Procurement with Unforeseen Contingencies[†]

Fabian Herweg[*]
University of Bayreuth, CESifo, and CEPR

Klaus M. Schmidt[**]
University of Munich, CESifo and CEPR

This version: June 4, 2018

Abstract: A buyer who wants to procure a complex good is often aware that there may be flaws in her initial design, but she does not know how they look like. Potential sellers who discover flaws have no incentive to reveal them early if the buyer uses a price-only auction. We derive the efficient mechanism that induces all sellers to report flaws early and that allocates the project to the seller with the lowest cost. We show that this can be implemented with a simple two-stage auction that does not require any prior knowledge of the set of possible flaws.

JEL classification numbers: D44; D82; D83; H57.

Keywords: Procurement; Renegotiation; Auctions; Design Flaws; Adaptation Costs; Behavioral Contract Theory.

[†]Financial support by Deutsche Forschungsgemeinschaft through CRC-TRR 190 is gratefully acknowledged. We would like to thank Dirk Bergemann, Jacques Crémer, Dominik Fischer, Drew Fudenberg, Daniel Garrett, Bob Gibbons, Oliver Hart, Bengt Holmström, Louis Kaplov, Paul Klemperer, Michael Kremer, Patrick Legros, Moritz Meyer-ter-Vehn, Benny Moldovanu, Takeshi Murooka, Christoph Schottmüller, Andrei Shleifer, Roland Strausz, and seminar participants at Ann Arbor, Harvard, Mannheim, MIT, Oxford, Stockholm, Toulouse, UCLA, Yale, Zurich, the NBER Organizations Workshop and the Theoretischer Ausschuss of the VfS for many helpful comments and suggestions. Part of this research was conducted while the first author visited University of Technology in Sydney and the second author visited the Economics Department at Harvard. We would like to thank our hosts for their great hospitality.

[*]Faculty of Law, Business and Economics, University of Bayreuth, Universitätsstr. 30, D-95440 Bayreuth, Germany, email: fabian.herweg@uni-bayreuth.de.

[**]Department of Economics, University of Munich, Ludwigstr. 28 (Rgb.), D-80539 München, Germany, email: klaus.schmidt@econ.lmu.de (corresponding author).

1 Introduction

The procurement of complex projects is often plagued by large cost overruns. One reason for unexpected additional costs are flaws in the initial design of the project. When these flaws are revealed after production has started, the design of the project needs to be changed and contracts have to be renegotiated which often leads to substantial adjustment costs. To minimize design flaws and the resulting adjustment costs the early collaboration with potential contractors is of crucial importance. However, if the project is procured by a standard price-only auction, potential contractors have no incentive to contribute their expertise. Each potential contractor who spotted a design flaw has a strong incentive to conceal this information, bid more aggressively in the auction in order to have a better chance of winning the contract, and then – after the award of the contract – renegotiate with the buyer in a bilateral monopoly position to fix the design flaw and thereby grab some share of the renegotiation surplus.

Bajari and Tadelis (2011, p. 132-133) report that *"it is widely believed [...] that when competitive tendering is used to award a fixed-price contract, the contractors strategically read the plans and specifications to determine where they will fail. [...] Competitive tendering may therefore lead to a problem of ex ante opportunism that is more problematic when projects are complex."*[1] The additional costs can be substantial. For example, during the Big Dig highway artery project in Boston \$ 1.1 billion unplanned construction costs can be traced back to more than 3200 cases where design flaws required contracts to be renegotiated.[2] However, the

[1]One of the authors (KS) participated in a workshop on public procurement rules organized by the Association of the Bavarian Construction Industry (BBIV). In private communication he learned from several representatives of construction companies that if there is a public procurement project the *only* thing they are interested in are the flaws in the design, because this is where they can make a profit. The parts of the project that are well specified don't earn any money because of the fierce competition in the auction. Leading representatives of a consultancy that specializes in the design of procurement auctions for large European companies confirmed that this problem is not restricted to the construction industry but widespread in many industries.

[2]See Boston Globe, 2/9/2003, http://archive.boston.com/news/specials/bechtel/part_1/. The article goes on to discuss the following case study as a typical example: "On July 15, 1997, state officials gathered to award a contract to build tunnels from Haymarket Square to North Station ... Bechtel estimated the job would cost about \$ 260 million to complete ... As it turned out, the low bid came in at \$ 218 million. Artery officials rejoiced. But their joy was short-lived. Today, the contract ... has grown \$ 128 million beyond the bid submitted that July day, an increase of nearly 60 percent." The cost increase was due to renegotiation of the initial contract necessary to fix several design flaws.

To be sure, there are many other contributors to the massive cost increases of more than \$ 10 billion in the Big Dig project, including unforeseen geological problems, design changes due to changing regulation and changing demands of the client, bureaucratic incompetence, corruption, etc. It is often difficult to disentangle empirical evidence is mostly anecdotal because a contractor will not admit – for legal reasons – that he knew about a design flaw early and concealed this information deliberately in order to renegotiate.[3]

In this paper we propose an informationally robust mechanism that allocates the contract to the seller with the lowest cost and that induces all potential sellers to reveal any information that they may have about possible design flaws early, i.e. before the contract is assigned. As a benchmark, we solve for the direct mechanism that implements the efficient allocation at the lowest possible cost to the buyer in ex post equilibrium. This optimal direct mechanism is ex post incentive compatible, so it does not depend on the priors (and higher order beliefs) of the involved parties about the likelihood of possible flaws and of the probabilities that these flaws have been spotted by each of the sellers. However, a crucial drawback of this direct mechanism is that it requires the parameters of the model to be common knowledge. In particular, the buyer has to know the set of potential design flaws and their payoff consequences. In the procurement context this is a very strong assumption. Buyers are often aware that the initial design may be flawed, but they have no idea how possible flaws look like and what their payoff implications are. After all, if they had this information, it would be easy for them to look for and detect the design flaws themselves.

To deal with this problem we propose an outcome equivalent indirect mechanism that does not require any knowledge about the set of possible flaws *ex ante*. The indirect mechanism uses an independent arbitrator who verifies flaws *ex post*. Sellers are asked to point out any design flaw they discovered to the arbitrator. The arbitrator verifies the flaws, estimates their payoff consequences, and assigns payments to the sellers according to a simple and intuitive rule: If a seller is the only one who reported a flaw, he will get a reward that is equal to the additional profit he would have made if he had kept the information to himself, won the contract, and renegotiated with the buyer. Otherwise, he gets nothing. Thereafter, the revealed flaws are

these effects.

[3]An instructive example is the planning and construction of several escalators at the still unfinished new airport in Berlin. The escalators were planned two feet too short. The construction company pointed out this flaw after winning the contract and proposed to renegotiate and fix the flaw at a stiff price increase. The Berlin airport company was outraged, refused to pay, and accused the construction company of withholding this information on purpose to inflate the price. In this case renegotiation failed, the escalator was built as planned, and a staircase with three steps had to be added to connect the escalator to the floor level. See Die Welt, 9/25/2015, https://www.welt.de/wirtschaft/article146839756/Das-ist-die-erschreckende-Fehlerliste-des-BER.html.

fixed and the improved design is allocated to the seller with the lowest cost in a simple Vickrey auction.

The indirect mechanism requires that design flaws can be verified ex post: Once a flaw has been pointed out, every industry expert (arbitrator) understands the flaw and knows how to fix it. She is also able to come up with an unbiased estimate of the payoff consequences if the flaw is fixed late rather than early. This assumption is weaker than the assumption that the buyer knows all potential flaws ex ante, and it is very plausible in the procurement context. Arbitrators are frequently used to assess damages if something goes wrong in a contractual relationship and to assign payments according to a rule (e.g. "expectation damages") based on their ex post assessments of what would have happened if the parties had behaved differently.

The contribution of our paper is twofold. First, we suggest a new, simple mechanism that can be used to solve the real world problem of "ex ante opportunism" in procurement contracts at the lowest possible cost for the buyer. Second, our paper contributes to the literature on "robust implementation". This literature (Bergemann and Morris, 2005) requires mechanisms to be "belief-free", i.e. not to depend on the unobservable beliefs and higher-order beliefs of the participants. Our model goes one step further and requires in addition that the mechanism should not require the mechanism designer to know all possible states of the world (i.e. the set of possible flaws and their payoff consequences). We show that this can be achieved by combining ex post incentive compatibility and ex post verifiability.

After discussing the relation of our paper to the literature in the next section, we set up the model in Section 3. In Section 4 we show that a standard price-only auction offers no incentives to sellers to reveal private information about possible design flaws early. This gives rise to three inefficiencies. First, fixing the flaw via renegotiation is more costly than fixing it early (Inefficient Renegotiation). Two further inefficiencies arise if the flaw is spotted only by the seller with higher production cost. Either this seller wins the auction – by bidding aggressively in anticipation of the renegotiation profit. In this case production is carried out at a too high cost (Inefficient Production). Or he does not win the auction – because the cost difference to the other seller is larger than the expected renegotiation profit. In this case the flaw will not be pointed out and the buyer suffers from the flawed design (Inefficient Design).

In Section 5 we assume that all parties know the underlying model, i.e. they know all potential design flaws and their payoff consequences. We solve for the least costly, ex post incentive compatible direct mechanism that induces all sellers to reveal any flaws that they observed to the buyer early and that allocates the project to the seller with the lowest cost.

In Section 6 we show that the allocation that is implemented by the optimal direct mechanism can also be implemented by a simple indirect mechanism that does not require the buyer (or any other party) to know the parameters of the underlying model. An important characteristic of this mechanism is that it separates the two problems of eliciting information about design flaws and assigning the contract to the seller with the lowest cost. We show that this is necessary to achieve efficiency.

So far we assumed that each seller observes a subset of the actual flaws with some exogenously given probability for free. In Section 7 we generalize this model and allow for search costs. Each seller has to actively search for possible design flaws which requires costly effort. The optimal mechanism of Section 5 does not offer efficient incentives to incur these costs. We derive the optimal mechanism that induces sellers not just to reveal their information about costs and flaws truthfully, but also to search efficiently for possible flaws. This mechanism has to pay a higher information rent to the seller, but it is also somewhat easier to specify because it does not depend on the bargaining power of the parties in the renegotiation game.

Section 8 concludes and discusses some possible directions of future research. All proofs are relegated to the Appendix.

2 Relation to the Literature

Our paper contributes to three strands of the literature. First, there is a large literature on optimal procurement auctions (McAfee and McMillan, 1986; Laffont and Tirole, 1993). The novel feature in our set-up is that sellers may have superior information about possible design flaws that the buyer would like to elicit. Our approach is closely related to the literature on scoring auctions (Asker and Cantillon, 2010; Che, 1993; Che, Iossa, and Rey, 2016), where sellers also make design proposals (submit bids on design and price). A scoring auction assigns the contract to the seller who comes up with the best proposal (the highest total score). In contrast, our mechanism combines the suggestions of several sellers to improve the design and assigns the contract for the improved design to the seller with lowest cost.

Second, our paper is related to the literature on "robust mechanism design". Bayesian mechanism design theory has often been criticized because the optimal mechanism crucially depends on the precise information that the agents and the mechanism designer have, in particular their prior beliefs and higher order beliefs that are not observable. Wilson (1987) pointed out that if the agents or the designer are mistaken in their beliefs, then the outcome of the supposedly optimal mechanism may be very different from the intended outcome. Bergemann and Morris (2005) require "robust implementation" that is independent of beliefs and higher order beliefs and depends only on payoff relevant types. They have shown that implementation is robust if the mechanism satisfies ex post incentive compatibility, i.e. if the strategy of each agent is optimal against the strategies of all other agents for every possible realization of types.[4] Our optimal mechanism satisfies ex post incentive compatibility and is therefore "informationally robust" in this sense. But, our mechanism is also "informationally robust" in a complementary and much stronger sense. It does not require any knowledge of the possible type spaces of the sellers! We are not aware of any other papers on robust mechanism design with this feature. This result requires that an arbitrator can verify reported flaws and evaluate their expected payoff consequences in order to "complete" the mechanism ex post. This is a novel assumption in the mechanism design literature that is plausible in many contexts and deserves further attention.

