (3) 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, (4) 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, (5) is a consistency requirement that the principal's posterior beliefs satisfy Bayes' rule. The final constraint (6) appears because of constrained ex-ante contracting, and does not appear in Bester and Strausz (2001). We use $\Gamma(\mathcal{A}, \mathcal{B})$ to label the set of incentive feasible mechanisms given $(\mathcal{A}, \mathcal{B})$.

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

$$W(\mathbf{x}, \boldsymbol{\Sigma}|\mathcal{A}, \mathcal{B}) = \textstyle\sum_{i=1}^{I} \sum_{j=1}^{I} \nu_i \sigma_{ji} W_i(x_j) \tag{7}$$

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 $\boldsymbol{\Sigma}$.

Complete commitment represents the default mode in mechanism design analysis where no additional contracting occurs after the agent has reported its type (pure ex-ante contracting). Bester and Strausz (2001) consider the alternative setting of exogenously incomplete commitment, i.e. for given $(\mathcal{A}, \mathcal{B})$ in the present context. Our paper attempts to bridge the cap 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 $W(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \mathcal{B})$. A mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ that solves this problem is *incentive optimal.*

There can be instances when an incomplete commitment mechanism can do as well as one with complete commitment, but no better. We stack the deck against ex-post contracting:

Definition 1 (Escape clauses are minimal) *An incentive feasible 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:*

$$W(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*) \geq W(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \mathcal{B}) \ \forall (\mathcal{A}, \mathcal{B}) \subset [\mathcal{I} \cup \emptyset] \times [\mathcal{I} \cup \emptyset], \ \mathcal{A} \cap \mathcal{B} = \emptyset, \tag{8}$$

and the escape clause is minimal in the following sense:

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

The incentive optimal mechanism entails *endogenous incomplete commitment* if $\mathcal{B}^* \neq \emptyset$. The incentive optimal escape clause is minimal in the sense that the principal cannot find an incentive feasible mechanism with a smaller or no escape clause that delivers the same expected surplus. By implication, the principal must *strictly* benefit from abandoning the complete commitment framework for the incentive optimal mechanism to contain an escape clause. Otherwise, pure ex-ante contracting is incentive optimal.

3 Mechanisms with complete commitment

This section analyzes the 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 ($\mathcal{B} = \emptyset$). These mechanisms establish benchmarks against which we are able to evaluate the merits and drawbacks of incomplete commitment mechanisms.

By the revelation principle, we can restrict attention to truth-telling mechanisms, i.e. incentive feasible mechanisms for which $\mathbf{\Sigma} = \mathbf{I}$, where $\mathbf{I}$ is the I-dimensional identity matrix. The principal then maximizes (7) over $\mathbf{x}_{\mathcal{A}}$ subject to (1), (2) and (6). Incentive compatibility implies that output is non-increasing in the agent's marginal cost. Hence, $\mathcal{A} = \{1, ..., A\}$, where θ_A is the marginal cost of the least efficient agent that produces positive output in the mechanism. The mechanism features full participation if $A = I$. Otherwise, $\mathcal{C} = \{A + 1, ..., I\}$.

Set $\mathcal{A}$ is partitioned into $\tilde{K} \leq K$ *cost groups*, $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_1, ..., \mathcal{A}_k, ..., \mathcal{A}_{\tilde{K}}\}$, because of constrained contracting. Each cost group consists of all cost types that operate under the same contract. Cost groups are convex because output is non-increasing in marginal cost. We denote by $x_{\mathcal{A}_k} = (q_{\mathcal{A}_k}, t_{\mathcal{A}_k})$ the contract for cost group $\mathcal{A}_k$. We will sometimes use $x_{A_k} = (q_{A_k}, t_{A_k})$ as a substitute for $x_{\mathcal{A}_k}$, where A_k is the least efficient cost type contained in $\mathcal{A}_k$. In particular, $x_{\mathcal{A}_{\tilde{K}}} = x_A = (q_A, t_A)$ since A is the upper boundary cost type in $\mathcal{A}_{\tilde{K}}$.

The principal minimizes transfer payments to minimize agency rent. By standard arguments (see the appendix), the IR constraint of the upper boundary cost type A in $\mathcal{A}$ and the IC constraints of all more efficient types are locally downward-binding, so the pure ex-ante mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\mathcal{A}, \emptyset)$ is incentive efficient only if transfers are set in a way that

$$U_i(\hat{x}_i) = \sum\nolimits_{j=i}^{A-1} (\theta_{j+1} - \theta_j)\hat{q}_{j+1} \ \forall i \in \{1, 2, ..., A-1\}, \ U_A(\hat{x}_A) = 0. \tag{10}$$

The rent of an agent with cost θ_i is found by adding up the rents for less efficient types, loosely the discrete type version of the well-known integral in the continuous type case.

Substituting the expressions for agency rent into (7) yields the principal's expected surplus

$$W(\hat{\mathbf{x}}, \mathbf{I}|\mathcal{A}, \emptyset) = \sum_{i=1}^{A} \nu_i W_i(\hat{x}_i) = \sum_{k=1}^{\tilde{K}} \nu_{\mathcal{A}_k} \tilde{W}_{\mathcal{A}_k}(\hat{q}_{\mathcal{A}_k}). \tag{11}$$

of the incentive efficient mechanism. In this expression, $\nu_{\mathcal{A}_k} = \sum_{i \in \mathcal{A}_k} \nu_i$ measures the ex-ante probability that agent's marginal cost belongs in cost group $\mathcal{A}_k$.

$$\tilde{W}_{\mathcal{A}_k}(q) = S(q) - [\sum_{i \in \mathcal{A}_k} \frac{\nu_i}{\nu_{\mathcal{A}_k}} (\theta_i + (1-\alpha)(\theta_{A_k} - \theta_i) + \frac{G_{A_{k-1}}}{\nu_{\mathcal{A}_k}} (1-\alpha)(\theta_{A_k} - \theta_{A_{k-1}})]q \tag{12}$$

is the principal's utility of output q in cost group $\mathcal{A}_k$ minus the virtual production cost of this output, where $\frac{G_{A_{k-1}}}{\nu_{\mathcal{A}_k}}$ is the hazard rate of cost group $\mathcal{A}_k$. We let $G_{A_0} = 0$.

Maximization of $\tilde{W}_{\mathcal{A}_k}(q)$ over q yields the incentive efficient output in cost group $\mathcal{A}_k$ as the solution to

$$S'(\hat{q}_{\mathcal{A}_k}) = \sum_{i \in \mathcal{A}_k} \frac{\nu_i}{\nu_{\mathcal{A}_k}} (\theta_i + (1-\alpha)(\theta_{A_k} - \theta_i)) + \frac{G_{A_{k-1}}}{\nu_{\mathcal{A}_k}} (1-\alpha)(\theta_{A_k} - \theta_{A_{k-1}}). \tag{13}$$

Output is downward distorted to extract rent from more efficient cost groups. The principal's expected surplus becomes

$$\tilde{W}_{\{i\}}(q) = W_i^{sb}(q) = S(q) - (\theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1}))q$$

if cost group $\mathcal{A}_k$ consists of one single cost type, $\mathcal{A}_k = \{i\}$. The incentive efficient output then is the second-best efficient output, $\hat{q}_{\{i\}} = q_i^{sb}$, solved by

$$S'(q_i^{sb}) = \theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1}). \tag{14}$$

We employ the standard regularity assumption

$$\theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1}) < \theta_{i+1} + \frac{G_i}{\nu_{i+1}}(1-\alpha)(\theta_{i+1} - \theta_i) \ \forall i \in \{1, ..., I-1\} \tag{15}$$

of increasing virtual marginal production cost. This implies that it is incentive efficient to take full advantage of all contractual flexibility, meaning that the incentive efficient number $\hat{K}$ of cost groups satisfies $\hat{K} = A$ for all $A \leq \min\{K; I\}$.

Contracting is unconstrained if the principal can offer at least as many contracts as there are cost types, $K \geq I$. Note that excluding some agent types from producing output by offering them the null contract could still be profitable if the cost savings on informational rent are sufficient. The principal's expected surplus associated with offering the second-best contract to an agent with marginal production cost θ_i equals $\nu_i w_i^{sb} = \nu_i W_i^{sb}(q_i^{sb})$ under unconstrained contracting. It is easy to verify that w_i^{sb} is strictly decreasing in marginal cost θ_i. We ensure that the second-best efficient mechanism $(\mathbf{x}^{sb}, \mathbf{I}|\mathcal{I}, \emptyset)$ with full participation is incentive optimal under unconstrained contracting by imposing the assumption that $w_I^{sb} > 0$.

Let $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ be a mechanism that maximizes the principal's expected surplus under complete commitment if contracting is constrained, $K < I$. $\hat{\mathcal{A}} = \{1, ..., \hat{A}\}$ is the subset of cost types that produce positive output in this mechanism. The above assumptions imply that the mechanism takes full advantage of all contractual flexibility, $\hat{K} = K \leq \hat{A}$.

To derive the properties of incentive efficient partitioning $\hat{\boldsymbol{\mathcal{A}}} = \{\hat{\mathcal{A}}_1, ..., \hat{\mathcal{A}}_k, ..., \hat{\mathcal{A}}_K\}$ of $\hat{\mathcal{A}}$ into cost groups for $K \geq 2$, 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 $\mathcal{A}_{k+1} = \hat{\mathcal{A}}_{k+1} \cup \hat{A}_k$, $k \leq K-1$. All other cost groups remain unchanged if $K \geq 3$. 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 $x_{\hat{A}_j} = \hat{x}_j$ for all $j \in \{\hat{A}_k + 1, ...I\}$. The difference in expected surplus between the two mechanisms simplifies to

$$W(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset) - W(\mathbf{x}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset) = \nu_{\hat{A}_k}[W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k}) - W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{A}_k + 1})] \geq 0. \quad (16)$$

By an analogous argument, an agent with marginal cost $\theta_{\hat{A}_k + 1}$ optimally belongs in cost group $\hat{\mathcal{A}}_{k+1}$ only if $W_{\hat{A}_k+1}^{sb}(\hat{q}_{\hat{A}_k+1}) \geq W_{\hat{A}_k+1}^{sb}(\hat{q}_{\hat{A}_k})$. Similar conditions arise for the boundary cost type $\hat{A}$ under partial participation. We summarize these findings as follows:

Lemma 1 *Assume that the mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ maximizes the principal's expected surplus under complete commitment and constrained contracting ($K < I$).*

1. *Output in cost group $\hat{\mathcal{A}}_k$ is characterized by (13). Under pooling, $|\hat{\mathcal{A}}_k| \geq 2$:*
   - (a) *Output is downward distorted relative to the second-best efficient output of the most efficient cost type in $\hat{\mathcal{A}}_k$, $\hat{q}_{\hat{\mathcal{A}}_k} < q_{\hat{A}_{k-1}+1}^{sb}$.*
   - (b) *Output is upward distorted relative to the second-best efficient output of the least efficient cost type in $\hat{\mathcal{A}}_k$, $\hat{q}_{\hat{\mathcal{A}}_k} > q_{\hat{A}_k}^{sb}$.*
2. *The upper boundary type $\hat{A}_k$ in interior cost group $k \in \{1, ..., K-1\}$, $K \geq 2$, satisfies the principal's local incentive compatibility constraint:*

$$W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{\mathcal{A}}_k}) - W_{\hat{A}_k}^{sb}(\hat{q}_{\hat{\mathcal{A}}_{k+1}}) \geq 0 \geq W_{\hat{A}_k+1}^{sb}(\hat{q}_{\hat{\mathcal{A}}_k}) - W_{\hat{A}_k+1}^{sb}(\hat{q}_{\hat{\mathcal{A}}_{k+1}}). \quad (17)$$

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:*

$$W^{sb}_{\hat{A}}(\hat{q}_{\hat{A}}) \geq 0 \geq W^{sb}_{\hat{A}+1}(\hat{q}_{\hat{A}}). \tag{18}$$

Notice the similarity between (17) and (18), the local downward and upward IC constraints, and the IR constraint of the agent. The difference is that the principal's constraints above are evaluated using the second-best welfare function $W^{sb}_i(q)$.

4 Properties of mechanisms with incomplete commitment

The principal cannot implement $(\mathbf{x}^{sb}, \mathbf{I}|\mathcal{I}, \emptyset)$ if $K < I$, by the assumption that $|\mathbf{x}^{sb}| = I$. If the principal still wants to maintain complete commitment, it can pool cost types into cost groups or reduce participation in the mechanism. We proceed in this section by considering the fundamental properties of incentive feasible and incentive optimal mechanisms under the assumption that these mechanisms also feature ex-post contracting ($\mathcal{B} \neq \emptyset$).