Finally, there is a small but growing literature on the inefficiencies of contract renegotiation. Several empirical studies emphasize that renegotiation is often costly and inefficient, including Crocker and Reynolds (1993), Chakravarty and MacLeod (2009), and Bajari, Houghton, and Tadelis (2014). Bajari et al. (2014, p. 1317) consider highway procurement contracts in California. They report that renegotiation costs are substantial and estimate that they "range from 55 cents to around two dollars for every dollar in change". A behavioral foundation based on loss aversion for inefficient renegotiation is developed by Herweg and Schmidt (2015). Other contributions like Bajari and Tadelis (2001) and Herweg and Schmidt (2017) start out from the assumption that renegotiation is costly and investigate the implications.

[4]See Bergemann and Morris (2012) for a survey of the literature on robust implementation. Bergemann and Morris distinguish between *partial* robust implementation and *full* robust implementation. We exclusively consider *partial* robust implementation in this paper and refer to it simply as robust implementation.

Bajari and Tadelis (2001) compare fixed-price to cost-plus contracts and show that standardized goods should be procured by fixed-price contracts that give strong cost-saving incentives to sellers, while complex goods should be procured by cost-plus contracts in order to avoid costly renegotiation. Herweg and Schmidt (2017) compare price-only auctions to bilateral negotiations. They show that negotiating with one selected seller may outperform an auction because the auction induces sellers to conceal private information about design improvements, which gives rise to inefficient renegotiation. These papers compare standard contracts and procurement procedures, while the current paper solves for the optimal procurement mechanism and proposes a new procedure.

3 The Model

A buyer (B, female) wants to procure a complex good from one of two sellers (male), denoted by $i \in \{1, 2\}$.[5] At date 0 the buyer comes up with a design proposal D_0 for the good. Seller i can produce design D_0 at cost $c^i \in [\underline{c}, \bar{c}]$, $0 \le \underline{c} \le \bar{c}$. These costs c^1 and c^2 are private information and drawn from some atomless cdf $H(c^1, c^2)$. If design D_0 is optimal, it generates utility v for the buyer. However, with some probability the design is plagued by one or multiple flaws which reduce the buyer's utility if design D_0 is implemented. In order to restore the buyer's utility to v, the flaws have to be fixed by adjusting the design.

We model the possibility of design flaws as follows. Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ denote the set of possible design flaws and let $\mathcal{P}(\mathcal{F})$ denote the power set of $\mathcal{F}$, i.e. the set of all possible subsets of $\mathcal{F}$ including $\emptyset$.[6] A typical element of $\mathcal{P}(\mathcal{F})$ is denoted by F. $F \in \mathcal{P}(\mathcal{F})$ is drawn from $\mathcal{P}(\mathcal{F})$ according to probability distribution $G(F)$. If $F = \emptyset$, there is no design flaw. If $F \neq \emptyset$, a non-empty subset of flaws has materialized.

Sellers are better able to detect design flaws than the buyer.[7] When the buyer proposes design D_0, each seller privately observes a subset of the realized flaws. Let $\mathcal{P}(F)$ denote the power set of F. Each seller observes a private signal $\hat{F}^i \in \mathcal{P}(F)$. If $\hat{F}^i = \emptyset$, seller i observes

[5] We restrict attention to the case of two sellers for notational simplicity only. It is straightforward to extend the analysis to the case of N sellers.

[6] If $\mid \mathcal{F} \mid$ is the cardinality of $\mathcal{F}$, then $\mid \mathcal{P}(\mathcal{F}) \mid = 2^{|\mathcal{F}|}$.

[7] After all, the sellers are the experts in producing the good. This is why the buyer turned to them in the first place and does not produce the good herself.

nothing. If $\hat{F}^i \neq \emptyset$, seller i observes some non-empty subset of the set of realized flaws F. The joint probability distribution over $(\hat{F}^1, \hat{F}^2)$ conditional on the set of realized flaws F is denoted by Q_F.

A seller can report only flaws that he observed, but he is free which of these flaws to report. Let $\tilde{F}^i$ denote the set of flaws reported by seller i, i.e., $\tilde{F}^i \in \mathcal{P}(\hat{F}^i)$. If $\tilde{F}^i = \emptyset$, seller i reports nothing. If $\tilde{F}^i \neq \emptyset$, seller i reports some (or all) of the flaws that he observed. His report can be partially verified:

Assumption 1 (Partial Verifiability)**.** *Each seller i can report any subset $\tilde{F}^i$ of the set of flaws $\hat{F}^i$ that he observed (including $\emptyset$). All flaws $f \in \tilde{F}^i$ can be verified. However, if $f_k \notin \tilde{F}^i$, it is impossible to verify whether $f_k \in \hat{F}^i$, i.e. whether seller i did or did not observe flaw f_k.*

Sellers can report flaws at date 1 ("early", before the contract is assigned) or at date 2 ("late", after the contract has been assigned and production has started). We assume that once a flaw has been pointed out, it can be fixed by any seller at the same cost. If a flaw f_k is fixed early (at date 1), the cost of fixing it is $\Delta c_k \geq 0$.[8] If it is fixed late (at date 2) the cost increases by $\Delta x_k \geq 0$. If f_k is not fixed, the buyer's utility is reduced to $v - \Delta v_k$.[9] We assume that $\Delta v_k > \Delta c_k + \Delta x_k$ for all $k \in \{1, \ldots, n\}$, so fixing a flaw is always efficient, but it is more efficient to fix it early rather than late. At date 2 the buyer has made already several other commitments (contracts with other suppliers, customers, etc.) that are based on the design of the initial contract. Furthermore, the parties may disagree on who is responsible for the design flaw and who should bear the cost of fixing it which may lead to haggling, aggrievement, and further delays. Thus, the surplus from fixing a flaw shrinks from $S_k = \Delta v_k - \Delta c_k$ if f_k is fixed at date 1 to $S_k^R = \Delta v_k - \Delta c_k - \Delta x_k$ if f_k is fixed via renegotiation at date 2.[10]

Let $D(\tilde{F})$ denote the design that fixes all flaws $f_k \in \tilde{F}$, where $\tilde{F} = \tilde{F}^1 \cup \tilde{F}^2$ is the set of flaws that have been reported to the buyer at date 1. If the set of actual flaws is F, $\tilde{F} \subseteq F$,

[8]The cost of fixing a flaw is small as compared to the cost of producing the full project. Thus, for simplicity we assume that all sellers have the same cost Δc_k to fix a flaw even though their cost c^i to produce the entire project D_0 may differ.

[9]The flaws may also be interpreted as "design improvements". If the design improvement is implemented early, it raises the buyer's valuation to $v + \Delta v_k$, if it is not implemented, the buyer's valuation remains at v. It is straightforward to reinterpret the model along these lines.

[10]The model assumes that flaws can be fixed independently of each other. The idea is that if two (or more) flaws interact with each other, then the discovery of one flaw will lead to the discovery of the other connected flaws as well. Thus, we can treat a set of interdependent flaws as a single flaw.

then the gross utility of the buyer derived from design $D(\tilde{F})$ is given by

$$V(D(\tilde{F})|F) = v - \sum_{\{k|f_k \in F\setminus\tilde{F}\}} \Delta v_k \tag{1}$$

while the cost of the seller to produce $D(\tilde{F})$ is

$$C^i(D(\tilde{F})) = c^i + \sum_{\{k|f_k \in \tilde{F}\}} \Delta c_k \ . \tag{2}$$

We assume that v is sufficiently large so that the buyer always wants to procure the good no matter how many flaws there are and when they are reported.[11]

Note that a seller may have an incentive to report a flaw late even though this raises the cost of fixing it. The reason is that he cannot simply "sell" his information to the buyer at date 1. Anybody can claim that there is a flaw. To prove his claim a seller has to point out what the flaw is, but once he does so, he gives his information away and the buyer no longer needs to pay for it (see Arrow (1962) for the seminal discussion of this problem). The situation changes at date 2; i.e., after the contract has been assigned to one of the sellers (the "contractor" C).[12] If the contractor reveals a flaw to the buyer now, the buyer cannot simply change the design in order to fix the flaw but has to renegotiate the initial contract with the contractor. Now the parties are in a bilateral monopoly position, and the contractor gets some share of the surplus from renegotiation.

We model the renegotiation game in reduced form by applying the Generalized Nash Bargaining Solution (GNBS). Suppose that the set of flaws $\tilde{F} = \tilde{F}^1 \cup \tilde{F}^2$ had been reported at date 1 and that seller i received the contract to produce $D(\tilde{F})$. The threatpoint of the renegotiation game is that renegotiation fails and that the initial contract is carried out. The surplus from renegotiating all flaws $f_k \in \hat{F}^i\setminus\tilde{F}$ at date 2, i.e. all flaws that seller i observed but that have not been reported at date 1, is given by

$$S^R(\hat{F}^i, D(\tilde{F})) = \sum_{\{k|f_k \in \hat{F}^i\setminus\tilde{F}\}} S_k^R \tag{3}$$

Let $\alpha \in (0,1)$ denote the bargaining power of the buyer. Then the seller's payoff from renegotiating is $(1-\alpha)S^R(\hat{F}^i, D(\tilde{F}))$.

[11] A sufficient condition for this to be the case is $V(D_0) - \sum_{k|f_k \in \mathcal{F}} \Delta v_k - \overline{c} > 0$.

[12] This is what Williamson (1985, p. 61-63), has termed the "fundamental transformation".

If a flaw f_k, $k \in \{1, \ldots, n\}$, exists but is not reported (either because no seller observed it or because a seller who observed it did not report it), then the flaw becomes apparent at date 3 when the project is (to a large degree) completed. If the buyer wants to change the design now, she can write a new contract with a (potentially) new seller. At this stage all sellers are equally good at fixing the problem and the initial contract no longer binds the buyer to seller i. We do not model this stage explicitly but assume that if this stage is reached, flaw f_k reduces the net utility of the buyer by Δv_k.

Finally, sellers are protected by limited liability, i.e. they can declare bankruptcy to avoid making negative profits. If the contractor declares bankruptcy, the project will not be completed and both parties get a payoff of zero.

The time structure of the model is summarized as follows.

Date 0: The buyer B announces that she wants to procure a good with design D_0. Nature determines the set of actual flaws F drawn from $\mathcal{P}(\mathcal{F})$ according to cdf $G(F)$, the set of flaws $\hat{F}^i$ that are observed by each seller $i \in \{1, 2\}$ according to cdf Q_F, and each seller's cost type c^i drawn from $[\underline{c}, \bar{c}]$ according to cdf $H(c^1, c^2)$.

Date 1: A procurement mechanism is proposed by B and executed. The mechanism may ask sellers to reveal their types $(c^i, \hat{F}^i)$. Given the set of all reported flaws $\tilde{F} = \tilde{F}^1 \cup \tilde{F}^2$ it determines the design $D(\tilde{F})$ of the good, which seller i becomes the contractor C, and which payments are made.

Date 2: B and C may engage in contract renegotiation if the contractor observed a flaw that has not been revealed at stage 1.

Date 3: The project is completed and payoffs are realized.