In any mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ that features incomplete commitment ($\mathcal{B} \neq \emptyset$), we let $\theta_{\underline{B}}$ be the minimal and θ_B the maximal cost report that yields ex-post contracting: $(\underline{B}, B) \in \mathcal{B} \times \mathcal{B}$, $\underline{B} \leq B$, and $\mathcal{B} \subseteq \{\underline{B}, ..., B\}$. We do not impose convexity on $\mathcal{B}$. Recall that θ_A is the maximal cost report that yields ex-ante contracting.

Lemma 2 (Fundamental properties of contracts) *Assume that the mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ is incentive feasible and features incomplete commitment ($\mathcal{B} \neq \emptyset$).*

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 ex-post contract offer $x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}})$.*
   (b) *Cost report θ_B yields ex-post contract offer x^{fb}_B.*

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

3. *This mechanism exploits all available flexibility: $|\mathbf{x}^*_{\mathcal{A}^*}| = K$.*
4. *All cost types contained in $\mathcal{B}^*$ are less efficient that those contained in $\mathcal{A}^*$.*
   (a) *All ex-post contracts have lower output than all ex-ante contracts: $q^*_{\underline{B}^*} < q^*_{A^*}$.*
   (b) *All ex-post contracts render the upper boundary cost type B^* indifferent between producing and not: $x^*_{\underline{B}^*} = (q^*_{\underline{B}^*}, \theta_{B^*} q^*_{\underline{B}^*})$ and $x^*_{B^*} = (q^*_{B^*}, \theta_{B^*} q^*_{B^*})$.*
5. *All cost types contained in $\mathcal{C}^*$ are less efficient that those contained in $\mathcal{B}^*$ if the mechanism features partial participation ($\mathcal{C}^* \neq \emptyset$).*

Ex-ante contracting occurs whenever the agent reports marginal cost equal to or below threshold θ_{A^*}, whereas ex-post contracting only occurs after the agent reports marginal cost above this threshold. Hence, the escape clause applies to circumstances when the cost of supplying output is reportedly high. It can also be the case that there is no production at all. This happens if $B^* \leq I-1$, and the agent reports marginal cost above θ_{B^*}.

Incomplete commitment increases the degrees of freedom in the mechanism by at most 2 contract offers compared to 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 defining incentive optimal mechanisms in terms of those with minimal escape clauses: the result applies to all incentive feasible mechanisms. Instead, flexibility is limited under ex-post contracting by the ratchet effect that renders incentive compatibility constraints both downward- and upward-binding.

The "no-distortion-at-the-bottom" contract x_B^{fb} for $|\mathbf{x}_{\mathcal{B}}| = 2$, follows from the discretionary nature of a mechanism with incomplete commitment. This mechanism still leaves an informational rent to a relatively efficient agent $\theta_i < \theta_B$. Unlike in the complete commitment setting, however, the transfer payments necessary to reach incentive compatibility are sunk after the agent has announced marginal cost θ_B in Stage 4 of the game. Consequently, there is no 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. Upon observing cost report θ_B, the principal's sequentially rational choice therefore is to offer the first-best efficient contract.

Lemma 2 contains a lot of information about the incentive optimal menu $\mathbf{x}^*$ of contracts, except for the sequentially rational quantity $q^*_{\underline{B}^*}$. This quantity depends on the incentive optimal reporting strategy $\mathbf{\Sigma}^*$. For its characterization, it is helpful to introduce some additional notation. Let $\underline{\mathcal{B}}^* = \{\underline{B}^*, ..., B^*-1\}$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ and $\underline{\mathcal{B}}^* = \mathcal{B}^*$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$. By this construction, any cost report θ_j, $j \in \underline{\mathcal{B}}^*$ activates the escape clause, leading the principal to offer the contract $x^*_{\underline{B}^*}$ in Stage 4 in the incentive optimal mechanism. The number $|\underline{\mathcal{B}}^*|$ of elements contained in $\underline{\mathcal{B}}^*$ is equal to $B^* - A^* - 1$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ and $B^* - A^*$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$.

Lemma 3 (Fundamental properties of reporting strategies) *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* $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ *that also features incomplete commitment* ($\mathcal{B}^* \neq \emptyset$)*, with the reporting strategy* $\mathbf{\Sigma}^*$ *fulfilling the following properties:*

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

   (a) *Relatively efficient types truthfully report their cost:* $\sigma_i^* = 1 \ \forall i \in \{1, ..., A^*-1\}$ *if* $A^* \geq 2$.

   (b) *The upper boundary type* A^* *may invoke the escape clause:* $\sigma^*_{A^*} \leq 1$, $\sigma^*_{jA^*} = \frac{1-\sigma^*_{A^*}}{|\underline{\mathcal{B}}^*|}$ $\forall j \in \underline{\mathcal{B}}^*$.

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 is minimal:* $\sigma^*_{\underline{B}^*} = 1$ *if* $\underline{B}^* = B^*$*;*

ii. *may choose not to invoke the escape clause if the clause is very small:* $\sigma^*_{\underline{B}^* A^*} = 1 - \sigma^*_{\underline{B}^*} \geq 0$ *if* $\underline{B}^* = B^* - 1$ *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* $\underline{B}^* = B^* - 1$ *and* $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$ *or if* $\underline{B}^* \leq B^* - 2$*.*

(b) *Intermediary types randomize across different cost reports that yield ex-post contracting:* $\sigma^*_{ji} = \frac{1}{|\underline{\mathcal{B}}^*|}$ $\forall (i,j) \in \{\underline{B}^* + 1, B^* - 1\} \times \underline{B}^*$ *if* $\underline{B}^* \leq B^* - 2$*.*

(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* $\mathcal{C}^* \neq \emptyset$*.*

Lemma 3 almost completely characterizes the incentive optimal reporting strategy under incomplete commitment, despite the potentially large set of cost types and cost reports. In this mechanism, any relatively efficient agent tends to choose its designated ex-ante contract in equilibrium. A possible exception occurs if the agent has boundary marginal cost θ_{A^*}. This agent potentially triggers the escape clause by exaggerating marginal cost.

If $\sigma_{A^*} < 1$, the only other agent type that would potentially choose anything other than its designated contract in equilibrium is an agent with marginal cost of θ_{B^*}. This agent understates marginal cost with positive probability by randomizing across all cost types in $\mathcal{B}^*$. Such behavior mitigates contractual opportunism by the principal and ensures that more efficient types in $\mathcal{B}^*$ receive an informational rent even under incomplete commitment.

By using the properties of incentive optimal ex-post contracts established in Item 4 of Lemma 2 and uniform randomization established in Lemma 3, we now apply (4) to characterize $q^*_{\underline{B}^*}$ for $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$:

$$S'(q^*_{\underline{B}^*}) = \frac{\sum_{i=A^*}^{B^*} \nu_i(\theta_i + (1-\alpha)(\theta_{B^*} - \theta_i)) - \nu_{A^*}\sigma^*_{A^*}(\theta_{A^*} + (1-\alpha)(\theta_{B^*} - \theta_{A^*})) - \nu_{B^*}\sigma^*_{B^*}\theta_{B^*}}{\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*} - \nu_{B^*}\sigma^*_{B^*}}, \tag{19}$$

where $\nu_{\mathcal{B}^*} = \sum_{i \in \mathcal{B}^*} \nu_i$ is the ex-ante probability that the agent has a cost type in $\mathcal{B}^*$. We can set $\sigma^*_{B^*} = 0$ in the above equation to get $q^*_{\underline{B}^*}$ for $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$.

To see how ex-post contracting affects the outcome, compare $q^*_{\underline{B}^*}$ to the incentive efficient quantity $\hat{q}_{\mathcal{B}^*}$ in an ex-ante mechanism where $\mathcal{B}^*$ constitutes a separate a cost group. We obtain this quantity by replacing $\mathcal{A}_k$ with $\mathcal{B}^*$, A_k with B^* and A_{k-1} with A^* in equation (13). The difference

$$S'(\hat{q}_{\mathcal{B}^*}) - S'(q^*_{\underline{B}^*}) = \frac{G_{A^*}}{\nu_{\mathcal{B}^*}}(1-\alpha)(\theta_{B^*} - \theta_{A^*}) + \sum_{i \in \mathcal{B}^*} \alpha\nu_i \frac{\nu_{A^*}(1-\sigma^*_{A^*})(\theta_i - \theta_{A^*}) + \nu_{B^*}\sigma^*_{B^*}(\theta_{B^*} - \theta_i)}{\nu_{\mathcal{B}^*}(\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*} - \nu_{B^*}\sigma^*_{B^*})}$$

in marginal utility is strictly positive, so $q^*_{\underline{B}^*} > \hat{q}_{\mathcal{B}^*}$. The ex-post contract is less distorted than the ex-ante one. Contrary to the ex-ante contract, the ex-post contract does not account for sunk informational rents to more efficient types. This is the first term of the right-hand side of the above expression. The relatively efficient cost type A^* may invoke the escape clause, so reducing the virtual marginal cost of output under the ex-post contract. This effect is captured by the terms involving $\sigma^*_{A^*}$. If $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$, the virtual marginal cost of the ex-post contract is further reduced when the inefficient cost type B^* is more likely to report its cost truthfully. This final effect is captured by the terms involving $\sigma^*_{B^*}$.[14] 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 in order to better adapt to 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 the ex-post contract offer; (ii) the agent may trigger the escape clause by exaggerating its cost.*

Vague escape clauses The transactions between the principal and the agent build on highly structured communication in the mechanisms described above. Those mechanisms describe for each possible cost report θ_j which contract x_j ($j \in \mathcal{A}$) of K specified options the agent shall receive, whether the cost report triggers the escape clause ($j \in \mathcal{B} \neq \emptyset$), or whether the agent shall not produce anything at all ($j \in \mathcal{C} \neq \emptyset$). Such high level of contractual detail can be costly to implement in practice, of which delineating the exact boundaries of the escape clause seems particularly challenging. In practice, such contract stipulations are often formulated in more ambiguous terms. An example of a vaguely stipulated escape clause is:

> *The agent has the right to obtain a new contract offer from the principal if the agent's costs are substantially higher than expected. All initial contract offers by the principal are void if the agent invokes this clause.*

Contrary to the escape clause that forms the foundation of the incentive optimal direct mechanism, the above clause does not state the precise circumstances under which it applies. The ambiguity that arises comes from the adverb "substantially", which is not defined in the contract.[15] The key question so moves to whether detailed communication would add real economic value in this context.

To gauge the value of communication, assume that the incentive optimal mechanism features *complete ex-post pooling* in the sense of $|\mathbf{x}^*_{B^*}| = 1$, so that $x^*_j = x^*_{\underline{B}^*}$ for all $j \in \mathcal{B}^*$. On the basis

[14] An agent with marginal cost $\theta_{\underline{B}^*}$ may understate marginal cost to θ_{A^*} under very particular circumstances; see Lemma 3. If $\underline{B}^* = B^* - 1$ and $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$, the incentive optimal quantity $q^*_{\underline{B}^*}$ is given by the solution to $S'(q^*_{\underline{B}^*}) = \frac{\nu_{\underline{B}^*}\sigma^*_{\underline{B}^*}(\theta_{\underline{B}^*}+(1-\alpha)(\theta_{B^*}-\theta_{\underline{B}^*}))+\nu_{B^*}(1-\sigma^*_{B^*})\theta_{B^*}}{\nu_{\underline{B}^*}\sigma^*_{\underline{B}^*}+\nu_{B^*}(1-\sigma^*_{B^*})}$.

[15] Maggi and Staiger (2011) and Gennaioli and Ponzetto (2017) develop rigorous models of vague contract stipulations and provide examples of vague contract provisions.

of uniform reporting strategies in Lemma 3, the principal forms posterior beliefs of

$$\mu^*_{jA^*} = \mu^*_{A^*} = \frac{\nu_{A^*}(1-\sigma^*_{A^*})}{\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*}}, \ \mu^*_{ji} = \mu^*_i = \frac{\nu_i}{\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*}}, \ \forall i \in \mathcal{B}^*. \tag{20}$$

after receiving a cost report of θ_j, $j \in \mathcal{B}^*$. All cost reports are equally (un)informative since they all yield the same posterior beliefs. The expected surplus to the principal of this mechanism is

$$\sum_{k=1}^{K}\sum_{i\in\mathcal{A}^*_k} \nu_i W_i(x^*_{\mathcal{A}^*_k}) + \nu_{A^*}(1-\sigma^*_{A^*})[W_{A^*}(x^*_{\underline{B}^*}) - W_{A^*}(x^*_{\mathcal{A}^*_K})] + \sum_{i\in\mathcal{B}^*} \nu_i W_i(x^*_{\underline{B}^*}) \tag{21}$$

as a function of the K contracts $\mathbf{x}^*_{\boldsymbol{\mathcal{A}}^*} = (x^*_{\mathcal{A}^*_1}, ..., x^*_{\mathcal{A}^*_k}, ..., x^*_{\mathcal{A}^*_K})$ presented to the agent ex ante, and the ex-post contract $x^*_{\underline{B}^*}$.