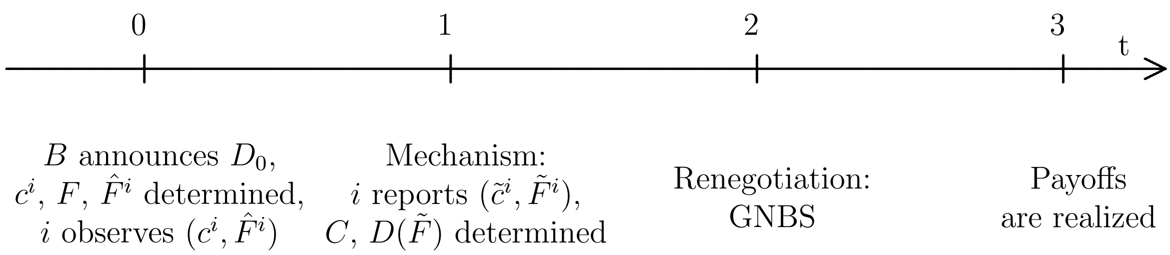

Figure 1: Time structure

Discussion of modeling assumptions:

1. *Common knowledge of the information structure.* The mechanism design literature typically assumes that the information structure of the underlying game is common knowledge of all players, i.e. all players know the set of possible flaws $\mathcal{F}$, their payoff consequences, and the probability distributions $G(F)$, $Q_F(\hat{F}^1, \hat{F}^2)$, and $H(c^1, c^2)$. In our set-up this does not make much sense. A sophisticated buyer may be aware that flaws may exist in the initial design, but she doesn't know how these flaws look like, otherwise she could find them herself.

   To deal with this problem we proceed in two steps. In the first step, we follow the standard mechanism design literature by assuming that the structure of the game is common knowledge and solve for the efficient, cost-minimizing mechanism. This yields a benchmark for what the buyer can achieve if she is very well informed. In a second step, we relax the informational requirements and assume that neither the buyer nor the sellers know the set of possible flaws and their payoff consequences. We show that there exists an indirect mechanism that implements the efficient allocation at the same cost to the buyer as the efficient, cost-minimizing mechanism of step 1.

2. *Timing of the discovery of flaws.* The model assumes that sellers observe flaws only at date 0, i.e. before the contract is assigned. Thus, if the contractor reports a flaw at date 2 in order to renegotiate, the buyer knows that the seller must have known this flaw early on. In the real world not all flaws are discovered early but some are discovered late. It would be straightforward (but notationally cumbersome) to extend the model by allowing for late discoveries of design flaws as well. In this extended model a seller who discovered a design flaw at date 0 will always claim that he observed it at date 2, and the buyer cannot find out when the flaw was discovered.[13] For simplicity we do not

[13]If the buyer could prove that the seller withheld the information on purpose and that he foresaw the harm caused to the buyer, a court of law could interpret this as a violation of the seller's "duty of care". In the extended version of the model proposed above the buyer is not able to prove this. In the extended version of the model sellers may bid more aggressively because they hope to find additional flaws later that can be renegotiated. As long as all sellers have the same expectations about finding flaws later this is not a problem because it will simply lower all bids by the same amount. We are grateful to Paul Klemperer for pointing this out.

present this extended model formally but assume that flaws are observed only at date 0. After all, these are the flaws that we care about because we want to induce sellers to report these flaws early in order to save the additional cost. If a flaw is discovered late, the higher cost to fix it is unavoidable.

3. *No commitment not to renegotiate.* If the buyer could commit never to renegotiate, the sellers would not be able to profit from withholding their information and they might as well report all observed design flaws early. However, if there is an opportunity for a Pareto improvement at date 2, the parties can always tear up the old contract and renegotiate a new one. Furthermore, if there are design flaws that are discovered late (as discussed above), a commitment not to renegotiate could be very harmful. We also exclude the possibility to "design" the renegotiation game in the mechanism at date 1 such that the buyer gets all the bargaining power.[14]

4 Three Inefficiencies

In this section we discuss a simple example to point out the inefficiencies that arise if the buyer naïvely uses a standard price-only auction to allocate the procurement contract. In this example we assume that there is only one possible flaw denoted by f that exists with probability p. If the flaw exists, each seller independently observes it with probability q in which case $\hat{F}^i = \{f\}$. With probability $1-q$ seller i observes nothing, so $\hat{F}^i = \emptyset$. We also assume that $\Delta c = 0$. Thus, if there is a design flaw and if the flaw is reported early (at date 1), then the problem can be solved at no additional cost. However, if the flaw is reported late (at date 2) and the parties have to renegotiate, then there is an inefficiency Δx, with $\Delta v > \Delta x > 0$.

Suppose that the buyer uses a sealed-bid, second-price auction. If the probability of a design flaw is 0, this auction implements the efficient allocation at the lowest possible cost for the buyer.

Before the auction is conducted, the buyer may ask the two sellers whether they have detected a flaw in design D_0. If none of the suppliers reports a flaw, the buyer auctions off

[14] See Hart (1995), p. 77-78, for a detailed discussion of these issues.

the procurement contract for design D_0. If at least one of the suppliers reports the flaw f, the buyer fixes the design and puts the improved design $D(f) = D_f$ up for auction.

Observation 1. *Suppose the buyer uses a sealed-bid, second-price auction to allocate the procurement contract. Given the outcome of the reduced form renegotiation game there is a unique Perfect Bayesian Equilibrium in weakly dominant bidding strategies. Any seller who detected flaw f conceals it ex ante and bids*

$$b(c^i, \hat{F}^i) = \begin{cases} c^i - (1-\alpha)S^R & \text{if } D = D_0 \text{ and } \hat{F}^i = \{f\} \\ c^i & \text{otherwise} \end{cases}.$$

Suppose that seller 1 is the only one who observed the flaw and that he reports it to the buyer ex ante. Then the buyer adjusts the design to D_f without rewarding seller 1. If, on the other hand, seller 1 conceals the flaw, a contract for design D_0 is awarded initially. Now, if seller 1 wins the contract he can make an additional profit ex post by renegotiating the contract. This allows him to bid more aggressively in the auction. Hence, by concealing the flaw seller 1 increases both the probability of winning the contract and the expected profit from being awarded the contract. Hence, a seller who spotted the flaw has a strict incentive to conceal this information ex ante.

This behavior of an informed seller can trigger three inefficiencies: First, because the flaw is only revealed at date 2, the parties have to renegotiate with positive probability and incur the renegotiation cost $\Delta x > 0$. Second, if $c^i > c^j$ and i observes the flaw while j does not, it may happen that $c^i - (1-\alpha)S^R < c^j$. In this case the seller with the higher cost gets the contract which is inefficient – i.e., the auction does not achieve efficient production. Finally, if $c^i > c^j$ and i observes the flaw while j does not, it may happen that $c^i - (1-\alpha)S^R > c^j$. In this case the lower cost seller gets the contract, but the flaw is not fixed and $S^R = \Delta v - \Delta x$ cannot be realized.

Proposition 1 (Three Inefficiencies)**.** *Suppose the buyer uses a sealed-bid, second-price auction to allocate the procurement contract. Then, three inefficiencies arise.*

1. *Inefficient Renegotiation:*
   With probability $pq\left[q + 2(1-q) \times Prob(c^i - (1-\alpha)S^R < c^j)\right]$ there is a design flaw that

is detected by at least one seller who wins the auction. In this case the design flaw is fixed via renegotiation which is inefficient because the parties have to incur $\Delta x > 0$.

2. *Inefficient Production:*
   With probability $2pq(1-q) \times Prob(c^j < c^i < c^j + (1-\alpha)S^R)$ there is a design flaw that is detected by one seller, this seller has higher production costs but wins the auction because of his expected profit in the renegotiation game. In this case production is carried out inefficiently by the seller with the higher cost.

3. *Inefficient Design:*
   With probability $2pq(1-q) \times Prob(c^j < c^i - (1-\alpha)S^R)$ there is a design flaw that is detected by one seller, but this seller does not win the auction. In this case the design flaw is not reported to the buyer and it cannot be fixed, so the buyer has to incur the loss of Δv which is inefficient.

The question is whether there is a mechanism that avoids these inefficiencies, i.e. a mechanism that induces both sellers to report all observed design flaws at date 1 and that always allocates the contract to the seller with the lowest cost. The next section shows how this can be done.

5 The Optimal Mechanism Design Problem

In this section we assume that the model is common knowledge, so all parties know the set of possible flaws, their payoff consequences and the underlying probability distributions. We derive the direct mechanism that induces both sellers to reveal their private information in ex post equilibrium at date 1 and that implements the efficient allocation at the lowest possible cost for the buyer. In the next section we show that there exists an indirect mechanism that implements the same allocation but that is informationally robust in the sense that it does not require the mechanism designer to know any of the parameters of the model when she sets up the mechanism.

In subsection 5.1 we focus on how to optimally induce sellers to report a design flaw early. We abstract from cost differences between firms and assume that $c^1 = c^2 = c$. Furthermore, we restrict attention to only one design flaw. In subsection 5.2 we return to the general model and allow for cost differences between sellers and multiple flaws.

5.1 Inducing Sellers to Report a Design Flaw Early

In this section we assume that both firms have identical costs c that are common knowledge and that there is only one possible flaw f that exists with probability p and is observed by each seller independently with probability q. Thus, the only problem is to induce sellers to report an observed flaw early. Consider the problem of a mechanism designer who knows the structure of the model and the values of v, c, Δv, Δc, and Δx, and the probabilities p and q. Each seller $i \in \{1, 2\}$ is one of two possible types. If he did not observe a design flaw, he is (with a slight abuse of notion) of type $\emptyset$. This is the case if either there is no flaw (in state $F = \emptyset$) or if there is a flaw (in state $F = \{f\}$) but the seller did not observe it. If the seller observed the flaw, he is of type f.

A direct mechanism asks each seller i to send a message $\tilde{F}^i \in \{\emptyset, f\}$. In words, a seller either claims that he did not observe anything ($\tilde{F}^i = \emptyset$) or he reports that he observed the flaw ($\tilde{F}^i = f$). While message $\tilde{F}^i = \emptyset$ can always be sent, message $\tilde{F}^i = f$ is feasible only if supplier i indeed observed the flaw f (is of type f). This is a mechanism design problem with "partially verifiable information" (Green and Laffont, 1986). In such a setup and with multiple agents, the *revelation principle* can be applied if the evidentiary structure (the set of feasible reports) is strongly normal (Bull and Watson, 2007). It is straightforward to check that this condition is satisfied in our model, so we can restrict attention to direct mechanisms.[15] A direct mechanism asks each seller to report his type. It specifies the design $D^i \in \{D_0, D_f\}$ of the good, probabilities ω^i with which seller i has to deliver good D^i, and transfers t^i, paid by the buyer and received by seller i, $i \in \{1, 2\}$, that depend on both messages. Because the

[15] By focusing on single agent problems, Green and Laffont (1986) have shown that in mechanism design problems with partially verifiable information the Revelation Principle applies if the so-called "Nested Range Condition (NRC)" is satisfied. This result has been extended to the class of mechanism design problems with n agents by Bull and Watson (2007). The equivalent to the NRC in the case with n agents is (strong) evidentiary normality. An evidentiary structure is called strongly normal if (i) there is a report that can be sent by any type ($\tilde{F}^i = \emptyset$ in our model), and (ii) if a type θ_1 can claim to be of type θ_2 and type θ_2 can claim to be of type θ_3, then also type θ_1 can claim to be of type θ_3. In our model, a seller can report a flaw only if he observed it. Thus, if $\hat{F}_2^i \in \mathcal{P}(\hat{F}_1^i)$ and $\hat{F}_3^i \in \mathcal{P}(\hat{F}_2^i)$, then we also have $\hat{F}_3^i \in \mathcal{P}(\hat{F}_1^i)$. Thus, strong evidentiary normality is satisfied.

problem is symmetric we restrict attention to symmetric mechanisms, i.e.

$$D^i = D(\tilde{F}^i, \tilde{F}^j), \quad \omega^i = \omega(\tilde{F}^i, \tilde{F}^j) \in [0,1], \quad \text{and} \quad t^i = t(\tilde{F}^i, \tilde{F}^j) \in \mathbb{R}$$

with $j \neq i$.

The mechanism designer wants to implement an efficient outcome. This requires:

(i) The specified design is optimal given the available information (efficient design, ED)

$$D(\tilde{F}^1, \tilde{F}^2) = \begin{cases} D_0 & \text{if } \tilde{F}^1 = \tilde{F}^2 = \emptyset \\ D_f & \text{otherwise} \end{cases} \tag{ED}$$

(ii) Production takes place with probability one (efficient production, EP)

$$\omega(\tilde{F}^1, \tilde{F}^2) + \omega(\tilde{F}^2, \tilde{F}^1) = 1 \quad \forall\, \tilde{F}^1, \tilde{F}^2 \in F \tag{EP}$$

(iii) An informed seller wants to reveal the state truthfully no matter what is reported by the other seller (ex post incentive compatibility, EPIC)

$$\begin{aligned} t(f,f) - \omega(f,f)(c + \Delta c) &\geq t(\emptyset, f) - \omega(\emptyset, f)(c + \Delta c) \\ t(f,\emptyset) - \omega(f,\emptyset)(c + \Delta c) &\geq t(\emptyset,\emptyset) - \omega(\emptyset,\emptyset)[c - (1-\alpha)S^R)] \end{aligned} \tag{EPIC}$$

(iv) Sellers always make non-negative profits because they are protected by limited liability (LL)

$$t(\tilde{F}^i, \tilde{F}^j) - \omega(\tilde{F}^i, \tilde{F}^j)\; c(D(\tilde{F}^1, \tilde{F}^2)) \geq 0 \quad \forall\, \tilde{F}^i, \tilde{F}^j \in F \tag{LL}$$

where $c(D(\tilde{F}^1, \tilde{F}^2)) = c$ if $\tilde{F}^1 = \tilde{F}^2 = \emptyset$ and $c(D(\tilde{F}^1, \tilde{F}^2)) = c + \Delta c$ otherwise. Note that (LL) implies that all sellers voluntarily participate in the mechanism (individual rationality).

We want to find a mechanism that implements the efficient allocation at the lowest possible cost to the buyer. Thus, the mechanism design problem can be stated as follows:

$$\min_{\omega(\cdot,\cdot), t(\cdot,\cdot)} \quad 2t(\emptyset,\emptyset)[1 - p + p(1-q)^2] + 2t(f,f)pq^2 + 2[t(f,\emptyset) + t(\emptyset,f)]p(1-q)q$$

subject to (ED), (EP), (EPIC) and (LL).

Let $u(\tilde{F}^i, \tilde{F}^j)$ denote the payoff that supplier i obtains if $\tilde{F}^i$ and $\tilde{F}^j$ are reported (not including any additional payoffs from renegotiation at date 2), i.e.

$$u(\tilde{F}^i, \tilde{F}^j) \equiv t(\tilde{F}^i, \tilde{F}^j) - \omega(\tilde{F}^i, \tilde{F}^j)c(D(\tilde{F}^i, \tilde{F}^j)). \tag{4}$$

Limited liability is satisfied if and only if for all $\tilde{F}^i, \tilde{F}^j \in \{\emptyset, f\}$ it holds that $u(\tilde{F}^i, \tilde{F}^j) \geq 0$. With this notation ex post incentive compatibility can be written as

$$\begin{aligned} u(f, f) &\geq u(\emptyset, f) \\ u(f, \emptyset) &\geq u(\emptyset, \emptyset) + \omega(\emptyset, \emptyset)(1-\alpha)S^R \end{aligned} \tag{EPIC}$$

Notice that symmetry together with (EP) implies that $\omega(\emptyset, \emptyset) = 1/2$. Obviously the buyer has an incentive to choose $u(\emptyset, \tilde{F}^j) = 0$ for all $\tilde{F}^j \in \{\emptyset, f\}$: Doing so relaxes the (EPIC) constraints and reduces the expected transfers to the sellers. Hence, the mechanism design problem simplifies to

$$\min_{u(f,f),\, u(f,\emptyset)} \quad 2qu(f, f) + 2(1-q)u(f, \emptyset)$$

subject to:

$$u(f, \emptyset) \geq (1-\alpha)S^R/2 \tag{EPIC}$$

$$u(f, f) \geq 0, \qquad u(f, \emptyset) \geq 0 \tag{LL}$$

The EPIC constraint must hold with equality in the optimal solution, so the following pair of payoffs solves the problem:

$$u^*(f, f) = 0, \qquad u^*(f, \emptyset) = (1-\alpha)S^R/2.$$

Proposition 2 (Optimal Direct Mechanism)**.** *The following efficient direct mechanism induces each seller of type f to report his type truthfully in ex post equilibrium at the lowest possible cost to the buyer:*

$$\begin{aligned}
&D^*(\emptyset, \emptyset) = D_0, && \omega^*(\emptyset, \emptyset) = \frac{1}{2}, && t^*(\emptyset, \emptyset) = \frac{c}{2} \\
&D^*(\emptyset, f) = D_f, && \omega^*(\emptyset, f) = 0, && t^*(\emptyset, f) = 0 \\
&D^*(f, \emptyset) = D_f, && \omega^*(f, \emptyset) = 1, && t^*(f, \emptyset) = c + \Delta c + \frac{(1-\alpha)S^R}{2} \\
&D^*(f, f) = D_f, && \omega^*(f, f) = \frac{1}{2}, && t^*(f, f) = \frac{c + \Delta c}{2}
\end{aligned}$$

Note that $t^*(\emptyset, \emptyset)$ and $t^*(f, f)$ are the expected transfers. If both sellers make the same announcement, each seller gets the contract with probability 0.5. In order to satisfy ex post limited liability, the seller who produces the good is reimbursed his cost, while the other seller receives nothing.

The mechanism of Proposition 2 is very intuitive. If one seller reports the flaw while the other one does not, then the former produces the good with the adjusted design D_f and gets a strictly positive rent, while the latter gets a utility of zero. This rent is necessary to induce the seller to reveal his information ex ante rather than to wait and renegotiate after having received the contract. Note that if the seller claims $\emptyset$, then – given that the other seller also claims $\emptyset$ – he gets the contract only with probability 1/2. Thus, the rent that has to be paid is only $\frac{1}{2}(1 - \alpha)S^R$. With N sellers this rent can be reduced to $\frac{1}{N}(1 - \alpha)S^R$ because the probability of getting the contract if all sellers report $\emptyset$ is only $1/N$. If two sellers (or more) report f, there is no need to pay a rent to these sellers because each of them would receive the contract with probability zero if he claimed to be of type $\emptyset$.

The mechanism of Proposition 2 is ex post incentive compatible, i.e. no seller has an incentive to change his report after observing what the other seller has reported.[16] Mechanisms that are ex post incentive compatible have the desirable feature that they do not depend on the beliefs or higher order beliefs of the players. Thus, in a model with a richer type space in which players have beliefs about the beliefs of their opponents, an ex post incentive compatible mechanism implements the desired allocation no matter how these beliefs look like.[17]

An interesting question is whether the buyer can implement the efficient allocation at a lower cost if she does not require ex post incentive compatibility but only Bayesian incentive compatibility. The answer is no. There are other optimal mechanisms that implement the efficient allocation that are Bayesian but not ex post incentive compatible. In fact, any pair of utilities $u(f, f), u(f, \emptyset) \geq 0$ satisfying

$$qu(f, f) + (1 - q)u(f, \emptyset) = (1 - q)(1 - \alpha)S^R/2 \tag{5}$$

does the job, but at the same expected cost as the mechanism of Proposition 2. This is because

[16]The mechanism of Proposition 2 does not implement the efficient allocation in dominant strategies because it relies on the application of the GNBS off the equilibrium path. Non-cooperative bargaining models that offer a foundation of the GNBS typically do not have equilibria in dominant strategies.

[17]See Bergemann and Morris (2005).

any Bayesian incentive compatible mechanism must give a seller who observed the flaw the expected renegotiation rent that he would receive if he did not reveal it. The mechanism of Proposition 2 gives him exactly this rent.

5.2 Multiple Flaws and Seller Heterogeneity

Now we generalize the optimal mechanism derived in Subsection 5.1 in two directions. First, sellers may have different costs which are private information. Second, we allow for multiple design flaws, i.e. each seller observes some subset $\hat{F}^i$ of the set of actual flaws F. Efficiency requires that the seller with the lowest cost produces the good and that both sellers are induced to reveal all flaws that they observed at date 1.

This multi-dimensional mechanism design problem is more intricate. Sellers have to be induced to report both the observed design flaws and their cost parameter truthfully. In Subsection 5.1 the optimal mechanism allocates the contract by a coin flip if both producers claim not to have observed any flaw. This minimizes the information rent that has to be paid to a seller. If both sellers have the same cost, a random allocation is efficient, but if sellers have different costs, efficiency requires that the seller with the lower cost gets the contract with probability 1. We show that this increases the information rent that has to be paid to the seller.

We return to the general model described in Section 3. As in Subsection 5.1 we assume that the mechanism designer knows all parameters of the model, i.e. the set of possible cost types $[\underline{c}, \bar{c}]$, the set of possible flaws $\mathcal{F} = \{f_1, \ldots, f_n\}$, the costs Δv_k, Δc_k and Δx_k associated with each potential flaw $k \in \{1, \ldots, n\}$, and the (conditional) probability distributions $G(F)$, $H(c)$, and Q_F. Note that the type of seller i is now multi-dimensional: it consists of a cost type c^i and an information type $\hat{F}^i$ (the set of flaws that seller i observed). Thus, the type of seller i is $(c^i, \hat{F}^i) \in [\underline{c}, \bar{c}] \times \mathcal{P}(F)$.[18]

By the revelation principle we can restrict attention to direct mechanisms that ask each seller i to send a message $(\tilde{c}^i, \tilde{F}^i) \in [\underline{c}, \bar{c}] \times \mathcal{P}(F)$. Put verbally, each seller i reports a cost $\tilde{c}^i$

[18]Note that the two dimensions of the seller's type are independent of each other. His cost type affects his valuation for completing project D_0. His "flaw type" affects his "outside option" utility if he does not report the flaw early and renegotiates, i.e. the transfer he must get to reveal all observed flaws early.

and a set of observed flaws $\tilde{F}^i$. Supplier i is free to report any cost $\tilde{c}^i \in [\underline{c}, \bar{c}]$ (i.e. costs cannot be verified), but he is restricted to report only flaws that he observed. A reported flaw can be verified, but the seller can always choose not to report some or all of the detected flaws, i.e. Assumption 1 applies.

The (symmetric) direct mechanism specifies for any announced types $((\tilde{c}^1, \tilde{F}^1), (\tilde{c}^2, \tilde{F}^2))$ a design $D = D((\tilde{c}^1, \tilde{F}^1), (\tilde{c}^2, \tilde{F}^2))$, a transfer $t^i = t((\tilde{c}^i, \tilde{F}^i), (\tilde{c}^j, \tilde{F}^j))$ paid by the buyer and received by seller i, and a probability $\omega^i = \omega((\tilde{c}^i, \tilde{F}^i), (\tilde{c}^j, \tilde{F}^j))$ with which seller i gets the contract – i.e., with probability ω^i seller i has to produce the good, for $i, j \in \{1, 2\}$ and $i \neq j$.