Consider now an alternative sequence of events:

Stage 1: The principal commits to the menu $\mathbf{x}^*_{\boldsymbol{\mathcal{A}}^*}$ of contracts, augmented by the vague escape clause (VEC):

> *The agent always has the right to obtain a new contract offer from the principal. All initial contract offers by the principal are void if the agent invokes this clause.*

Stage 2: Depending on its marginal cost θ_i, the agent:

- Selects $x^*_{\mathcal{A}^*_k}$ if $i \in \mathcal{A}^*_k$ and $i < A^*$.
- Selects $x^*_{\mathcal{A}^*_K}$ with probability $\sigma^*_{A^*}$ and invokes VEC with probability $1-\sigma^*_{A^*}$ if $i = A^*$.
- Invokes VEC if $i \in \mathcal{B}^*$.
- Outright rejects the offer if $i \in \mathcal{C}^* \neq \emptyset$.

Stage 3: The principal offers $x^*_{\underline{B}^*}$ if the agent has invoked VEC in Stage 2.

Communication in the above mechanism is restricted in the sense that the agent never directly reports its cost to the principal. The agent either self-selects one of the ex-ante contracts, invokes the escape clause, or completely rejects the offer, after which the game ends. The escape clause here represents an option the agent can exercise after receiving $\mathbf{x}^*_{\boldsymbol{\mathcal{A}}^*}$. The agent does not need to communicate anything to the principal other than its decision. Clearly, this *restricted communication mechanism augmented by VEC* delivers the expected surplus (21) to the principal if the agent and the principal play the sequence of events specified in Stage 2 and in Stage 3.

In Stage 3, the principal does not have any detailed cost reports upon which to form beliefs after the agent has invoked the escape clause. Instead, the principal uses the likelihood by which the agent invokes the escape clause, given its marginal cost, to derive posterior beliefs of

$$\mu^{VEC}_{A^*} = \frac{\nu_{A^*}(1-\sigma^*_{A^*})}{\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*}}, \ \mu^{VEC}_i = \frac{\nu_i}{\nu_{A^*}(1-\sigma^*_{A^*}) + \nu_{\mathcal{B}^*}}, \ \forall i \in \mathcal{B}^*.$$

concerning the agent's marginal cost. This is precisely the same distribution of beliefs (20) as in the incentive optimal direct mechanism. Hence, $x^*_{\underline{B}^*}$ is sequentially rational in Stage 3.

In Stage 2, all cost types $i \in \mathcal{I}$, except possibly A^*, obtain $U_i(x_i^*)$. Rationality for these cost types then follows from incentive compatibility (2) of $\mathbf{x}^*$. An agent with marginal cost θ_{A^*} either selects $x^*_{\mathcal{A}^*_K}$ and obtains $U_{A^*}(x^*_{A^*})$ or invokes VEC and obtains $U_{A^*}(x^*_{\underline{B}^*})$. Randomizing among the two is rational by (3). Hence, the behaviors in Stage 2 and Stage 3 form a PBE.

Communication has economic value if the incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|A^*, B^*)$ features *partial ex-post pooling*, $|\mathbf{x}^*_{\underline{B}^*}| = 2$. The principal now draws different inferences about the agent's true marginal cost, depending on whether the agent reports marginal cost θ_{B^*} or some other marginal cost θ_j, $j \in \underline{\mathcal{B}}^*$. Moreover, sequential rationality of a reduced communication mechanism does not follow from incentive feasibility of the incentive optimal mechanism.

Proposition 1 *Assume that the incentive optimal (direct) 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 $\nu_{B^*} \to 0$ the same expected surplus to the principal as the incentive optimal (direct) mechanism.*

Proposition 1 shows that the value of communication is small in the context of this model. For instance, it is negligible in a large type space (so that ν_i is small for all $i \in \mathcal{I}$). A policy implication is that for the initial menu of contracts $\mathbf{x}^*_{\mathcal{A}^*}$, the principal has little to gain from specifying a detailed escape clause. The vague escape clause (VEC) formulated as an option on behalf of the agent can do nearly as well in equilibrium.

Effectively, the properties of the ex ante contracts determine the content of the escape clause. Suppose the principal offers one single contract $x = (t, q)$ up front. Absent any escape clause, $\frac{t}{q}$ defines the threshold marginal cost for accepting this contract if $\theta_1 < \frac{t}{q} < \theta_I$. Under a VEC, the two equations

$$t - \theta_A q = (\theta_I - \theta_A) q^{VEC}$$

and

$$S'(q^{VEC}) = \sum_{i=A+1}^{I} \frac{\nu_i}{1 - G_A} (\theta_i + (1-\alpha)(\theta_I - \theta_i))$$

characterize the lower cut-off θ_A for activating the escape clause and the ex-post quantity q^{VEC} as functions of x in a full participation mechanism. The first expression is an incentive compatibility constraint that renders an agent with marginal cost θ_A indifferent between x and the anticipated ex-post contract $x^{VEC} = (q^{VEC}, \theta_I q^{VEC})$. The second expression characterizes the sequentially rational output in the ex-post contract. The contract x augmented by VEC can potentially sustain multiple equilibria. However, this is also the case in the direct mechanism. The methodology applied in this paper assumes that the mechanism selects the equilibrium that maximizes the expected surplus of the principal.

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. To answer this question in the affirmative and gain additional intuition about the nature of incentive optimal contracts under endogenously incomplete commitment, we now consider some specific examples of the more general model.

5 Severely constrained contracting

Consider the polar extreme case where the principal can offer at most one ex-ante contract: $K = 1$. 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 by providing insights into the relative merits of offering contracts ex ante relative to ex post. In the first case, the principal commits to one single contract $x = (q, t)$. An agent with marginal cost $\theta_i \leq \frac{t}{q}$ accepts this contract, but rejects it if less efficient. Under pure ex-post contracting, the principal does not propose any contract up front. Instead, the principal states an upper bound θ_B 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. 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 prefers pure ex-post over pure ex-ante contracting for $K = 1$.

The surplus-maximizing ex-ante contract By the results in Section 3, the incentive efficient pure ex-ante contract $\hat{x}_A = (\hat{q}_A, \theta_A \hat{q}_A)$ that yields production for all cost types $i \in \{1, ..., A\}$ and no production for the less efficient cost types $i \in \{A + 1, ..., I\}$ under partial participation, $A \leq I - 1$, has output requirement $\hat{q}_A$ given by

$$S'(\hat{q}_A) = \sum_{i=1}^{A} \frac{\nu_i}{G_A}(\theta_i + (1 - \alpha)(\theta_A - \theta_i)). \tag{22}$$

The principal's expected surplus of this contract equals:

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

Denote by $\hat{x}_{\hat{A}}$ the contract offer that maximizes the principal's expected surplus across all ex ante mechanisms under $K = 1$, and assume that the maximum is unique:

$$\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}}) > \tilde{W}_A(\hat{q}_A) \ \forall A \in \mathcal{I} \cup \emptyset, \ A \neq \hat{A}. \tag{23}$$

Incentive feasible ex-post contracts Assume that ex-post contracting occurs for all cost reports $j \in \{1, ..., B\}$, $B \geq 2$. In light of the results in Section 4, we consider a menu of two ex post contracts $(x_{\underline{B}}, x_B^{fb})$. All cost reports θ_j, $j \in \{1, ..., B - 1\}$, yield the same contract offer $x_{\underline{B}} = (q_{\underline{B}}, \theta_B q_{\underline{B}})$, where $q_{\underline{B}}$ is characterized by

$$S'(q_{\underline{B}}) = \sum_{i=1}^{B} \frac{\nu_i}{G_B - \nu_B \sigma_B}(\theta_i + (1 - \alpha)(\theta_B - \theta_i)) - \frac{\nu_B \sigma_B \theta_B}{G_B - \nu_B \sigma_B} < \theta_B = S'(q_B^{fb}). \tag{24}$$

In this expression, $\sigma_B \in (0,1)$ is the probability that an agent with marginal cost θ_B truthfully reports its cost, whereas this agent reports marginal cost θ_j with probability $\frac{1-\sigma_B}{B-1}$ for all $j < B$. An agent with marginal cost $\theta_i < \theta_B$ randomizes uniformly across all cost reports θ_j, $j \in \{1, ..., B-1\}$. If $B \leq I-1$, then an agent with marginal cost $\theta_i > \theta_B$ truthfully reports marginal cost and receives the null contract. By comparison of (24) with (22), we see that the ex-post quantity $q_{\underline{B}}$ converges to $\hat{q}_B$ when $\sigma_B \to 0$. The ex-post quantity increases as the agent becomes more truthful,

$$\frac{dq_{\underline{B}}}{d\sigma_B} = \frac{-\alpha\nu_B}{S''(q_{\underline{B}})} \frac{\sum_{i=1}^{B-1} \nu_i(\theta_B - \theta_i)}{(G_B - \nu_B\sigma_B)^2} > 0$$

because the perceived marginal cost of an agent that reports $\theta_j < \theta_B$ is smaller when σ_B is larger.

The expected surplus of the principal equals

$$\begin{aligned}\tilde{\Omega}_B(q_{\underline{B}}, \sigma_B) &= \sum_{i=1}^{B-1} \nu_i[S(q_{\underline{B}}) - (\theta_i + (1-\alpha)(\theta_B - \theta_i))q_{\underline{B}}] + \nu_B[(1-\sigma_B)W_B^{fb}(q_{\underline{B}}) + \sigma_B w_B^{fb}] \\ &= \tilde{W}_B(q_{\underline{B}}) + \nu_B\sigma_B[w_B^{fb} - W_B^{fb}(q_{\underline{B}})]\end{aligned}$$

in the ex-post mechanism. The principal benefits from the agent being more truthful,

$$\frac{d\tilde{\Omega}_B(q_{\underline{B}}, \sigma_B)}{d\sigma_B} = \nu_B[w_B^{fb} - W_B^{fb}(q_{\underline{B}})] > 0,$$

because an agent with marginal cost θ_B is more likely to receive a contract better suited to its particular circumstances if σ_B is larger. The marginal effect on $q_{\underline{B}}$ of an increase in σ_B has only a second-order effect on the principal's expected surplus.

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

$$\frac{d}{d\sigma_B} \sum_{i=1}^{B} \nu_i U_i(x_{\underline{B}}) = \sum_{i=1}^{B-1} \nu_i(\theta_B - \theta_i) \frac{dq_{\underline{B}}}{d\sigma_B} > 0.$$

The agent is indifferent between truthfully reporting its cost θ_B and understating it to $\theta_j < \theta_B$, all else equal. However, the agent benefits from the indirect effect $\frac{dq_{\underline{B}}}{d\sigma_B} > 0$ because the higher output increases informational rent whenever the agent has marginal cost $\theta_i < \theta_B$. Both the principal and the agent therefore agree ex ante that more truthful behavior would be better under pure ex-post contracting.

There is an upper bound to the truthfulness that the principal can implement. If an agent with marginal cost θ_B is too honest, then it can become sequentially rational for the principal to exclude this cost type after receiving a cost report $\theta_j < \theta_B$. Doing so would then allow the principal to save on informational rent without sacrificing too much efficiency. We denote by $x_A^d = (q_A^d, \theta_A q_A^d)$ the surplus maximizing ex-post contract that just leaves an agent with marginal cost θ_A, $A \in \{1, ...B-1\}$, indifferent between accepting or rejecting the ex-post contract offer. Under pure ex-post contracting, this deviation contract is given by $x_A^d = \hat{x}_A$. The expected deviation surplus is then proportional to $\tilde{W}_A(\hat{q}_A)$. Instead, the expected surplus of offering $x_{\underline{B}}$

is proportional to $\tilde{W}_B(q_{\underline{B}}) - \nu_B\sigma_B W_B^{fb}(q_{\underline{B}})$. Hence, $x_{\underline{B}}$ is sequentially rational if and only if

$$\tilde{W}_B(q_{\underline{B}}) - \nu_B\sigma_B W_B^{fb}(q_{\underline{B}}) \geq \tilde{W}_A(\hat{q}_A),\ \forall A \in \{1, ...B-1\}. \tag{25}$$

Conversely,

$$\begin{aligned}&\tilde{W}_{B-1}(\hat{q}_{B-1}) + \nu_B W_B^{fb}(q_{\underline{B}}) - \tilde{W}_B(q_{\underline{B}})\\ &\quad = \tilde{W}_{B-1}(\hat{q}_{B-1}) - \tilde{W}_{B-1}(q_{\underline{B}}) + G_{B-1}(1-\alpha)(\theta_B - \theta_{B-1})q_{\underline{B}} > 0\end{aligned}$$

implies that there is an upper bound to the agent's honesty σ_B in the sequentially rational ex post contract, as argued.