The mechanism designer seeks to induce an efficient outcome. The mechanism has to satisfy the following constraints:

(i) The design is optimal given the available information (efficient design, ED)

$$D((c^1, \hat{F}^1), (c^2, \hat{F}^2)) = D(\hat{F}^1 \cup \hat{F}^2) \tag{ED}$$

(ii) The good is produced by the seller with the lowest cost (efficient production, EP)

$$\omega((c^i, \hat{F}^i), (c^j, \hat{F}^j)) = \begin{cases} 1 & \text{if } c^i < c^j \\ 1/2 & \text{if } c^i = c^j \\ 0 & \text{if } c^i > c^j \end{cases} \tag{EP}$$

(iii) Sellers make non-negative profits (limited liability, LL)

$$\forall\, i \in \{1, 2\}, \hat{F}^i, \hat{F}^j \subseteq F,\ c^i, c^j \in [\underline{c}, \bar{c}]\ :$$
$$t((c^i, \hat{F}^i), (c^j, \hat{F}^j)) - \omega((c^i, \hat{F}^i), (c^j, \hat{F}^j)) \left[c^i + \sum_{\{k | f_k \in \hat{F}^i \cup \hat{F}^j\}} \Delta c_k \right] \geq 0. \tag{LL}$$

Note that (LL) implies that all sellers voluntarily participate in the mechanism (individual rationality).

(iv) It is ex post optimal for each seller to reveal his type truthfully (ex post incentive compatibility, EPIC)

$$\begin{aligned}
&\forall\ i \in \{1,2\},\ \hat{F}^i, \hat{F}^j \subseteq F,\ \text{and } c^i, c^j \in [\underline{c}, \bar{c}]\ :\\
&\qquad t((c^i,\hat{F}^i),(c^j,\hat{F}^j)) - \omega((c^i,\hat{F}^i),(c^j,\hat{F}^j))\left[c^i + \sum_{\{k|f_k\in\hat{F}^i\cup\hat{F}^j\}} \Delta c_k\right] \geq \\
&\max_{\tilde{F}^i\subseteq\hat{F}^i,\tilde{c}^i\in[\underline{c},\bar{c}]} t((\tilde{c}^i,\tilde{F}^i),(c^j,\hat{F}^j)) - \omega((\tilde{c}^i,\tilde{F}^i),(c^j,\hat{F}^j))\left[c^i + \sum_{\{k|f_k\in\tilde{F}^i\cup\hat{F}^j\}} \Delta c_k\right]\\
&\qquad\qquad + \omega((\tilde{c}^i,\tilde{F}^i),(c^j,\hat{F}^j))(1-\alpha)S^R(\hat{F}^i, D(\tilde{F}^i\cup\hat{F}^j))\ . \qquad \text{(EPIC)}
\end{aligned}$$

Ex post incentive compatibility (EPIC) ensures that no seller wants to change his report once he learns the report of the other seller. Furthermore, it implies that the mechanism does not depend on the beliefs or higher order beliefs of the participants.

Proposition 3 (Efficient Direct Mechanism)**.** *The following ex post incentive compatible direct mechanism implements the efficient allocation; i.e., it satisfies (ED), (EP), (LL), and (EPIC):*

$$\begin{aligned}
&D^*((c^1,\hat{F}^1),(c^2,\hat{F}^2)) = D^*(\hat{F}^1,\hat{F}^2) = D(\hat{F}^1\cup\hat{F}^2)\\
&\omega^*((c^i,\hat{F}^i),(c^j,\hat{F}^j)) = \omega^*(c^i,c^j) = \begin{cases} 1 & \textit{if}\ \ c^i < c^j \\ 1/2 & \textit{if}\ \ c^i = c^j \\ 0 & \textit{if}\ \ c^i > c^j \end{cases}\\
&t^*((c^i,\hat{F}^i),(c^j,\hat{F}^j)) = \omega^*(c^i,c^j)\left[c^j + \sum_{\{k|f_k\in\hat{F}^1\cup\hat{F}^2\}} \Delta c_k\right] + (1-\alpha)S^R(\hat{F}^i, D(\hat{F}^j))
\end{aligned}$$

Each seller i has to be induced to report his information $\hat{F}^i$ and his cost c^i truthfully. Note that the direct mechanism of Proposition 3 separates these two problems. It induces the seller to report his information $\hat{F}^i$ by paying him $(1-\alpha)S^R(\hat{F}^i, D(\hat{F}^j))$ if he reports $\hat{F}^i$. This is exactly the rent that seller i can obtain by revealing all observed flaws that have not been revealed by seller j ex post – at the renegotiation stage – rather than ex ante. It induces the seller to report his cost truthfully by allocating the contract to him if he reports the lower cost at a price that is equal to the cost of the second lowest bidder (as in a Vickrey auction). Thus, the mechanism is ex post incentive compatible: no seller has an incentive to revise his decision after learning the announcement $\tilde{c}^j, \tilde{F}^j$ of the other seller.

It is instructive to compare this mechanism to the mechanism of Proposition 2. Suppose that there is at most one flaw and this flaw is reported by seller i but not by seller j. Then, according to the mechanism of Proposition 3, seller i obtains a rent of $(1-\alpha)S^R$. This rent is larger than the rent $\frac{1}{2}(1-\alpha)S^R$, which is paid by the direct mechanism of Proposition 2. The reason is that in Proposition 2 the contract was allocated randomly if both sellers claim to be of type $\emptyset$, while the mechanism of Proposition 3 must allocate the contract to the seller with the lowest cost with probability one.

The next proposition shows that among all ex post incentive compatible mechanisms the direct mechanism of Proposition 3 is indeed optimal.

Proposition 4 (Optimal Mechanism)**.** *The mechanism of Proposition 3 is optimal, i.e. there does not exist any other ex post incentive compatible direct mechanism that implements the efficient outcome at a lower expected cost for the buyer.*

The intuition for this result is simple. The mechanism has to be ex post incentive compatible, i.e. seller i has to be induced to report all observed flaws, no matter what type $(c^j, \hat{F}^j)$ is reported by seller j. In particular, seller i has to report $\hat{F}^i$ truthfully if he wins the contract. With the mechanism of Proposition 3 all incentive constraints are binding, i.e. a seller who wins the contract is just indifferent whether to report any or all of the flaws that he observed. Thus, it is not possible to reduce the payment to a seller for revealing his flaws any further. Moreover, each seller is induced to report his cost type truthfully by using a Vickrey auction. It is well known that a Vickrey auction implements the efficient allocation at the lowest possible cost.

Would it be possible to implement the efficient allocation at a lower cost to the buyer if we imposed Bayesian rather than ex post incentive compatibility? This time the answer is yes. The buyer can exploit the fact that sellers types are correlated. Note that types must be correlated because a seller can observe a flaw only if a flaw exists, i.e. if seller i observes a flaw, his updated belief that the other seller also observed the flaw must increase. The optimal Bayesian incentive compatible mechanism punishes a seller who did not report flaw f_k, if this flaw was reported by his competitor.[19] This implies that it is no longer optimal to separate the elicitation of costs and the elicitation of flaws. Note, however, that, in order to optimally exploit the correlation the buyer needs to know the underlying probability distributions and

[19]This is similar to Crémer and McLean (1988). However, the buyer has to respect the limited liability constraints of the sellers.

the set of possible flaws and payoff consequences.

6 The Arbitration Mechanism

So far we assumed that the model is common knowledge, in particular that the mechanism designer knows all parameters of the problem when she sets up the mechanism. But, a typical buyer does not have this information. She is not aware of the design flaws, she does not know their payoff consequences, and she does not know the probability distributions over the realizations of these flaws nor how likely it is that each seller observed any given flaw. However, a sophisticated buyer is aware that she is unaware. She knows that mistakes happen and that they can be very costly if they are not fixed early. Thus, she would like to prepare for this possibility and to give incentives to sellers to reveal possible flaws at date 1 already.

In this section we assume that the buyer and the sellers do note know the set of possible flaws and their payoff consequences, nor do they have Bayesian priors $H(c^1, c^2)$, $G(F)$ and $Q_F(\hat{F}^1, \hat{F}^2)$. Nevertheless, we show that it is possible to implement the same allocation as the optimal direct mechanism of Proposition 3 if flaws are ex post verifiable.

Assumption 2 (Ex Post Verifiability)**.** *If a flaw f_k is pointed out by any seller, every industry expert can verify the flaw ex post, i.e. she establish that it is indeed a flaw and she can form an unbiased estimate of the payoff consequences of the flaw f_k: Δc_k, Δx_k, and Δv_k.*

Note that we still assume that flaws are only partially verifiable (Assumption 1, i.e. if a flaw is not reported by a seller it is impossible to verify whether the seller observed this flaw. In Section 5 we assumed that all parties (including the courts) know all possible flaws and their payoff consequences. Here we assume instead that if a flaw is pointed out all involved parties understand this flaw and can form an unbiased estimate of its payoff consequences.

If Assumption 2 holds the buyer can employ the following indirect mechanism that uses an industry expert as an independent arbitrator:

1) The buyer publicizes her initial design proposal D_0 and invites all potential sellers to evaluate the proposal and to report possible design flaws in sealed envelopes to an independent arbitrator.

2) Let $\tilde{F}^i$ denote the set of flaws reported by seller $i \in \{1,2\}$. The arbitrator evaluates all flaws $f_k \in \tilde{F} \equiv \tilde{F}^1 \cup \tilde{F}^2$ and their expected payoff consequences, i.e., for any reported flaw f_k she estimates Δv_k, Δc_k and Δx_k. She awards the following reward to each seller i:

$$T_1^i(\tilde{F}^i, \tilde{F}^j) = (1-\alpha)E[S^R(\tilde{F}^i, D(\tilde{F}^j))] \tag{6}$$

3) The buyer uses the information on the reported design flaws $\tilde{F} = \tilde{F}^i \cup \tilde{F}^j$ to redesign the good to $D(\tilde{F})$, and then runs a sealed-bid, second-price auction. Each seller $i \in \{1,2\}$ submits a bid b^i, the contract is allocated to the lowest bidder, and seller i receives

$$T_2^i(b^i, b^j) = \begin{cases} b^j & \text{if } b^i < b_j \\ b^j/2 & \text{if } b^i = b_j \\ 0 & \text{if } b^i > b^j \end{cases} . \tag{7}$$

If both sellers place the same bid $b^1 = b^2 = b$, one seller is selected at random and obtains the contract at price b.

4) If seller i got the contract and if he observed design flaws f_k that have not been revealed to the buyer at stage 1, he may renegotiate the contract with the buyer.

Proposition 5 (Arbitration Mechanism)**.** *If an independent arbitrator can verify flaws after they have been pointed out to her and assess their payoff consequences, then the Arbitration Mechanism implements the same allocation as the optimal mechanism of Proposition 3. Furthermore, it is informationally robust in the sense that it is ex post incentive compatible and that it does not require any prior knowledge of the set of possible flaws $\mathcal{F}$ and their payoff consequences Δv_k, Δc_k and Δx_k for all $k \in \{1, \ldots, n\}$, nor the probability distributions G, Q_F and H.*

The proof of Proposition 5 is straightforward. The Arbitration Mechanism mimics the optimal mechanism of Proposition 3 and yields the same payoffs in all states of the world. Therefore, it gives the same incentives to reveal the flaw and it implements the same allocation.

Informational robustness is a highly desirable property of the Arbitration Mechanism. Of course, the buyer must be aware that there could be design flaws, but she does not have to know how these flaws look like, what payoff consequences they imply, and what the probability distributions G, Q_F and H are. However, the buyer has to be able to assess her bargaining power α if the initial contract is renegotiated.

The independent arbitrator does not need to know these parameters of the model either. However, pointing out a flaw to her is an "eye-opener" as in (Tirole, 2009). Once the flaw has been pointed out, the arbitrator understands the flaw, she knows what is necessary to fix it, and she is able to assess $E[\Delta v_k]$, $E[\Delta c_k]$ and $E[\Delta x_k]$ for all k with $f_k \in \tilde{F}$. Note that it is not necessary that she assesses these parameters perfectly. If the sellers are risk neutral, it is sufficient that the arbitrator forms an unbiased estimate of these parameters.