Comparison of ex-post and ex-ante contracting Assume that $\hat{A} \geq 2$ and compare the incentive efficient ex-ante contract $\hat{x}_{\hat{A}}$ to the menu $(x_{\underline{\hat{A}}}, x_{\hat{A}}^{fb})$ of ex-post contracts ($B = \hat{A}$):

$$\tilde{\Omega}_{\hat{A}}(q_{\underline{\hat{A}}}, \sigma_{\hat{A}}) - \tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}}) = \tilde{\Omega}_{\hat{A}}(q_{\underline{\hat{A}}}, \sigma_{\hat{A}}) - \tilde{\Omega}_{\hat{A}}(\hat{q}_{\hat{A}}, \sigma_{\hat{A}}) + \nu_{\hat{A}}\sigma_{\hat{A}}[w_{\hat{A}}^{fb} - W_{\hat{A}}^{fb}(\hat{q}_{\hat{A}})] > 0\ \forall\sigma_{\hat{A}} > 0. \tag{26}$$

In the above expression, $\tilde{\Omega}_{\hat{A}}(q_{\underline{\hat{A}}}, \sigma_{\hat{A}}) > \tilde{\Omega}_{\hat{A}}(\hat{q}_{\hat{A}}, \sigma_{\hat{A}})$ because $q_{\underline{\hat{A}}}$ represents a better trade-off between efficiency and rent extraction than $\hat{q}_{\hat{A}}$, given $\sigma_{\hat{A}}$. In addition, ex-post contracting allows to supply a tailor-made contract to an agent with marginal cost $\theta_{\hat{A}}$. Seeing as $\hat{x}_{\hat{A}}$ maximizes the principal's expected surplus across all incentive feasible pure ex-ante contracts, we conclude that $(x_{\underline{\hat{A}}}, x_{\hat{A}}^{fb})$ strictly outperforms all pure ex-ante contracts.

Still, we have not yet established sequential rationality of $x_{\underline{\hat{A}}}$. To do so, we reproduce the necessary and sufficient condition (25) with the appropriate change in notation:

$$\tilde{W}_{\hat{A}}(q_{\underline{\hat{A}}}) - \nu_{\hat{A}}\sigma_{\hat{A}} W_{\hat{A}}^{fb}(q_{\underline{\hat{A}}}) \geq \tilde{W}_A(\hat{q}_A),\ \forall A \in \{1, ...\hat{A}-1\}.$$

The left-hand side of this expression converges to $\tilde{W}_{\hat{A}}(\hat{q}_{\hat{A}})$ as $\sigma_{\hat{A}} \to 0$ because then $q_{\underline{\hat{A}}} \to \hat{q}_{\hat{A}}$. By way of (23), it follows that $x_{\underline{\hat{A}}}$ is sequentially rational for $\sigma_{\hat{A}} > 0$ sufficiently close to zero. We can then conclude:

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

Under severely constrained contracting, ex-post contracts generally offer superior adaptability to the economic environment compared to a pure ex-ante contract, despite strategic manipulation of cost reports by the agent.[16] Since under ex-post contracting the principal can always add an ex-ante contract without reducing expected surplus, the following result follows directly:

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

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

6 Constrained contracting

So far we have established the general incentive optimality of introducing an escape clause if contracting is severely constrained in the sense that the principal offers a one-size-fits-all contract under ex-ante contracting ($K = 1$). The incremental 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 include more contingencies into the menu of contracts already beforehand. However, it is reasonable to assume that the principal might have insufficient degrees of freedom to be able to include all potential contingencies ex ante. This occurs for any K if the type space is sufficiently large. This plausible scenario leads to the question whether escape clauses are incentive optimal for more complex mechanisms $K \geq 1$? The next result establishes a simple sufficient condition for this to be the case:

Lemma 4 *Denote by $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ the mechanism that maximizes the principal's expected surplus in the set of mechanisms with complete commitment ($\mathcal{B} = \emptyset$). Let $\hat{q}_{\hat{A}}$ be the output of the least efficient agent that delivers positive output in this mechanism. The incentive optimal mechanism $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ features incomplete commitment ($\mathcal{B}^* \neq \emptyset$) if $\hat{q}_{\hat{A}} > \hat{q}^{fb}_{\hat{A}}$.*

Proof. If $|\hat{\mathcal{A}}_K| = 1$, then $\hat{q}_{\hat{A}} = q^{sb}_{\hat{A}} < q^{fb}_{\hat{A}}$; see Section 3. Hence, $\hat{q}_{\hat{A}} > \hat{q}^{fb}_{\hat{A}}$ 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 $\mathcal{B} = \hat{A}$. $\mathbf{x}$ has the following properties: $x_j = (\hat{q}_j, \hat{t}_j - (\theta_{\hat{A}} - \theta_{\hat{A}-1})(\hat{q}_{\hat{A}} - q^{fb}_{\hat{A}}))$ for all $j \in \mathcal{A}$ and $x_{\hat{A}} = x^{fb}_{\hat{A}}$. We first check incentive feasibility of the modified mechanism. Individual rationality (1) and incentive compatibility (2) follow from

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

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

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

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

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

This mechanism trivially satisfies (3) because all types truthfully report cost with probability 1. $x^{fb}_{\hat{A}}$ is sequentially rational (4) because the only type that reports $\theta_{\hat{A}}$ is an agent with marginal cost $\theta_{\hat{A}}$. By truthfulness, the posterior probabilities (5) are $\mu_{jj} = 1$ for all $j \in \mathcal{I}$. The mechanism satisfies the contracting constraint (6) 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. ■

Pooling a subset of cost types into cost group $\hat{\mathcal{A}}_K$ implies that the output of the least efficient cost type $\hat{A}$ is inefficiently high from a second-best perspective, $\hat{q}_{\hat{A}} > q_{\hat{A}}^{sb}$; see Lemma 1. This output could potentially be upward distorted even compared to the first-best output, $\hat{q}_{\hat{A}} > q_{\hat{A}}^{fb}$. Even a principal that completely discards informational rent in the choice of $q_{\hat{A}}$ would then from an ex ante perspective prefer the first-best efficient contract for an agent with marginal cost $\theta_{\hat{A}}$. One way to reach this efficiency gain is to include an escape clause in the mechanism.

To illustrate the usefulness of Lemma 4, subtract (13) from $S'(q_{\hat{A}}^{fb}) = \theta_{\hat{A}}$ to get:

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

. The first term on the right-hand side is the effect of pooling cost types into a cost group. The second term is the adjustment for informational rent. The pooling effect tends to increase and the rent effect tends to reduce $\hat{q}_{\hat{A}}$ relative to $q_{\hat{A}}^{fb}$. The rent effect is zero by $G_{\hat{A}_{K-1}} = 0$ if contracting is severely constrained ($K = 1$). It vanishes also in the limit as $\alpha \to 1$ because rent extraction plays a negligible role for 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).*

It is sometimes the case that the principal can offer more complex contracts than one-size-fits-all. It is also reasonable to assume that the principal sometimes places a lot of weight on rent extraction in the design of the mechanism. Note that the above results do not apply when $K \geq 2$ and α is small. However, there are still plausible circumstances under which the incentive optimal mechanism features incomplete commitment:

Proposition 5 *Assume that the mechanism $(\hat{\mathbf{x}},\mathbf{I}|\hat{\mathcal{A}},\emptyset)$ that maximizes the principal's expected surplus in the set of mechanisms with complete commitment features partial participation ($\hat{A} \leq I-1$). Assume also that the incremental difference in marginal production costs is small for boundary cost types ($\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$ is close to zero). The incentive optimal mechanism contains an escape clause if $W_{\hat{A}}^{sb}(q_{\hat{A}}^{fb}) > 0$.*

Proof. If $\hat{A} \leq I-1$ and $\theta_{\hat{A}+1} - \theta_{\hat{A}-1}$ is close to zero, then $W^{sb}_{\hat{A}}(\hat{q}_{\hat{A}}) \approx W^{sb}_{\hat{A}+1}(\hat{q}_{\hat{A}}) \approx 0$; see (18). If also $W^{sb}_{\hat{A}}(q^{fb}_{\hat{A}}) > 0$, then $W^{sb}_{\hat{A}}(q^{sb}_{\hat{A}}) > W^{sb}_{\hat{A}}(q^{fb}_{\hat{A}}) > W^{sb}_{\hat{A}}(\hat{q}_{\hat{A}})$. Strict concavity of $W^{sb}_{\hat{A}}(q)$, $q^{fb}_{\hat{A}} > q^{sb}_{\hat{A}}$ and $\hat{q}_{\hat{A}} > q^{sb}_{\hat{A}}$ (Lemma 1) then imply $\hat{q}_{\hat{A}} > q^{fb}_{\hat{A}}$. ■

If the type space is large and the mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ features partial participation, then the principal is indifferent between including an agent with marginal cost θ_A in the mechanism and excluding the agent, i.e. $W^{sb}_{\hat{A}}(\hat{q}_{\hat{A}}) = 0$. It is then worthwhile to include this agent via an escape clause if $W^{sb}_{\hat{A}}(q^{fb}_{\hat{A}}) > 0$.

This section and the previous have established circumstances under which the incentive optimal mechanism features incomplete commitment in terms of an escape clause. 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 delineate the boundaries of the escape clause in incentive optimal mechanisms, we assume that the number I of potential cost realizations is large, so that the probability ν_i of any single cost realization θ_i is small. In this case, truth-telling by an agent with marginal cost θ_{A^*}, uniform randomization across $\mathcal{B}^*$ also for an agent with marginal cost θ_{B^*}, and a single ex-post contract $x^*_{\underline{B}^*}$ form part of an (approximately) incentive optimal mechanism with incomplete commitment. This mechanism delivers expected surplus

$$w^* = \sum_{i \in \mathcal{A}^*} \nu_i W_i(x^*_i) + \Omega_{\mathcal{B}^*}(q^*_{\underline{B}^*}).$$

to the principal. 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 $x^*_{\underline{B}^*} = (q^*_{\underline{B}^*}, \theta_{B^*} q^*_{\underline{B}^*})$.

Assume that $|\mathcal{B}^*| \geq 2$. Compare the expected surplus w^* to a modified mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ in which the boundary type $\underline{B}^*$ is included in the least efficient cost group, $\mathcal{A}_K = \mathcal{A}^*_K \cup \underline{B}^*$, and the escape clause is correspondingly reduced to $\mathcal{B} = \mathcal{B}^* \backslash \underline{B}^*$. All other cost groups remain the same as before, $\mathcal{A}_k = \mathcal{A}^*_k$ for all $k \in \{1, ...K-1\}$ if $K \geq 2$. The modification of the escape clause reduces ex-post output $q_{\underline{B}} < q^*_{\underline{B}^*}$ 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. The ex-ante contracts are modified as follows. Every contract $x_{\mathcal{A}_k} = (q^*_{\mathcal{A}^*_k}, t_{\mathcal{A}_k})$, $k \in \{1, ...K\}$, has the same output requirement as in the incentive optimal contract, and all transfer payments are adjusted by the same amount: $t_{\mathcal{A}_k} - t^*_{\mathcal{A}^*_k} = (\theta_{\underline{B}^*} - \theta_{A^*})(q^*_{A^*} - q^*_{\underline{B}^*}) - (\theta_{B^*} - \theta_{\underline{B}^*})(q^*_{\underline{B}^*} - q_{\underline{B}})$.[17]

We can decompose the net benefit to the principal of the incentive optimal mechanism over

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

the modified one into three separate effects:

$$w^* - W(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B}) = \nu^*_{\underline{B}}[W^{sb}_{\underline{B}^*}(q^*_{\underline{B}^*}) - W^{sb}_{\underline{B}^*}(q^*_{A^*})] \\ - G_{\underline{B}^*}(1-\alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q^*_{\underline{B}^*} - q_{\underline{B}}) - [\Omega_{\mathcal{B}}(q_{\underline{B}}) - \Omega_{\mathcal{B}}(q^*_{\underline{B}^*})].$$

The first term on the right-hand side is the effect on the principal's expected surplus of an agent with marginal cost $\theta_{\underline{B}^*}$ producing ex-post output $q^*_{\underline{B}^*}$ instead of ex-ante output $q^*_{A^*}$, evaluated on the basis of the second-best welfare function $W^{sb}_{\underline{B}^*}(q)$. Ex-post output $q^*_{\underline{B}^*}$ is larger under escape clause $\mathcal{B}^*$ compared to $q_{\underline{B}}$ under the smaller escape clause $\mathcal{B} = \{\underline{B}^*+1, ..., B^*\}$ 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 informational rent is sunk. The dynamic inefficiency of a larger escape clause is measured by the first negative term on the second row above. The final effect is the inefficiency of output $q^*_{\underline{B}^*}$ relative to $q_{\underline{B}}$ under the ex-post welfare function $\Omega_{\mathcal{B}}(q)$.[18]

A mechanism with complete commitment requires $W^{sb}_{\underline{B}^*}(q^*_{\underline{B}^*}) \geq W^{sb}_{\underline{B}^*}(q^*_{A^*})$ for efficiency; see (17). However, a mechanism is incentive optimal only if:

$$W^{sb}_{\underline{B}^*}(q^*_{\underline{B}^*}) - W^{sb}_{\underline{B}^*}(q^*_{A^*}) \geq \frac{G_{\underline{B}^*}}{\nu^*_{\underline{B}}}(1-\alpha)(\theta_{B^*} - \theta_{\underline{B}^*})(q^*_{\underline{B}^*} - q_{\underline{B}}).$$

We interpret the left-hand side of this inequality as the *marginal benefit of flexibility* associated with an agent that has marginal cost $\theta_{\underline{B}^*}$ operating under 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 marginal benefit must be sufficiently large to outweigh the *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 marginal benefit of a better adapted contract against the marginal increase in dynamic inefficiency.*

8 Discussion

Contract complexity We have interpreted K literally as the number of contracts contained in the ex-ante menu offered to the agent. Many real-life mechanisms have this property. 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 $x(\theta) = (q(\theta), t(\theta))$ with the property that all $x_{\mathcal{A}_k}$, $k \in \{1, ..., K\}$, lie somewhere on $x(\theta)$. With this continuous formulation, it is not meaningful to discuss constrained contracting in terms of the number of contracts.

[18]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 say that a local incentive compatibility constraint $U_i(x_i) = U_i(x_{i+1})$ is *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. 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 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 here briefly explore a different avenue to explain $K < I$. Assume now that the number K of contracts is endogenous, but 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 the agent actually invokes this contract at a later stage, 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, $j \in \mathcal{B}$.

This 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 is likely to be large relatively to the cost of writing the contract. If so, contracts should be close to being complete (K is close to I in this setting). From that perspective, costs of writing contracts cannot explain the prevalence of incomplete contracting.

We subscribe to the idea that adding an arbitrary contract to an initial menu of contracts is unlikely to be very costly. However, not all contract additions will generate economic value to the principal. In our setting, any additional contract must be incentive compatible relative to the initial menu of contracts. Second, the incremental contract must increase the principal's expected surplus relative to the initial menu. Identifying an incentive compatible, surplus increasing 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 incremental value smaller as the number of initial contracts is larger. Hence, we assume that contracting costs are non-negligible. Still, we will characterize circumstances under which the principal would constrain the number of ex-ante contracts and include an escape clause rather than increase the number of ex-ante contracts, even for small $C > 0$.

Let us consider the simplest possible framework with two cost types: $\mathcal{I} = \{1, 2\}$. The principal has four options under pure ex-ante contracting. The first is a single contract $\hat{x}_2 = (\hat{q}_2, \theta_2\hat{q}_2)$ that is acceptable to the agent regardless of its marginal cost. This mechanism yields expected surplus

$$\tilde{W}_2(\hat{q}_2) = \nu_1 W_1^{fb}(\hat{q}_2) + \nu_2 W_2^{sb}(\hat{q}_2) - C,$$

where the output $\hat{q}_2 > q_2^{fb}$ is characterized by $S'(\hat{q}_2) = \nu_1(\theta_1 + (1-\alpha)(\theta_2 - \theta_1)) + \nu_2\theta_2$. We assume that contracting costs are small relative to the value of contracting in the sense that $W_2^{sb}(\hat{q}_2) > C$. The second option is a single contract that only the most efficient agent will accept:

$$\tilde{W}_1(q_1^{fb}) = \nu_1 w_1^{fb} - C.$$

The third option is to supply the second-best mechanism at the expense of increased contracting costs:

$$w^{sb} = \nu_1 w_1^{fb} + \nu_2 w_2^{sb} - 2C.$$

The fourth option, null contracting, is dominated by the first option by the assumption of small contracting costs.

Consider now the mechanism with incomplete commitment. The principal offers the contract $x_1 = (q_1^{fb}, \theta_1 q_1^{fb} + (\theta_2 - \theta_1)q_2^{fb})$ up front. The agent receives this contract by reporting marginal cost θ_1. The agent invokes the escape clause by reporting marginal cost θ_2, after which the principal offers the ex-post contract x_2^{fb}. The agent truthfully reports its cost even in this case. This mechanism is incentive feasible and yields an expected surplus of

$$w_B = \nu_1 w_1^{fb} - C + \nu_2[W_2^{sb}(q_2^{fb}) - C].$$

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

$$w_B - w^{sb} = \nu_1 C - \nu_2[w_2^{sb} - W_2^{sb}(q_2^{fb})].$$

On the one hand, the principal reduces expected contracting costs by including an escape clause in the mechanism. On the other, the second-best contract x_2^{sb} offers a better trade-off between efficiency and rent extraction from an ex-ante perspective than the discretionary contract x_2^{fb} when the agent is inefficient. Importantly, 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 value of contracting is large. For instance, reduced contracting costs dominate increased contractual efficiency for arbitrary $C > 0$ if the likelihood of a high cost event is small, i.e. ν_2 is small. 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_1 is sufficiently close to zero. The net benefit to the principal of second-best relative to first-best contracting is small if $\nu_1(1-\alpha)$ is close to zero because then the principal mainly cares about efficiency in the choice of q_2^{sb}.[19]

The mechanism with an escape clause beats the ex-ante mechanism with production only by

[19]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 $\lim_{\nu_1 \to 0} \frac{w_2^{sb} - W_2^{sb}(q_2^{fb})}{\nu_1} = \lim_{\nu_1 \to 0}[S'(q_2^{sb}) - \theta_2]\frac{dq^{sb}}{d\nu_1} = 0$ and therefore $\lim_{\nu_1 \to 0} \frac{w_B - w^{fb}}{\nu_1} = C > 0$.

the efficient agent, $w_B > \tilde{W}_1(q_1^{fb})$, as $W^{sb}(q_2^{fb}) > W^{sb}(\hat{q}_2) > C$. However, it does not necessarily beat the pure ex-ante contract $\hat{x}_2$. The difference

$$w_B - \tilde{W}_2(\hat{q}_2) = \nu_1[w_1^{fb} - W^{fb}(\hat{q}_2)] + \nu_2[W_2^{sb}(q_2^{fb}) - W^{sb}(\hat{q}_2) - C].$$

in expected surplus can be positive or negative, depending on the circumstances. It is positive if ν_2 is small or if α is close to one and $C < w_2^{fb} - W^{fb}(\hat{q}_2)$. It is not optimal to modify $\hat{x}_1$ by adding an escape clause if ν_1 is small.[20]

Remark 2 *Constrained contracting ($K < I$) can be justified even on the basis of small contracting costs, for instance 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 $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathcal{A}}, \emptyset)$ that maximizes the principal's expected surplus in the set of incentive feasible mechanisms with complete commitment. The general message of the paper is that incentive feasible mechanisms sometimes exist that strictly improve upon the complete commitment benchmark under constrained contracting. All such improvements must necessarily involve some form of incomplete commitment.

We have interpreted incomplete commitment as the inclusion of an escape clause that the agent can trigger by reporting marginal cost θ_j, $j \in \mathcal{B}$, where the subset $\mathcal{B}$ is specified in the mechanism offered to the agent at the initial stage of interaction. All initial contract offers are void if the agent invokes the escape clause. 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, whereas the ex-post contract merely is required to outperform the outside option under the escape clause.[21]

Under the escape clause, the value of the agent's outside option is zero, regardless of the agent's marginal cost. In turn, the value of the outside option is type dependent and therefore private information under the renegotiation clause. 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, $j \in \mathcal{B} = \{A+1, ...B\}$, $B \geq A+1$. Finally,

[20] However, pure ex-post contracting always yields strictly higher expected surplus than the pure ex-ante contract $\hat{x}_2$ by Proposition 2. Adding contractual costs to the equations does not matter for the comparison in (26) because the expected contracting cost equals C in either mechanism under full participation. Under partial participation, ex-post contracting not only is more efficient but also reduces the expected contracting cost under severely constrained contracting.

[21] 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).

the agent receives the null contract for all cost reports θ_j, $j \in \mathcal{C} = \{B+1, ...I\}$ if $B \leq I-1$. Suppose an agent with marginal cost θ_i, $i \in \mathcal{B}$, has invoked the renegotiation clause. This agent will accept the ex-post contract x_j if and only if

$$U_i(x_j) \geq U_i(x_A).$$[22]

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 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 escape clauses. 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, in the sense that 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 it might be in a principal's best interest to allow some discretion when it comes to future contracting, even if the principal has access to a very general reward structure with which to incite agent behavior. In a setting where the principal cannot cover every possible pay-off relevant contingency by an ex ante contract, the added flexibility associated with ex-post contracting can be sufficiently valuable to dominate the dynamic inefficiency associated with discretionary contracting. The principal constrains its own incentive to abuse the escape clause by delegating the choice whether to activate the clause to the agent.

Many contractual arrangements feature endogenously incomplete commitment, even if not always an escape clause. A prime example is multi-period contracting. In regulatory and service procurement agreements, optimal contract length is a major design issue. A longer-term agreement implies stronger commitment, whereas a sequence of shorter-term agreements means less commitment. It would be interesting to analyze the trade-off between flexibility and dynamic efficiency also in a multi-period context.

[22]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 $(i,h) \in \mathcal{B} \times \mathcal{A}$.

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.

Athey, Susan, Andrew Atkeson, and Patrick J. Kehoe, "The optimal degree of discretion in monetary policy," *Econometrica*, 2005, *73* (5), 1431–1475.

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.

Baron, David P. and Roger B. Myerson, "Regulating a monopolist with unknown costs," *Econometrica*, 1982, *50* (4), 911–930.

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

Bergemann, Dirk and Juuso Välimäki, "Dynamic mechanism design: An introduction," *Journal of Economic Literature*, 2019, *57* (2), 235–74.

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

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," 2020. Mimeo, Queensland University.

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," 2021. Mimeo, Columbia Business School.

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

Fiocco, Raffaele and Roland Strausz, "Consumer standards as a strategic device to mitigate ratchet effects in dynamic regulation," *Journal of Economics & Management Strategy*, 2015, *24* (3), 550–569.

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.

Gennaioli, Nicola and Giacomo Ponzetto, "Optimally vague contracts and the law," 2017. Mimeo, Bocconi University.

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.

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.

Kumar, Praveen, "Consistent mechanism design and the noisy revelation principle." PhD dissertation, Stanford University 1985.

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.

Maggi, Giovanni and Robert W. Staiger, "The role of dispute settlement procedures in international trade agreements," *Quarterly Journal of Economics*, 2011, *126* (1), 475–515.

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

We first prove some general results concerning incentive feasible mechanisms in this specific context, and then provide some characterizations of incentive efficient mechanisms with complete commitment. In particular, these claims verify that the standard conditions of locally downward-binding incentive compatibility constraints, individual rationality of the least efficient cost type and output declining in cost apply also to the present setting.

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

$$U_I(x_I) \geq 0, \tag{27}$$

$$U_i(x_i) \geq U_i(x_{i+1}) \ \forall i \in \{1, ..., I-1\}, \tag{28}$$

$$U_i(x_i) \geq U_i(x_{i-1}) \ \forall i \in \{2, ..., I\}, \tag{29}$$

$$q_i \geq q_{i+1} \ \forall i \in \{1, ..., I-1\}, \tag{30}$$

Proof. Necessity of (27)-(29) 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

$$U_i(x_i) - U_i(x_j) = \sum_{h=i}^{j-1} [U_h(x_h) - U_h(x_{h+1}) + (\theta_{h+1} - \theta_h)(q_{h+1} - q_j)] \geq 0 \ \forall i \in \{1, ..., I-1\}, \tag{31}$$

where non-negativity follows from the assumptions of local downward incentive compatibility (28) and monotonicity (30). The net benefit of truthfully reporting cost θ_i relative to understating it to θ_j, $j \in \{1, ..., i-1\}$, equals

$$U_i(x_i) - U_i(x_j) = \sum_{h=j+1}^{i} [U_h(x_h) - U_h(x_{h-1}) + (\theta_h - \theta_{h-1})(q_j - q_{h-1})] \geq 0 \ \forall i \in \{2, ..., I\}, \quad (32)$$

where non-negativity follows from the assumptions of local upward incentive compatibility (29) and monotonicity (30). Individual rationality (1) then follows from

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

■

Claim 2 *Let* $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ *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 (31), the net benefit of truthfully reporting cost θ_i, relative to exaggerating it to θ_j, $j \in \{i+1, ..., I\}$, satisfies

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

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

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

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

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 (32), 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

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

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

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 (2) 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 $x_i = x_j$. ∎

We now turn to incentive efficient mechanisms under complete commitment.

Claim 5 *A complete commitment mechanism* $(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \emptyset)$ *with* $\hat{K}$ *cost groups is incentive efficient only if*

$$[U_{A_k}(\hat{x}_{A_k}) - U_{A_k}(\hat{x}_{A_k+1})][U_{A_k+1}(\hat{x}_{A_k+1}) - U_{A_k+1}(\hat{x}_{A_k})] = 0 \; \forall k \in \{1,...,\hat{K}-1\}, \; \hat{K} \geq 2. \quad (33)$$

Equation (33) holds also for $k = \hat{K}$ *if* $\hat{K} \leq I - 1$.