Arbitrators are frequently used in the real world to settle conflicts in contractual relationships. To do this they have to be able to verify the state of the world ex post and to assess payoff consequences. Consider for example arbitration in the construction industry. A buyer may claim that a completed building suffers from construction flaws. An independent arbitrator has to evaluate whether these claims are justified and what has to be done to fix the flaws or to compensate the buyer. Thus, the arbitrator verifies the flaws and assesses their payoff consequences as in our model. The arbitrator will then assign payments according to a rule such as "specific performance" or "expectation damages" that has been agreed upon beforehand by the contracting partners or that is imposed by law. Our mechanism uses a different rule to assign payments, but the basic principle is the same.

The Arbitration Mechanism is a two-stage mechanism that separates the problems of (i) inducing sellers to reveal observed design flaws early, and (ii) allocating the contract to the seller with the lowest cost. This separation is necessary if the buyer wants to implement an efficient allocation. However, if the buyer does not want to implement the efficient allocation but rather wants to maximize her expected profits, she may want to tie the allocation of the contract to the revelation of design flaws (e.g. by offering bonus points that create an advantage in the auction for sellers who reveal flaws early) in order to reduce the rent that has to be paid to sellers. But, of course, this comes at the cost that the allocation is inefficient with positive probability. Furthermore, the buyer needs to know the possible flaws, their payoff consequences, and the underlying probability distributions to do this optimally.

The Arbitration Mechanism requires the commitment of the buyer to pay sellers for the information on design flaws that they provide. The simplest and most transparent way to do this is the use of an independent third party. However, there are also other ways how this commitment can be achieved. For example, if the buyer frequently procures similar projects, i.e., if she is in a repeated relationship with the sellers, and if the allocation procedure is fully transparent, then she may be able to credibly commit to paying out $T_1^i(\tilde{F}_i, \tilde{F}_j) = (1-\alpha)S^R(\tilde{F}_i, D(\tilde{F}_j))$ herself. If parties are sufficiently patient, this commitment is sustained by the threat of the sellers not to reveal any design flaws in the future if the buyer ever reneges on her promise.

7 Incentives to Invest in Finding Design Flaws

So far, we assumed that sellers receive the signal about design flaws for free. In reality finding flaws requires effort and other costly resources. Thus, the question arises whether the Arbitration Mechanism provides optimal incentives to invest into finding flaws.

We analyze the investment incentives of the two sellers for the baseline model with at most one flaw; i.e., $\mathcal{F} = \{\emptyset, \{f\}\}$. The flaw exists with probability $p \in (0,1)$. If the flaw exists and is detected and revealed ex ante, this creates a social surplus of $S = \Delta v - \Delta c$. If the flaw is revealed only at the renegotiation stage, the social surplus is reduced to $S^R = \Delta v - \Delta c - \Delta x > 0$.

Each seller i can invest resources in order to increase the probability of detecting the flaw (if it exists). The probability of detecting the flaw if it exists is q when seller i invests the amount $\phi^i(q)$, where $\phi^i(\cdot)$ is strictly increasing and convex. Given the flaw exists, the detection probabilities are assumed to be uncorrelated across sellers.

Recall that the Arbitration Mechanism separates the problems of inducing sellers to reveal flaws and of allocating the contract to the most efficient seller. Thus, we can focus on the profits a seller obtains from detecting and revealing the flaw. The expected profit of seller i under the Arbitration Mechanism (ignoring potential profits from production) is

$$\pi^i(q^i) = pq^i(1-\hat{q}^j)(1-\alpha)S^R - \phi^i(q^i), \tag{8}$$

where $\hat{q}^j$ with $j \neq i$ is the investment that seller i expects his competitor j to make. In the

Nash Equlibrium of the investment game seller i chooses

$$q^i_{NE} \in \arg\max_{q^i} \left\{ pq^i(1-\hat{q}^j)(1-\alpha)S^R - \phi^i(q^i) \right\} \tag{9}$$

Consider now the problem of a social planner who can choose q^1 and q^2 in order to maximize social welfare

$$W(q^1,q^2) = p(q^1+q^2-q^1q^2)S - \phi^1(q^1) - \phi^2(q^2). \tag{10}$$

The welfare optimal investments are

$$(q^1_W, q^2_W) \in arg \max_{q^1,q^2} \; W(q^1,q^2). \tag{11}$$

We assume that there exists a unique welfare maximizing tuple of investment levels $(q^1_w, q^2_W) \gg 0$.

The welfare maximizing investment levels do not coincide with the investment levels that sellers choose in Nash equilibrium. Suppose seller 1 expects that seller 2 invests efficiently. Then, the investment level of seller 1 must solve $p(1-q^2_W)(1-\alpha)S^R = d\phi^1/dq^1$. However, the welfare optimal investment level q^1_W solves the following first-order condition $p(1-q^2_W)S = d\phi^1/dq^1$. Thus, given that seller 2 invests efficiently, seller 1 has an incentive to invest too little. The reason is that with the Arbitration Mechanism seller i receives less than the social value generated by his efforts. He receives only share $(1-\alpha) < 1$ of the *renegotiation* surplus, which is the social surplus reduced by the cost of renegotiation. Thus, the Arbitration Mechanism induces sellers to underinvest into finding design flaws. This problem can be fixed by replacing T^i_1 in the Arbitration Mechanism by the full social surplus S; i.e., if seller i reports the flaw and seller j reports nothing, seller i obtains a payment of S and nothing otherwise. Now, the expected profit of seller i amounts to

$$\pi^i(q^i) = pq^i(1-\hat{q}^j)S - \phi^i(q^i) \tag{12}$$

Proposition 6 (Investment Incentives)**.** *Suppose the Arbitration Mechanism specifies $\hat{T}^i_1(f,\emptyset) = S$ and $\hat{T}^i_1(f,f) = \hat{T}^i_1(\emptyset,f) = \hat{T}^i_1(\emptyset,\emptyset) = 0$. This mechanism induces both sellers to reveal their information truthfully, it allocates the contract to the seller with the lowest cost, and it induces both sellers to invest efficiently; i.e., $(q^1,q^2) = (q^1_W, q^2_W)$.*

The mechanism proposed by Proposition 6 is essentially a Groves mechanism: Each seller is made residual claimant for his contribution to the social surplus. Thus, there is always an equilibrium in which all sellers invest efficiently.[20]

It is interesting to note that the mechanism of Proposition 3 is not a Groves mechanism. If a seller reports flaw f_k early rather than late, he increases the social surplus by Δx_k. However, his reward for doing so is given by $(1-\alpha)S_k^R = (1-\alpha)[\Delta v_k - \Delta x_k - \Delta c_k]$. Thus, if Δx_k increases, the social surplus generated by the seller's action increases, but his reward decreases because the renegotiation surplus shrinks. The mechanism of Proposition 3 does not pay each seller his marginal contribution to the social surplus, but rather the increase of his outside option utility.

The mechanism of Proposition 7 implements the efficient investment levels in Bayesian Nash equilibrium. It does not require that the buyer knows the investment cost functions. Each seller, however, has to be able to anticipate the behavior of the other seller correctly. Thus, each seller has to know his competitor's investment cost functions, as well as the likelihood that the flaw exists and what its payoff consequences are.

The question of how to incentivize sellers to search for flaws efficiently raises many additional interesting questions. For example, it may be optimal to limit the number of sellers who are incentivized to search for flaws, or it may be optimal to let sellers search sequentially in order not to duplicate search efforts. However, all of these design choices require a detailed knowledge of what sellers are searching for, i.e. the buyer must know the parameters of the model and the underlying probability distributions. This is why these problems are orthogonal to the problem addressed in this paper.

8 Conclusions

An important problem for real world mechanism design is the fact that the mechanism designer is often unaware of some possible contingencies. An experienced mechanism designer

[20] Proposition 6 is very similar to Theorem 1 in Bergemann and Välimäki (2002). However, Bergemann and Välimäki (2002) assume that the signals observed by each seller are stochastically independent which is not the case in our model.

understands that she may have overlooked something (she is "aware that she is unaware"), but she does not know what it is that she has overlooked. The traditional mechanism design literature ignores this problem by assuming that "the model" is common knowledge. All involved parties know what the possible states of the world are, and they know the probability distributions with which nature determines the actual state (including the information structure). Following Wilson (1987) the literature on "robust implementation" focuses on mechanisms that do not depend on the probability distributions, i.e. on the beliefs and higher-order beliefs of the involved parties. Our paper goes one step further. We allow for the possibility that the mechanism designer has only partial knowledge of the physical state-space, i.e. she is unaware of some contingencies. Therefore, it is impossible for her to describe these contingencies in a contract or mechanism ex ante.

We have shown that, nevertheless, the mechanism designer can implement the efficient allocation if information that is revealed by the agents can be verified ex post, i.e., once a contingency has been pointed out, an industry expert can verify it and understands its payoff consequences. The mechanism designer can appoint an expert as an independent arbitrator who verifies the contingency and completes the contract ex post according to a general rule that does not require the ex ante knowledge of the state space. In the procurement context this is a reasonable assumption. Courts of arbitration are a well established instrument to settling disputes out of court. They frequently have to verify the state of the world ex post and to assign payments according to rules that have been specified in general terms ex ante. This is nothing new. The new idea here is to use an *Arbitration Mechanism* to induce sellers to reveal all observed design flaws early in order to implement the efficient allocation.

One potential problem of the *Arbitration Mechanism* is collusion. A seller who reports a set of flaws is not rewarded for all flaws that he detected, but only for those that have not been reported by other sellers. Thus, if sellers coordinated their reports so that no flaw is reported by more than one seller, they could benefit substantially. In some sense this is the same problem of collusion that arises in any auction. However, there is an interesting twist to it that makes collusion more difficult in our setup. In order to coordinate their behavior each seller has to point out the flaws that he observed to the other seller. Suppose that seller 1 observed a flaw that seller 2 did not observe. By pointing out the flaw to seller 2, seller 1 gives away his information. Seller 2 may now claim that he also observed the flaw and threaten to reveal it to the buyer in order to obtain concessions from seller 1 in the collusion game. An interesting question for future research is to model this in more detail and to ask whether the mechanism can be modified to make collusion more difficult.

The *Arbitration Mechanism* separates the problem of inducing sellers to report design flaws early and assigning the contract to the seller with the lowest cost. This separation is necessary to achieve efficiency. However if the buyer is not interested in implementing the efficient allocation but rather in maximizing profits, separation need no longer be optimal. In this case the buyer may be able to increase her expected profits by tying the assignment of the contract to the revelation of flaws, e.g. by offering "bonus points" in the auction in exchange for pointing out design improvements. Note, however, that in order to design the mechanism optimally, the buyer needs to know the probabilities of possible flaws and their payoff consequences. If she does not know this, she may get it wrong which can be very costly if sellers keep the information on design flaws to themselves. Thus, even for a profit maximizing buyer the *Arbitration Mechanism* is attractive, because it induces sellers to report all flaws they observed at a minimum cost to the buyer.

The *Arbitration Mechanism* relies on ex post verifiability which plays an important role in many other settings as well. For example, large parts of the legal system rely on the courts to verify the state of the world ex post and to assign an allocation according to a general rule specified in the law. Similarly, many companies use subjective performance evaluation to determine bonus payments for employees, i.e. the employee's performance is verified ex post and a bonus payment is determined according to a general rule. Thus, exploring the potential and the limits of ex post verifiability in implementation theory is an important topic for future research.

A Appendix

Proof of Observation 1. First, we investigate the bidding stage and thereafter the decision of a seller whether to reveal the flaw.

Suppose that at least one seller reported the flaw, so $D = D_f$. In this case there can be no renegotiation at date 2 and we have a standard second-price auction for design D_f.