Proof. Suppose both the local IC constraints are slack for some $k \in \{1, ..., \hat{K}\}$. By Claim 2, the downward IC constraints are slack for all cost types θ_i and cost reports θ_j, $(i, j) \in \{1, ..., A_k\} \times \{A_k + 1, ..., I\}$ as well. Hence, $\hat{\sigma}_{ji} = 0$ for all those combinations. By Claim 3, the upward IC constraints are slack for all cost types θ_i and cost reports θ_j, $(i, j) \in \{A_k + 1, ..., I\} \times \{1, ..., A_k\}$. Hence, $\hat{\sigma}_{ji} = 0$ even for all these combinations. A marginal reduction in the transfer payment $\hat{t}_j$ by the same amount for all types $j \in \{1, ..., A_k\}$ then increases the principal's expected surplus while maintaining incentive feasibility. Then the proposed mechanism cannot be incentive efficient. ■

Claim 6 *A complete commitment mechanism* $(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \emptyset)$ *with* $\hat{K}$ *cost groups is incentive efficient only if*

$$U_{A_k}(\hat{x}_{A_k}) = U_{A_k}(\hat{x}_{A_k+1}) \ \forall k \in \{1, ..., \hat{K} - 1\}, \ \hat{K} \geq 2, \ U_A(\hat{x}_A) = 0. \quad (34)$$

Proof. We first show that $U_{A_k+1}(\hat{x}_{A_k+1}) > U_{A_k+1}(\hat{x}_{A_k})$ for all $k \in \{1, ..., \hat{K} - 1\}$ if $\hat{K} \geq 2$ and for $k = \hat{K}$ if $\hat{K} \leq I - 1$. Suppose instead the local upward IC constraint is binding for some k. Then the local downward IC constraint in (33) is slack by $\hat{q}_{A_k} > \hat{q}_{A_k+1}$. An agent with marginal cost equal to or below θ_{A_k} will strictly prefer to truthfully report its cost rather than exaggerate it to θ_{A_k+1} or above, by Claim 2. By $\hat{q}_{A_k} > \hat{q}_{A_k+1}$ and Claim 3, an agent with marginal cost equal to or above θ_{A_k+2} strictly prefers to truthfully report its cost rather than understate it to θ_{A_k} or below. Finally, $\hat{\sigma}_{j(A_k+1)} = 0$ for all $j \in \{1, ..., A_{k-1}\}$ if $k \geq 2$, again by monotonicity $\hat{q}_{A_{k-1}} > \hat{q}_{A_k}$.

Construct a perturbed mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \emptyset)$ by setting $t_j = \hat{t}_j - \epsilon$, $\epsilon > 0$, for all $j \in \{1, ..., A_k\}$ and setting $\sigma_{A_k+1} = \sum_{j \in A_k} \hat{\sigma}_{j(A_k+1)} + \hat{\sigma}_{A_k+1}$. Everything else is held equal to the original mechanism. This perturbed mechanism is incentive feasible for all ϵ sufficiently small. The difference in expected principal surplus between the two mechanisms is:

$$W(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \emptyset) - W(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \emptyset) = \sum_{i=1}^{A_k} \nu_i(1 - \alpha)\epsilon + \sum_{j \in A_k} \nu_{A_k+1}\hat{\sigma}_{j(A_k+1)}[W_{A_k+1}(\hat{x}_{A_{k+1}}) - W_{A_k+1}(\hat{x}_{A_k})],$$

which is strictly positive. The inequality follows from $\hat{\sigma}_{A_k+1} = 0$ if $W_{A_k+1}(\hat{x}_{A_k}) > W_{A_k+1}(\hat{x}_{A_{k+1}})$, which violates the incentive feasibility condition $\hat{\sigma}_{A_k+1} > 0$. Sine the upward IC condition in (33) is slack, then the local downward IC constraint in (33) necessarily is binding. To complete the proof, we need to establish $U_A(\hat{x}_A) = 0$. If $A \leq I - 1$, then $\hat{x}_{A+1} = x_0$. The binding downward IC condition then implies $U_A(\hat{x}_A) = U_A(\hat{x}_{A+1}) = U_A(x_0) = 0$. Assume next that $A = I$. If $U_I(\hat{x}_I) > 0$, then the principal could reduce the transfer for all cost types $j \in \mathcal{I}$ by $\epsilon > 0$ without violating incentive feasibility. Hence, $U_A(\hat{x}_A) = 0$ also in this final case. ■

Proof of Lemma 1

Claim 6 implies $U_i(\hat{x}_i) = U_i(\hat{x}_{i+1}) = U_{i+1}(\hat{x}_{i+1}) + (\theta_{i+1} - \theta_i)\hat{q}_{i+1}$ for all $i \in \{1, ..., A - 1\}$ and $U_A(\hat{x}_A) = 0$ in any incentive efficient mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\mathcal{A}, \emptyset)$ with complete commitment. We can then derive the property (10) of agency rent. Substituting these expressions into (7) produces

(11). This objective function is strictly concave in output. Hence, a mechanism $(\hat{\mathbf{x}}, \mathbf{I}|\mathcal{A}, \emptyset)$ with transfer payments that yield (10) and where output is characterized by (13) for all $\hat{K}$ cost groups, is incentive efficient if this mechanism is also incentive feasible. The mechanism satisfies (6) by $\hat{K} \leq K$. We then need to verify individual rationality (1) and incentive compatibility (2). To do this, we start by demonstrating monotonicity of output. Using a summation by parts,

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

we can write $\tilde{W}_{\mathcal{A}_k}(q)$ more compactly as

$$\tilde{W}_{\mathcal{A}_k}(q) = S(q) - \sum_{i\in\mathcal{A}_k}\frac{\nu_i}{\nu_{\mathcal{A}_k}}[\theta_i + \frac{G_{i-1}}{\nu_i}(1-\alpha)(\theta_i - \theta_{i-1})]q.$$

Comparing $\hat{q}_{\mathcal{A}_k}$ derived from this expression with (14) delivers:

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

Strict concavity of $S(q)$ implies $q^{sb}_{A_{k-1}+1} \geq \hat{q}_{\mathcal{A}_k}$ with strict inequality if $|\mathcal{A}_k| \geq 2$. Next,

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

implies $\hat{q}_{\mathcal{A}_k} \geq q^{sb}_{A_k}$ with strict inequality if $|\mathcal{A}_k| \geq 2$. We then get $\hat{q}_{\mathcal{A}_k} > \hat{q}_{\mathcal{A}_{k+1}}$ by $q^{sb}_{A_k} > q^{sb}_{A_k+1}$. These properties establish 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 $U_I(\hat{x}_I) = 0$, 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 (1) and (2) by Claim 1. This completes the proof of Item 1 of the Lemma.

As for Item 2 and Item 3, we first verify that $W^{sb}_{A_k}(\hat{q}_{\mathcal{A}_k}) \geq W^{sb}_{A_k}(\hat{q}_{\mathcal{A}_{k+1}})$ for all $k \in \{1, ...\hat{K}-1\}$, $\hat{K} \geq 2$ in any incentive efficient mechanism. This holds trivially if $|\mathcal{A}_k| = 1$ because then $W^{sb}_{A_k}(\hat{q}_{\mathcal{A}_k}) = w^{sb}_{A_k} \geq W^{sb}_{A_k}(\hat{q}_{\mathcal{A}_{k+1}})$. Assume that $|\mathcal{A}_k| \geq 2$. The modified mechanism $(\mathbf{x}, \mathbf{I}|\hat{\mathbf{A}}, \emptyset)$ described in Section 3 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 (1) and (2) by Claim 1. Moreover, $|\mathbf{x}_{\hat{\mathcal{A}}}| = |\hat{\mathbf{x}}_{\hat{\mathcal{A}}}| = \hat{K} \leq K$. These results verify incentive feasibility of $(\mathbf{x}, \mathbf{I}|\hat{\mathbf{A}}, \emptyset)$. Incentive efficiency of $(\hat{\mathbf{x}}, \mathbf{I}|\hat{\mathbf{A}}, \emptyset)$ therefore implies $W^{sb}_{\hat{A}_k}(\hat{q}_{\hat{\mathcal{A}}_k}) \geq W^{sb}_{\hat{A}_k}(\hat{q}_{\hat{\mathcal{A}}_{k+1}})$ by (16). One can use the same recipe to establish also the other properties of Item 2 and Item 3.

Proof of Lemma 2

We prove the Lemma through a sequence of 7 claims. Assume throughout that $K < I$ so that the second-best mechanism $(\mathbf{x}^{sb}, \mathbf{I}|\mathcal{I}, \emptyset)$ is infeasible. Let $z_j \in \{j, ..., I\}$ be the maximal cost type that reports θ_j with positive probability in the incentive feasible 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 7 *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 informational 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 (3), 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 (4) and maximizing over q_j leads to

$$q_j = S'^{-1}\left(\sum_{i=1}^{I} \mu_{ji}(\theta_i + (1-\alpha)(\theta_{z_j} - \theta_i))\right) \geq S'^{-1}(\theta_I) = q_I^{fb} > 0,$$

where $q_I^{fb} > 0$ by assumption, and $q_j \geq q_I^{fb}$ by $S'' < 0$ and

$$\theta_I - \sum_{i=1}^{I} \mu_{ji}(\theta_i + (1-\alpha)(\theta_{z_j} - \theta_i)) = \sum_{i=1}^{I} \mu_{ji}(\alpha(\theta_I - \theta_i) + (1-\alpha)(\theta_I - \theta_{z_j})) \geq 0.$$

■

Let $\underline{B} \in \mathcal{I}$ be the minimal cost type and $B \in \mathcal{I}$ the maximal cost type contained in $\mathcal{B}$ in a mechanism with incomplete commitment, i.e. $\underline{B} \in \mathcal{B}$, $B \in \mathcal{B}$, $\underline{B} \leq B$ and $\mathcal{B} \subseteq \{\underline{B}, ..., B\}$. In particular, the escape clause $\mathcal{B}$ need not be convex.

Claim 8 *A mechanism* $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ *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 (3), it follows that $U_{z_j}(x_{z_j}) = U_{z_j}(x_j) = t_j - \theta_{z_j} q_j = 0$. $U_{z_j}(x_{I_B}) > U_{z_j}(x_{z_j})$ then follows, which is a violation of (2). 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 (2),

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

Invoking incentive compatibility (2) again, plus individual rationality (1) and Claim 7 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 9 *A mechanism* $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}, \mathcal{B})$ *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 $\underline{B} \leq B - 1$. Obviously, $|\mathbf{x}_{\mathcal{B}}| = 1$ if $\underline{B} = B$. ■

Claim 9 establishes Item 1 of Lemma 2. Consider Item 2. $|\mathbf{x}_B| = 1$ if $z \geq B + 1$ by Claim 8. Hence, $|\mathbf{x}_B| = 2$ implies $z = B$. Claims 7, 8 and $z = B$ then imply $x_j = (q_j, \theta_B q_j)$ for all $j \in \mathcal{B}$. Invoking Claim 9 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 8. $U_i(x_i) > U_i(x_B)$ for all $i \neq B$ implies $\sigma_{Bi} = 0$ for all $i \neq B$ by (3). Hence, $\mu_{BB} = 1$ by (5) 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 2. To prove items 3-5, we now characterize additional properties of incentive efficient and incentive optimal mechanisms with incomplete commitment. The next claim states that local incentive compatibility constraints are binding even in mechanisms with incomplete commitment.

Claim 10 *Consider a mechanism* $(\hat{\mathbf{x}}, \hat{\mathbf{\Sigma}}|\mathcal{A}, \mathcal{B})$ *that entails ex-ante contracting* ($\mathcal{A} \neq \emptyset$) *and incomplete commitment* ($\mathcal{B} \neq \emptyset$). *Let the mechanism have the following properties:* $\underline{B} \geq 2$,

$\hat{q}_{A_k} > \hat{q}_{\underline{B}}$ *for some cost group* $\mathcal{A}_k$*, and* $\hat{x}_j = \hat{x}_{\underline{B}}$ $\forall j \in \{A_k + 1, ..., \underline{B} - 1\}$ *if* $A_k \leq \underline{B} - 2$*. This mechanism is incentive efficient only if*

$$[U_{A_k}(\hat{x}_{A_k}) - U_{A_k}(\hat{x}_{\underline{B}})][U_{A_k+1}(\hat{x}_{\underline{B}}) - U_{A_k+1}(\hat{x}_{A_k})] = 0, \tag{35}$$

the local incentive compatibility constraint A_k *is downward binding for* $A_k \leq \underline{B} - 2$*,*

$$U_{A_k}(\hat{x}_{A_k}) = U_{A_k}(\hat{x}_{\underline{B}}), \tag{36}$$

and

$$U_{A_l}(\hat{x}_{A_l}) = U_{A_l}(\hat{x}_{A_{l+1}}) \; \forall l \in \{1, ..., k-1\}, k \geq 2. \tag{37}$$

Proof. The proof of identity (35) is analogous to the proof of Claim 5 and the proofs of identities (36) and (37) are analogous to the proof of Claim 6. ■

We finally prove three claims of incentive optimal mechanisms.

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* $|\mathbf{x}^*_{\mathcal{A}^*}| = K$ *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 $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}, \mathcal{B})$ 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 $\mathcal{B} \subset \mathcal{B}^* \cup \emptyset$, $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ is not minimal in the sense of (9), and therefore 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 $\mathcal{B} = \emptyset$. By way of an identical arguments as above, the proposed mechanism is not minimal in the sense of (9), and therefore cannot be incentive optimal. ■

An immediate implication of Claim 11 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 8, which violates Claim 11.

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 12 *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 8. In particular, $q_{B^*+1}^* = q_{\underline{B}^*}^*$, which violates the necessary condition of incentive optimality established in Claim 11. Invoking Claim 8 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 13 *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 8. Let $B^* \leq I-1$. The proof proceeds as follows: We first show that $x_{B^*+1}^* = x_{B^*+1}^{fb}$ if $q_{B^*+1}^* > 0$. We then show that the principal in that case can obtain strictly higher expected surplus than in the proposed mechanism by modifying the escape clause. Hence, incentive optimality implies $q_{B^*+1}^* = 0$. We already showed in the proof of Claim 12 that $z^* \in \{B^*, B^*+1\}$. The final part of the proof establishes that $z^* \neq B^*+1$ if $q_{B^*+1}^* = 0$. This leaves $z^* = B^*$ as the only remaining possibility for $B^* \leq I-1$.

It cannot be the case that $q_{B^*+1}^* = q_{B^*}^*$, because this would violate Claim 11. If $q_{B^*+1}^* \in (0, q_{B^*}^*)$, then $z^* = B^*+1$ by Claim 12. Hence, $x_j^* = x_{\underline{B}^*}^*$ for all $j \in \mathcal{B}^*$ by Claim 8. Moreover, $\mathcal{A}_K^* = \{B^*+1\}$ identifies the maximal cost group in $\boldsymbol{\mathcal{A}}^*$ because $x_j^* = x_0$ for all $j \in \{B^*+2, ..., I\}$ if $B^* \leq I-2$; see Claim 8. 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 informational rent by distorting $q_{B^*+1}^*$, it follows that $x_{B^*+1}^* = x_{B^*+1}^{fb}$. Finally, $\sigma_{j(B^*+1)}^* = 0$ for all $j \in \{1, ..., \underline{B}^*-1\}$ 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:

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

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 $x_{B^*+1} = x_{B^*+1}^{fb} = x_{B^*+1}^*$ is sequentially rational following the cost report θ_{B^*+1} in the modified mechanism: $U_i(x_i^*) > U_i(x_{B^*+1}^*)$ for all $i \neq B^*+1$ implies $\sigma_{(B^*+1)i} = 0$ for all $i \neq B^*+1$, which in turn implies that the principal attaches posterior probability equal to 1 to the event that the agent has cost θ_{B^*+1} after observing that particular cost report. The expected surplus to the principal of the modified mechanism equals

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

The difference

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

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

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

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

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 $q^*_{\underline{B}^*} > q^{fb}_{B^*+1}$.

The expected surplus in the modified mechanism is

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

The difference in expected surplus between the two mechanisms is

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

which contradicts the assumed incentive optimality of $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$. We conclude that $q^*_{B^*+1} = 0$ implies $z^* \neq B^* + 1$. ■

We can now draw conclusions about incentive optimal mechanisms with incomplete commitment. Claim 11 proves Item 3 of Lemma 2. By way of $z^* = B^*$ and Item 2 of Claim 8, $\mathcal{C}^* = \{B^* + 1, ..., I\}$ 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 2. Item 4(a) follows from Claim 11 and monotonicity. Item 4(b) follows from Claim 7, Claim 8 and $z^* = B$. Obviously, $U_{B^*}(x^*_{\underline{B}^*}) = U_{B^*}(x^*_{B^*}) = 0$.

Proof of Lemma 3

We first demonstrate some general properties of $\mathbf{\Sigma}^*$ in incentive optimal mechanisms $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ that feature incomplete commitment ($\mathcal{B}^* \neq \emptyset$). This is done in the following claims.

Claim 14 *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:

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

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 $j \in \{\underline{B}^*, ..., I\}$.

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 $j \in \{B^*, ...I\}$.

Item 3: If $h = B^*$, then the rightmost expression in (38) is strictly positive for all $(i, j) \in \{1, ..., B^* - 1\} \times \mathcal{C}^*$ 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 $(i, j) \in \mathcal{C}^* \times \mathcal{A}^*$ by $q^*_j \geq q^*_{A^*} > q^*_{\underline{B}^*} \geq q^*_{B^*}$ for all $j \in \mathcal{A}^*$. $U_i(x^*_i) = 0 > -(\theta_i - \theta_{B^*})q^*_j = U_i(x^*_j)$ for all $(i, j) \in \mathcal{C}^* \times \mathcal{B}^*$ completes the proof. ■

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

Proof. We consider two cases separately. In case one, $q^*_{\underline{B}^*} > q^*_{B^*}$. By Claim 14, $\sigma^*_{I_{B^*}i} = 0$ 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 $x^*_{B^*} = x^{fb}_{B^*}$. This holds for any $\sigma^*_{B^*} > 0$. The expected surplus of the principal equals

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

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

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

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

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 informational rent by excluding one or more of the least efficient cost types. The maximal surplus the principal can achieve by offering a deviation contract $x^{*d}_{jh} = (q^{*d}_{jh}, \theta_h q^{*d}_{jh})$ 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

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

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

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

characterizes the optimal output given the reporting strategy $\boldsymbol{\sigma}_{jh} = (\sigma_{jA^*}, ..., \sigma_{jh})$. The contract x^{*d}_{jh} 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

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

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

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

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 $q_{\underline{B}^*}^* = q_{B^*}^*$. ■

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

Proof. Item 1 of Claim 14 implies $\sum_{j=1}^{A^*-1} \sigma_{jB^*}^* = 0$ if $A^* \geq 2$. Claim 15 implies $\sum_{j=B^*+1}^{I} \sigma_{jB^*}^* = 0$ if $B^* \leq I - 1$. Hence, $\sigma_{A^*B^*}^* + \sigma_{B^*}^* = 1$ if $\underline{B}^* = B^*$. $\sigma_{A^*B^*}^* > 0$ only if $U_{B^*}(x_{B^*}^*) = U_{B^*}(x_{A^*}^*)$. In that case, $U_{A^*}(x_{A^*}^*) > U_{A^*}(x_{B^*}^*)$ 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 principal's expected surplus then equals

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

in an incentive optimal mechanism where $\underline{B}^* = B^*$ and $\sigma_{B^*}^* < 1$. Consider a modified mechanism $(\mathbf{x}, \mathbf{\Sigma} | \mathcal{A}^*, \mathcal{B}^*)$ 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

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

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

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

Claim 17 *For any incentive optimal mechanism* $(\mathbf{x}^*, \mathbf{\Sigma}^{**}|\mathcal{A}^*, \mathcal{B}^*)$ *that features incomplete commitment* ($\mathcal{B}^* \neq \emptyset$)*, there exists an incentive optimal mechanism* $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ *where the reporting strategy* $\mathbf{\Sigma}^*$ *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 $\mathbf{\Sigma}^*$ 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 $\sigma_i^* = 1 \ \forall i \in C^*$ if $C^* \neq \emptyset$. The modification from $\mathbf{\Sigma}^{**}$ to $\mathbf{\Sigma}^*$ does not affect posterior beliefs for any reported $j \in \mathcal{B}^*$ in Stage 4 of the game. Hence, $(\mathbf{x}^*, \mathbf{\Sigma}^*|A^*, B^*)$ 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:

$$\begin{aligned} &W(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*) - W(\mathbf{x}^*, \mathbf{\Sigma}^{**}|\mathcal{A}^*, \mathcal{B}^*) \\ &\qquad = \textstyle\sum_{i=1}^{\underline{B}^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**})W_i(x_j^*) \\ &\qquad = \textstyle\sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**})W_i(x_j^*) \\ &\qquad = \textstyle\sum_{i=1}^{A^*} \sum_{j=1}^{A^*} \nu_i(\sigma_{ji}^* - \sigma_{ji}^{**})[W_i(x_j^*) - W_i(x_i^*) + W_i(x_i^*)] \\ &\qquad = \textstyle\sum_{i=1}^{A^*} \nu_i[\sum_{j=1}^{A^*} \sigma_{ji}^* W_i(x_j^*) - \sum_{j=1}^{A^*} \sigma_{ji}^{**} W_i(x_i^*)] \\ &\qquad = \textstyle\sum_{i=1}^{A^*} \nu_i[\sigma_i^* - \sum_{j=1}^{A^*} \sigma_{ji}^{**}]W_i(x_i^*) = 0. \end{aligned} \tag{43}$$

The third row follows from

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

where

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

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 obviously holds for $\sigma_{ji}^{**} = 0$, but also 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 the condition that $\sigma_i^{**} > 0$ in an incentive feasible mechanism. 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 14 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 14 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 15 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, $1 = \sum_{j=1}^{I} \sigma_{jA^*}^{**} = \sum_{j=1}^{A^*} \sigma_{jA^*}^{**} = \sigma_{A^*}^*$.

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

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

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

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

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

Claim 17 still differs from Lemma 3 in a number of aspects. One difference is that Lemma 3 is specific about the randomization strategies an agent with marginal cost θ_i, $i \in \{A^*, ..., B^*\}$, uses for cost reports θ_j, $j \in \{\underline{B}^*, ..., B^* - 1\}$ if $|\mathbf{X}_{\mathcal{B}^*}^*| = 2$ and $j \in \mathcal{B}^*$ if $|\mathbf{X}_{\mathcal{B}^*}^*| = 1$, that cause the principal to implement the ex post contract $x_{\underline{B}^*}^*$. We next establish incentive optimality of uniform randomization strategies.

Claim 18 *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* $(\mathbf{x}^*, \boldsymbol{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ *in which the reporting strategy* $\boldsymbol{\Sigma}^*$ *has the following properties:*

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

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

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

Proof. Using the results in Claim 17, we can write the principals' expected surplus as

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

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, $j \in \underline{\mathcal{B}}^* = \{\underline{B}^*, ..., B^*-1\}$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$ and in $\underline{\mathcal{B}}^* = \mathcal{B}^*$ if $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$. 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 $\bar{\Omega}_h(\mathbf{\Sigma}^{**}_h) \geq \Omega_h$.

Consider the problem of minimizing $\bar{\Omega}_h(\mathbf{\Sigma}_h)$ over $\mathbf{\Sigma}_h$ subject to $0 \leq \sigma_{ji} \leq 1$ for all $\sigma_{ji} \in \mathbf{\Sigma}_h$, $\sum_{j=\underline{B}^*}^{B^*-1} \sigma_{ji} = 1 - \sigma^{**}_{A^*i}$ 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:

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

with the cross-partial derivative of

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

If we define

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

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:

$$\mathbf{H}_{jh} = \frac{-\boldsymbol{y}_{jh} \boldsymbol{y}^T_{jh}}{\sum_{i=A^*}^{h} \nu_i \sigma_{ji} S''(q^d_{jh})}.$$

By implication:

$$\boldsymbol{\sigma}^T_{jh} \mathbf{H}_{jh} \boldsymbol{\sigma}_{jh} = \frac{-\boldsymbol{\sigma}^T_{jh} \boldsymbol{y}_{jh} \boldsymbol{y}^T_{jh} \boldsymbol{\sigma}_{jh}}{\sum_{i=A^*}^{h} \nu_i \sigma_{ji} S''(q^d_{jh})} = \frac{-(\boldsymbol{\sigma}^T_{jh} \boldsymbol{y}_{jh})^2}{\sum_{i=A^*}^{h} \nu_i \sigma_{ji} S''(q^d_{jh})} \geq 0.$$

Positive definiteness of $\mathbf{H}_{jh}$ implies that $\Omega_{jh}(\boldsymbol{\sigma}_{jh})$ is a convex function. Consequently, $\bar{\Omega}_h(\mathbf{\Sigma}_h)$ 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

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

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 $\bar{\Omega}_h(\mathbf{\Sigma}_h)$. 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, $\mathbf{\Sigma}^*_h$, $\underline{\xi}_{ji}=\xi_{ji}=0$ and $\lambda_{ih}=\nu_i W_i(x^d_h)$ jointly solve (47)-(49). Hence, $\Omega^*_{B^*}\geq \bar{\Omega}_h(\mathbf{\Sigma}^{**}_h)\geq\bar{\Omega}_h(\mathbf{\Sigma}^*_h)=\Omega_h$ for all $h\in\{A^*,...,B^*-1\}$. ■