Now suppose that neither of the two sellers reported the flaw, so $D = D_0$. The auction is again a standard second-price auction but now sellers may benefit from contract renegotiation. If a seller observed the flaw (but concealed it) and wins the contract, then there is scope for renegotiation.[21] Given the outcome of the reduced form renegotiation game, the seller's payoff from contract renegotiation is $(1-\alpha)[\Delta v - \Delta x] = (1-\alpha)S^R$. Let $c^i - (1-\alpha)S^R$ be seller i's *pseudo cost type* if he observed the flaw and c^i if he did not observe it. By applying standard arguments for bidding behavior in second-price auctions it can readily be established that it is a weakly dominant strategy to bid the pseudo cost type; i.e.

$$b(c^i, \hat{F}^i) = \begin{cases} c^i & \text{if } \hat{F}^i = \emptyset \\ c^i - (1-\alpha)S^R & \text{if } \hat{F}^i = \{f\} \end{cases}. \tag{A.1}$$

As a next step we analyze whether a seller who observed the flaw has an incentive to report it before the auction takes place (given that both suppliers use their weakly dominant bidding strategies at the subsequent auction stage). Consider seller 1 who observed the flaw and has cost c^1. Let $H(c^2|c^1)$ be the c.d.f. of seller 2's cost, conditional on seller 1's cost being c^1.

If seller 1 reports the flaw, design D_f is specified ex ante. The expected profit from revealing (REV) the flaw thus is

$$\Pi^1(REV|c^1) = \int_{c^1}^{\bar{c}} (c^2 - c^1) dH(c^2|c^1). \tag{A.2}$$

If seller 1 does not report the flaw, the subsequent auction game depends on the report of seller 2. Let $q(c^2) := Prob(\hat{F}^2 = \{f\}|c^2 \wedge \tilde{F}^2 = \emptyset)$ be the conditional probability that seller

[21]Note that the threatpoint of the renegotiation game in the GNBS is given by the initial contract. The ex post payoffs resulting from the GNBS are all positive, so the limited liability constraint does not become relevant. See Binmore, Rubinstein, and Wolinsky (1986, p. 185) on how to model the threatpoint of a negotiation game with the Nash Bargaining Solution.

2 with cost c_2 observed the flaw, conditional on not having reported it in equilibrium. Note that $0 \leq q(c^2) \leq q < 1$ for all c^2. With this notation, the expected profit of seller 1 from concealing (CON) the flaw is

$$\Pi^1(CON|c^1) = \int_{c^1}^{\bar{c}} q(c^2)(c^2 - c^1)dH(c^2|c^1) + \int_{\max\{c^1-(1-\alpha)S^R, \underline{c}\}}^{\bar{c}} [1 - q(c^2)][c^2 - c^1 + (1-\alpha)S^R]dH(c^2|c^1). \quad \text{(A.3)}$$

Notice that

$$\int_{\max\{c^1-(1-\alpha)S^R, \underline{c}\}}^{\bar{c}} (c^2 - c^1)dH(c^2|c^1) \geq \int_{c^1}^{\bar{c}} (c^2 - c^1)dH(c^2|c^1) \quad \text{(A.4)}$$

and

$$c^2 - c^1 + (1-\alpha)S^R > c^2 - c^1. \quad \text{(A.5)}$$

Thus, $\Pi^1(CON|c^1) > \Pi^1(REV|c^1)$ for all $c^1 \in [\underline{c}, \bar{c}]$, which completes the proof.

□

Proof of Proposition 1. The result is shown in the main text. □

Proof of Proposition 2. We have shown in the text above that the proposed mechanism satisfies all constraints and minimizes the expected cost of the buyer. Note that it is weakly optimal for each seller to report his type truthfully no matter what the announced type of his opponent is. Thus, the mechanism is ex post incentive compatible. □

Proof of Proposition **??**. The Arbitration Mechanism gives rise to the same monetary outcomes and the same incentives for each seller as the mechanism of Proposition 2. Thus, given that it is an equilibrium in the direct mechanism for each seller to reveal the design flaw early, it is also an equilibrium in the Arbitration Mechanism. The mechanism is informationally robust because it is independent of the probabilities p and q and because the reward for reporting a design flaw is determined *ex post* by the independent arbitrator. □

Proof of Proposition 3. The mechanism in the proposition statement obviously satisfies (ED), (EP), and (LL).Using the expression for the transfer payments, (EPIC) can be written as

$$\begin{aligned}\omega^*(c^i,c^j)[c^j-c^i]+(1-\alpha)S^R(\hat{F}^i,D(\hat{F}^j))\geq\\ \omega^*(\tilde{c}^i,c^j)[c^j-c^i+(1-\alpha)S^R(\hat{F}^i,D(\tilde{F}^i\cup\hat{F}^j))]+(1-\alpha)S^R(\tilde{F}^i,D(\hat{F}^j))\end{aligned}\quad (A.6)$$

for all $\tilde{c}^i\in[\underline{c},\bar{c}]$, $\tilde{F}^i\subseteq\hat{F}^i$. First, note that

$$\begin{aligned}&S^R(\tilde{F}^i,D(\hat{F}^j))+S^R(\hat{F}^i,D(\tilde{F}^i\cup\hat{F}^j))\\ &=\sum_{\{k|f_k\in\tilde{F}^i\setminus\hat{F}^j\}}\Delta v_k-\Delta c_k-\Delta x_k+\sum_{\{k|f_k\in\hat{F}^i\setminus(\hat{F}^j\cup\tilde{F}^i)\}}\Delta v_k-\Delta c_k-\Delta x_k\\ &=\sum_{\{k|f_k\in\hat{F}^i\setminus\hat{F}^j\}}\Delta v_k-\Delta c_k-\Delta x_k=S^R(\hat{F}^i,D(\hat{F}^j)).\end{aligned}\quad (A.7)$$

Using (A.7), (A.6) reduces to

$$\omega^*(c^i,c^j)[c^j-c^i]\ \geq\ \omega^*(\tilde{c}^i,c^j)[c^j-c^i]-[1-\omega^*(\tilde{c}^i,c^j)](1-\alpha)S^R(\hat{F}^i,D(\tilde{F}^i\cup\hat{F}^j))\quad (A.8)$$

for all $\tilde{c}^i\in[\underline{c},\bar{c}]$, $\tilde{F}^i\subseteq\hat{F}^i$.

Thus, given that $\omega^*\leq 1$ it never pays off for a seller to misreport the observed flaws. The remaining question is whether a seller can benefit from misreporting his cost. Given that the set of observed flaws is revealed truthfully, inequality (A.8) simplifies to

$$\omega^*(c^i,c^j)[c^j-c^i]\geq\omega^*(\tilde{c}^i,c^j)[c^j-c^i]\quad\forall\tilde{c}^i\in[\underline{c},\bar{c}].\quad (A.9)$$

If $c^i<c^j$, we have $\omega^*(c^i,c^j)=1$, so misreporting cannot be beneficial. If $c^i>c^j$, we have $\omega^*(c^i,c^j)=0$, so misreporting can lead to seller i getting the contract but this is not in seller i's interest because $[c^j-c^i]<0$. For the knife-edge case $c^j=c^i$, seller i is indifferent between all potential cost reports – i.e., reporting truthfully is a best response. □

Proof of Proposition 4. The optimal mechanism design problem of the buyer is given by:

$$\min_{t(\cdot,\cdot)}\quad\mathbb{E}[t((c^1,\hat{F}^1),(c^2,\hat{F}^2))+t((c^2,\hat{F}^2),(c^1,\hat{F}^1))]\quad (A.10)$$

subject to:

$$D((c^1, \hat{F}^1), (c^2, \hat{F}^2)) = D(\hat{F}^1, \hat{F}^2) = D(\hat{F}^1 \cup \hat{F}^2) \tag{ED}$$

$$\omega((c^1, \hat{F}^1), (c^2, \hat{F}^2)) = \omega(c^1, c^2) = \begin{cases} 1 & \text{if } c^1 < c^2 \\ 1/2 & \text{if } c^1 = c^2 \\ 0 & \text{if } c^1 > c^2 \end{cases} \tag{EP}$$

$$t((c^1, \hat{F}^1), (c^2, \hat{F}^2)) - \omega(c^1, c^2)\left[c^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] \geq 0 \tag{LL}$$

$$t((c^1, \hat{F}^1), (c^2, \hat{F}^2)) - \omega(c^1, c^2)\left[c^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] \geq \tag{EPIC}$$

$$t((\tilde{c}^1, \tilde{F}^1), (c^2, \hat{F}^2)) - \omega(\tilde{c}^1, c^2)\left[c^1 + \sum_{\{k|f_k \in \tilde{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] + \omega(\tilde{c}^1, c^2) \sum_{\{k|f_k \in \hat{F}^1 \setminus (\tilde{F}^1 \cup \hat{F}^2)\}} (1-\alpha) S_k^R$$

All constraints have to hold for both sellers and for all seller types.

Efficient Design (ED) and Efficient Production (EP) imply directly

$$D^*((c^1, \hat{F}^1), (c^2, \hat{F}^2)) = D^*(\hat{F}^1, \hat{F}^2) = D(\hat{F}^1 \cup \hat{F}^2)$$

$$\omega^*((c^i, \hat{F}^i), (c^j, \hat{F}^j)) = \omega^*(c^i, c^j) = \begin{cases} 1 & \text{if } \; c^i < c^j \\ 0 & \text{if } \; c^i > c^j \end{cases}$$

Consider now the implications of Ex Post Incentive Compatibility (EPIC) by focusing on the incentives of seller 1 to misreport his type (the incentives of seller 2 are symmetric). (EPIC) requires that for all $(c^1, \hat{F}^1) \in [\underline{c}, \bar{c}] \times \mathcal{P}(\mathcal{F})$ and for all $(c^2, \hat{F}^2) \in [\underline{c}, \bar{c}] \times \mathcal{P}(\mathcal{F})$ it must hold that:

$$\begin{aligned} & t((c^1, \hat{F}^1), (c^2, \hat{F}^2)) - \omega(c^1, c^2)\left[c^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] \geq \\ & t((\tilde{c}^1, \tilde{F}^1), (c^2, \hat{F}^2)) - \omega(\tilde{c}^1, c^2)\left[c^1 + \sum_{\{k|f_k \in \tilde{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] + \omega(\tilde{c}^1, c^2) \sum_{\{k|f_k \in \hat{F}^1 \setminus (\tilde{F}^1 \cup \hat{F}^2)\}} (1-\alpha) S_k^R \end{aligned} \tag{EPIC}$$

In the following, we derive conditions on transfers that need to be satisfied so that seller 1 has no incentive to misreport his type. We have to distinguish four cases.