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

Claim 19 *Consider an incentive optimal mechanism* $(\mathbf{x}^{**},\mathbf{\Sigma}^{**}|\mathcal{A}^{**},\mathcal{B}^{**})$ *that features incomplete commitment* ($\mathcal{B}^{**}\neq\emptyset$)*, and where* $\sigma^{**}_{ji}>0$ *for some* $(i,j)\in\mathcal{B}^{**}\times\mathcal{A}^{**}$*. Assume that this mechanism is characterized either by (i)* $\underline{B}^{**}=B^{**}-1$ *and* $|\mathbf{x}^{**}_{\mathcal{B}^{**}}|=1$*, or (ii)* $\underline{B}^{**}\leq B^{**}-2$*. Then there exists an incentive optimal mechanism* $(\mathbf{x}^*,\mathbf{\Sigma}^*|\mathcal{A}^*,\mathcal{B}^*)$ *that also features incomplete commitment* ($\mathcal{B}^*\neq\emptyset$)*, where* $\sigma^*_{ji}=0$ *for all* $(i,j)\in\mathcal{B}^*\times\mathcal{A}^*$*.*

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

$$\mu^{**}_{j\underline{B}^{**}}=\mu^{**}_{\underline{B}^{**}}=\frac{\nu_{\underline{B}^{**}}(1-\sigma^{**}_{A^{**}\underline{B}^{**}})}{\sum_{i\in\mathcal{B}^{**}}\nu_i-\nu_{\underline{B}^{**}}\sigma^{**}_{A^{**}\underline{B}^{**}}-\nu_{B^{**}}\sigma^{**}_{B^{**}}}$$
$$\mu^{**}_{ji}=\mu^{**}_{i}=\frac{\nu_i}{\sum_{i\in\mathcal{B}^{**}}\nu_i-\nu_{\underline{B}^{**}}\sigma^{**}_{A^{**}\underline{B}^{**}}-\nu_{B^{**}}\sigma^{**}_{B^{**}}}\forall i\in\{\underline{B}^{**}+1,...,B^{**}-1\}$$
$$\mu^{**}_{jB^{**}}=\mu^{**}_{B^{**}}=\frac{\nu_{B^{**}}(1-\sigma^{**}_{B^{**}})}{\sum_{i\in\mathcal{B}^{**}}\nu_i-\nu_{\underline{B}^{**}}\sigma^{**}_{A^{**}\underline{B}^{**}}-\nu_{B^{**}}\sigma^{**}_{B^{**}}}$$

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

$$\mu^{**}_{j\underline{B}^{**}}=\mu^{**}_{\underline{B}^{**}}=\frac{\nu_{\underline{B}^{**}}(1-\sigma^{**}_{A^{**}\underline{B}^{**}})}{\sum_{i\in\mathcal{B}^{**}}\nu_i-\nu_{\underline{B}^{**}}\sigma^{**}_{A^{**}\underline{B}^{**}}}$$
$$\mu^{**}_{ji}=\mu^{**}_{i}=\frac{\nu_i}{\sum_{i\in\mathcal{B}^{**}}\nu_i-\nu_{\underline{B}^{**}}\sigma^{**}_{A^{**}\underline{B}^{**}}}\forall i\in\{\underline{B}^{**}+1,...,B^{**}\}$$

for all $j \in \mathcal{B}^{**}$ if $|\mathbf{x}^{**}_{\mathcal{B}^{**}}| = 1$. 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 $|\mathbf{x}^{**}_{\mathcal{B}^{**}}| = 2$, then $\sigma^*_{j\underline{B}^{**}} = \frac{1-\sigma^*_{\underline{B}^{**}}}{B^{**}-\underline{B}^{**}-2}$, $\sigma^*_{ji} = \frac{1}{B^{**}-\underline{B}^{**}-2}$, $i \in \{\underline{B}^{**}+1, ..., B^{**}-1\}$, and $\sigma^*_{jB^{**}} = \frac{1-\sigma^{**}_{B^{**}}}{B^{**}-\underline{B}^{**}-2}$ for all $j \in \{\underline{B}^{**}+1, ..., B^{**}-1\}$, whereas $\sigma^*_{B^{**}} = \sigma^{**}_{B^{**}}$. If $|\mathbf{x}^{**}_{\mathcal{B}^{**}}| = 1$, then $\sigma^*_{j\underline{B}^{**}} = \frac{1-\sigma^*_{\underline{B}^{**}}}{B^{**}-\underline{B}^{**}-1}$, $\sigma^*_{ji} = \frac{1}{B^{**}-\underline{B}^{**}-1}$, $i \in \{\underline{B}^{**}+1, ..., B^{**}\}$, 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 modified mechanism is incentive feasible if $\mathbf{x}^{**}_{\mathcal{B}^{**}}$ is sequentially rational. It is easy to verify that the uniform distribution of cost reports $\mathbf{\Sigma}^*$ yields posterior beliefs $\mu^*_{ji} = \mu^{**}_i$ for all $(i,j) \in \mathcal{I} \times \{\underline{B}^{**}+1, ..., B^{**}\}$. $\mathbf{x}^{**}_{\mathcal{B}^{**}}$ is sequentially rational under the reporting strategy $\mathbf{\Sigma}^*$ because it is sequentially rational under $\mathbf{\Sigma}^{**}$. ■

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

Proof of Proposition 1

Let $(\mathbf{x}^*, \mathbf{\Sigma}^*|\mathcal{A}^*, \mathcal{B}^*)$ be an incentive optimal mechanism with incomplete commitment ($\mathcal{B}^* \neq \emptyset$). We elaborated on the $|\mathbf{x}^*_{\mathcal{B}^*}| = 1$ case in the main text, so we focus here on $|\mathbf{x}^*_{\mathcal{B}^*}| = 2$. There are two subcases.

Subcase 1: $\sigma^*_{A^*\underline{B}^*} = 0$. We start the proof by demonstrating incentive feasibility of a modified mechanism $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ in which the menu $\mathbf{x}$ of contracts has the following properties: $x_j = (q^*_j, t^*_j - (\theta_{B^*} - \theta_{A^*})(q^*_{\underline{B}^*} - q_{\underline{B}^*}))$ for all $j \in \mathcal{A}^*$, $x_j = x_{\underline{B}^*} = (q_{\underline{B}^*}, \theta_{B^*} q_{\underline{B}^*})$ for all $j \in \mathcal{B}^*$, and

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

As for the reporting strategies, $\sigma_i = \sigma^*_i = 1$ for all $i \in \{1, ...A^*-1\}$ if $A^* \geq 2$, and $\sigma_i = \sigma^*_i = 1$ for all $i \in \mathcal{C}^*$ if $\mathcal{C}^* \neq \emptyset$. Moreover, $\sigma_{jA^*} = \frac{1-\sigma^*_{A^*}}{|\mathcal{B}^*|}$ for all $j \in \mathcal{B}^*$ and $\sigma_{ji} = \frac{1}{|\mathcal{B}^*|}$ for all $(i,j) \in \mathcal{B}^* \times \mathcal{B}^*$.

By subtracting (19) from the above equation, we obtain

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

Hence, $q^*_{\underline{B}^*} > q_{\underline{B}^*}$ by strict concavity of $S(q)$.

Since $U_i(x_i) - U_i(x_j) = U_i(x^*_i) - U_i(x^*_j)$ 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 $q^*_{A^*} > q_{\underline{B}^*}$, one can apply Item 2 of Claim 2 and Item 1 of Claim 3 to verify individual rationality (1) and incentive compatibility (2) of $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$. The mechanism obviously satisfies the stochastic rationality constraint (3). We calculate the posterior beliefs (5) of the modified mechanism as:

$$\mu_{jA^*} = \frac{\nu_{A^*}(1-\sigma^*_{A^*})}{\nu_{A^*}(1-\sigma^*_{A^*})+\nu_{\mathcal{B}^*}}\ \forall j \in \mathcal{B}^*,\ \mu_{ji} = \frac{\nu_i}{\nu_{A^*}(1-\sigma^*_{A^*})+\nu_{\mathcal{B}^*}}\ \forall (i,j) \in \mathcal{B}^* \times \mathcal{B}^*.$$

Moreover $|\mathbf{x}_{\mathcal{A}^*}| = |\mathbf{x}^*_{\mathcal{A}^*}|$ implies (6). We then need to verify sequential rationality (4) 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

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

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

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

By this construction, $\tilde{q}(\sigma^*_{B^*}) = q^*_{\underline{B}^*}$, $\tilde{\Omega}(\sigma^*_{B^*}) = \Omega^*_{B^*}$, see the proof of Claim 18 for a definition, and $\tilde{q}(0) = q_{\underline{B}^*}$. 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 $x^d_h = (q^d_h, \theta_h q^d_h)$, $h \in \{A^*, ..., B^*-1\}$, equals $\frac{\Omega_h}{\nu_{A^*}(1-\sigma^*_{A^*})+\nu_{\mathcal{B}^*}}$, see the proof of Claim 18 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

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

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 15. 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 following reduced communication mechanism augmented by a vague escape clause (VEC) can be sustained as a PBE: The principal offers $\mathbf{x}_{\boldsymbol{\mathcal{A}}^*}$ in Stage 1. In Stage 2, an agent with marginal cost $\theta_i < \theta_{A^*}$ selects $x_{\mathcal{A}^*_k}$, $i \in \mathcal{A}^*_k$, an agent with marginal cost θ_{A^*} selects $x_{\mathcal{A}^*_K}$ with probability $\sigma^*_{A^*}$ and invokes

VEC with probability $1-\sigma^*_{A^*}$, an agent with marginal cost θ_i, $i \in \mathcal{B}^*$, triggers VEC, an agent with marginal cost $\theta_i > \theta_{B^*}$ 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 the same expected surplus to the principal as $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$, namely:

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

where $\nu_{A^*} = \sum_{i\in\mathcal{A}^*}\nu_i$. The net benefit of choosing the incentive optimal mechanism over $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$ can be written as

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

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

Subcase 2: $\sigma^*_{A^*\underline{B}^*} > 0$. This possibility occurs only if $\underline{B}^* = B^* - 1$ and $\sigma^*_{A^*} = 1$; see Lemma 3. In this case, $\sigma^*_{A^*\underline{B}^*} = 1 - \sigma^*_{\underline{B}^*}$. 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 as follows: $\sigma_i = \sigma^*_i = 1$ for all $i \in \mathcal{A}^*$. If $\mathcal{C}^* \neq \emptyset$, then $\sigma_i = \sigma^*_i = 1$ also for all $i \in \mathcal{C}^*$. Moreover, $\sigma_{A^*\underline{B}^*} = 1 - \sigma^*_{\underline{B}^*}$, $\sigma_{\underline{B}^*} = \sigma_{B^*\underline{B}^*} = \frac{\sigma^*_{\underline{B}^*}}{2}$ and $\sigma_{\underline{B}^* B^*} = \sigma_{B^*} = \frac{1}{2}$. 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 (1)-(3) and (6). We calculate the posterior beliefs (5) of the modified mechanism as:

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

We finally verify sequential rationality (4) of $x_{\underline{B}^*}$. In the incentive optimal mechanism, $x^*_{\underline{B}^*}$ is sequentially rational if and only if

$$\Omega^*_{B^*} = \nu_{\underline{B}^*}\sigma^*_{\underline{B}^*}[S(q^*_{\underline{B}^*})-(\theta_{\underline{B}^*}+(1-\alpha)(\theta_{B^*}-\theta_{\underline{B}^*}))q^*_{\underline{B}^*}]+\nu_{B^*}(1-\sigma^*_{B^*})[S(q^*_{\underline{B}^*})-\theta_{B^*}q^*_{\underline{B}^*}] \geq \nu_{\underline{B}^*}\sigma^*_{\underline{B}^*}w^{fb}_{\underline{B}^*}.$$

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

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

One can then construct a similar function to $\tilde{\Omega}(\sigma)$ above to verify $\Omega_{B^*} > \Omega^*_{B^*}$, but we omit this step.

By the properties of $(\mathbf{x}, \mathbf{\Sigma}|\mathcal{A}^*, \mathcal{B}^*)$, the following reduced communication mechanism sustained by a vague escape clause (VEC) can be sustained as a PBE: The principal offers $\mathbf{x}_{\mathcal{A}^*}$ in Stage 1. In Stage 2, any agent with marginal cost $\theta_i \leq \theta_{A^*}$ selects $x_{\mathcal{A}^*_k}$, $i \in \mathcal{A}^*$, an agent with marginal cost $\theta_{\underline{B}^*}$ selects $x_{\mathcal{A}^*_K}$ with probability $1-\sigma^*_{\underline{B}^*}$ and invokes the escape clause with probability $\sigma^*_{\underline{B}^*}$, an agent with marginal cost θ_{B^*} triggers the escape clause with probability 1, an agent with marginal cost $\theta_i > \theta_{B^*}$ 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

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

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

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

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