Case (i) $[c^1 < c^2$ **and** $\tilde{c}^1 < c^2]$**:** Seller 1 is more efficient than seller 2. He must have no incentive to misreport his type by claiming to have a different cost $\tilde{c}^1 \neq c^1$ such that he is still selected as the contractor. This is the case iff

$$t(c^1, \hat{F}^1, \cdot) - [c^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k] \geq t(\tilde{c}^1, \tilde{F}^1, \cdot) - [c^1 + \sum_{\{k|f_k \in \tilde{F}^1 \cup \hat{F}^2\}} \Delta c_k] + \sum_{\{k|f_k \in \hat{F}^1 \setminus (\tilde{F}^1 \cup \hat{F}^2)\}} (1-\alpha) S_k^R$$

for all $\tilde{c}^1 < c^2$ and all $\tilde{F}^1 \subseteq \hat{F}^1$. Rearranging yields

$$t(c^1, \hat{F}^1, \cdot) - t(\tilde{c}^1, \tilde{F}^1, \cdot) \geq \sum_{\{k|f_k \in \hat{F}^1 \setminus (\tilde{F}^1 \cup \hat{F}^2)\}} \left[\Delta c_k + (1-\alpha) S_k^R\right] \quad \forall c^1, \tilde{c}^1 < c_2,\ \tilde{F}^1 \subseteq \hat{F}^1. \tag{A.11}$$

Case (ii) [$c^1 < c^2$ **and** $\tilde{c}^1 > c^2$]**:** If Seller 1 is more efficient than seller 2 it must also be the case that he has no incentive to report to be less efficient than seller 2, which is the case iff

$$t(c^1, \hat{F}^1, \cdot) - t(\tilde{c}^1, \tilde{F}^1, \cdot) \geq c^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k \quad \forall c^1 < c^2,\ \tilde{c}^1 > c^2,\ \tilde{F}^1 \subseteq \hat{F}^1. \tag{A.12}$$

Case (iii) [$c^1 > c^2$ **and** $\tilde{c}^1 > c^2$]**:** Seller 1 is less efficient than seller 2. He must have no incentive to report $\tilde{c}^1 \neq c^1$ such that he is still less efficient. In this case (EPIC) reduces to

$$t(c^1, \hat{F}^1, \cdot) - t(\tilde{c}^1, \tilde{F}^1, \cdot) \geq 0 \quad \forall c^1, \tilde{c}^1 > c^2,\ \tilde{F}^1 \subseteq \hat{F}^1. \tag{A.13}$$

The reverse condition has to hold to deter type $\tilde{c}^1$ from reporting to be type c^1. Furthermore these conditions have to hold for all $\hat{F}^1$ and $\tilde{F}^1 \subseteq \hat{F}^1$, so in particular for $\tilde{F}^1 = \hat{F}^1$. Thus, different types of seller 1 that report the same set of flaws and have different costs that are both higher than the cost of seller 2 must receive the same transfer: For all $c^1, \tilde{c}^1$ so that $\omega(c^1, \cdot) = \omega(\tilde{c}^1, \cdot) = 0$ we must have

$$t(c^1, \hat{F}^1, \cdot) = t(\tilde{c}^1, \hat{F}^1, \cdot) \quad \forall \hat{F}^1, \forall c^1, \tilde{c}^1 > c^2. \tag{A.14}$$

Case (iv) [$c^1 > c^2$ **and** $\tilde{c}^1 < c^2$]**:** If seller 1 is less efficient than seller 2 he must not have an incentive to report to be more efficient. The (EPIC) constraint in this case is equivalent to

$$t(c^1, \hat{F}^1, \cdot) - t(\tilde{c}^1, \tilde{F}^1, \cdot) \geq - \left[c^1 + \sum_{\{k|f_k \in \tilde{F}^1 \cup \hat{F}^2\}} \Delta c_k\right] + \sum_{\{k|f_k \in \hat{F}^1 \setminus (\tilde{F}^1 \cup \hat{F}^2)\}} (1-\alpha) S_k^R \quad \forall c^1 > c^2, \tilde{c}^1 < c^2,\ \tilde{F}^1 \subseteq \hat{F}^1. \tag{A.15}$$

We now turn to the limited liability constraint (LL). Suppose the seller reports $\tilde{F}^1 = \emptyset$. Reducing $t(\tilde{c}^1, \emptyset, \cdot)$ relaxes (EPIC), so the buyer wants to reduce the transfer until the (LL) constraint holds with equality. Thus, if $\tilde{c}^1 < c^2$ it is optimal to set

$$t(\tilde{c}^1, \emptyset, \cdot) = \tilde{c}^1 + \sum_{\{k|f_k \in \hat{F}^2\}} \Delta c_k.$$

Using this in inequality (A.11) for $\tilde{F}^1 = \emptyset$ yields

$$\begin{aligned} t(c^1, \hat{F}^1, \cdot) \;&\geq\; t(\tilde{c}^1, \emptyset, \cdot) + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} [\Delta c_k + (1-\alpha) S_k^R] \\ &=\; \tilde{c}^1 + \sum_{\{k|f_k \in \hat{F}^2\}} \Delta c_k + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} [\Delta c_k + (1-\alpha) S_k^R] \\ &=\; \tilde{c}^1 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \;. \end{aligned} \tag{A.16}$$

This inequality must hold for all $\tilde{c}^1$ arbitrarily close to c^2. Thus, a necessary condition for ex post incentive compatibility to hold for a type $(c^1, \hat{F}^1)$ with $c^1 < c^2$ is:

$$t(c^1, \hat{F}^1, \cdot) \geq c^2 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \quad \forall c^1 < c^2. \tag{A.17}$$

Now suppose that $\tilde{c}^1 < c^2$ and $\tilde{F}^1 = \emptyset$. Again, $t(\tilde{c}^1, \emptyset, \cdot)$ satisfies (LL) with equality if

$$t(\tilde{c}^1, \emptyset, \cdot) = \tilde{c}^1 + \sum_{\{k|f_k \in \hat{F}^2\}} \Delta c_k \;.$$

Using this in inequality (A.15) yields

$$\begin{aligned} t(c^1, \hat{F}^1, \cdot) \;&\geq\; t(\tilde{c}^1, \emptyset, \cdot) - [c^1 + \sum_{\{k|f_k \in \hat{F}^2\}} \Delta c_k] + \sum_{\{k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \\ &= \tilde{c}^1 - c^1 + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \quad \forall c^1 > c^2, \tilde{c}^1 < c^2. \end{aligned} \tag{A.18}$$

Inequality (A.18) must hold for any $\tilde{c}^1$ arbitrarily close to c^2 which implies

$$t(c^1, \hat{F}^1, \cdot) \geq c^2 + \epsilon - c^1 + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \quad \forall c^1 > c^2,\ \epsilon \geq 0. \tag{A.19}$$

Furthermore, by (A.14) it has to hold that $t(c^1, \hat{F}^1, \cdot) = t(\tilde{c}^1, \hat{F}^1, \cdot)$ for all $c^1, \tilde{c}^1 > c^2$; i.e., if the seller does not execute production, his transfer is independent of the reported cost type. Hence, a necessary condition for ex post incentive compatibility is

$$t(c^1, \hat{F}^1, \cdot) \geq \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R \quad \forall c^1 > c^2. \tag{A.20}$$

The next result follows immediately from equations (A.17) and (A.20).

Lemma 1. *Consider a mechanism satisfying constraints (EPIC), (EP), (ED), and (LL). Then, the transfer schedule must satisfy*

$$t(c^1, \hat{F}^1, \cdot) \geq \begin{cases} \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R & \text{if } c^1 > c^2, \\ c^2 + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k + \sum_{\{k|f_k \in \hat{F}^1 \setminus \hat{F}^2\}} (1-\alpha) S_k^R & \text{if } c^1 < c^2. \end{cases} \tag{A.21}$$

If we extend (A.21) to $c^1 = c^2$, the ex post utility of seller 1 with cost type $c^1 = c^2$ is the same, irrespective of whether he has to produce the good and the transfer is given by the lower bound of the term for $c^1 < c^2$ or he does not obtain the contract and the transfer is given by the lower bound of the term for $c^1 > c^2$.

The mechanism of Proposition 3 satisfies (ED), (EP), (LL) and (EPIC), and it satisfies the condition provided in Lemma 1 with equality. Thus, this mechanism implements the efficient allocation at the lowest possible transfers, i.e. at the lowest possible cost to the buyer.

□

Proof of Proposition 5. The Arbitration Mechanism gives rise to the same monetary outcomes and the same incentives for each seller as the mechanism of Proposition 3. Thus, given that it is an equilibrium in the direct mechanism for each seller to reveal all observed design flaws early, it is also an equilibrium in the Arbitration Mechanism. Furthermore, given that all observed flaws have been revealed it is optimal for each seller to bid his true cost in the sealed-bid, second-price auction. Hence, with the Arbitration Mechanism the total payment that seller i receives is given by

$$T_1^i + T_2^i = (1-\alpha) S^R(\tilde{F}^i, D(\tilde{F}^j)) + \begin{cases} c_j + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k & \text{if } c^i < c^j \\ (1/2) \left(c_j + \sum_{\{k|f_k \in \hat{F}^1 \cup \hat{F}^2\}} \Delta c_k \right) & \text{if } c^i = c^j \\ 0 & \text{if } c^i > c^j \end{cases}, \tag{A.22}$$

where $T_1^i + T_2^i$ denotes the expected total payment. As the mechanism of Proposition 3, the Arbitration Mechanism is ex post incentive compatible. Furthermore, it does not require any ex ante knowledge knowledge of the parameters of the model. Finally, by Proposition 4 and the revelation principle there does not exist any other informationally robust mechanism that implements the efficient allocation at a lower cost to the buyer. □

Proof of Proposition 6. By the definition and the uniqueness of the welfare optimal investment levels, we can write

$$q_W^i = arg\max_{q^i} W(q^i, q_W^j) = arg\max_{q^i} \left\{p(q^i + q_W^j - q^i q_W^j)S - \phi^i(q^i) - \phi^j(q_W^j)\right\}, \tag{A.23}$$

for $i, j \in \{1, 2\}$ and $i \neq j$.

If seller i expects that seller j invests efficiently – i.e., $q^j = q_W^j$, then the expected profit of i is

$$\pi^i(q^i) = pq^i(1 - \hat{q}_W^j)S - \phi^i(q^i). \tag{A.24}$$

The above expression is maximized at the investment level given in equation (A.23). Hence, investing efficiently is a mutually best response, which completes the proof.

□

References

Arrow, K. J. (1962): "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Incentive Activity: Economic and Social Factors*, ed. by Richard R. Nelson, pp. 609–626. Princeton Univerisity Press, Princeton.

Asker, J., and E. Cantillon (2010): "Procurement when Price and Quality Matter," *The RAND Journal of Economics*, 41(1), 1–34.

Bajari, P., S. Houghton, and S. Tadelis (2014): "Bidding for Incomplete Contracts: An Empirical Analysis of Adaptation Costs," *American Economic Review*, 104(4), 1288–1319.

Bajari, P., and S. Tadelis (2001): "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *RAND Journal of Economics*, 32(3), 387–407.

——— (2011): "Incentives and Award Procedures: Competitive Tendering vs. Negotiations in Procurement," in *Handbook of Procurement*, ed. by N. Dimitri, G. Piga, and G. Spagnolo, pp. 121–142. Cambridge University Press, Cambridge.

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1007–1035.

——— (2012): "An Introduction to Robust Mechanism Design," *Foundations and Trends in Microeconomics*, 8(3), 169–230.

Bergemann, D., and J. Välimäki (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70(3), 1007–1033.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986): "The Nash Bargaining Solution in Economic Modelling," *RAND Journal of Economics*, 17(2), 176–188.

Bull, J., and J. Watson (2007): "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, 58(1), 75–93.

Chakravarty, S., and W. B. MacLeod (2009): "Contracting in the Shadow of the Law," *The RAND Journal of Economics*, 40(3), 533–557.

CHE, Y.-K. (1993): "Design Competition Through Multidimensional Auctions," *The RAND Journal of Economics*, 24(4), 668–680.

CHE, Y.-K., E. IOSSA, AND P. REY (2016): "Prizes versus Contracts as Incentives for Innovation," CEIS Research Paper, 358.

CRÉMER, J., AND R. P. MCLEAN (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247–1257.

CROCKER, K. J., AND K. J. REYNOLDS (1993): "The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement," *RAND Journal of Economics*, 24(1), 126–146.

GREEN, J. R., AND J.-J. LAFFONT (1986): "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53(3), 447–456.

HART, O. (1995): *Firms, Contracts, and Financial Structure*, Clarendon Lectures in Economics. Clarendon Press, Oxford, paperback ed., reprint. edn.

HERWEG, F., AND K. M. SCHMIDT (2015): "Loss Aversion and Inefficient Renegotiation," *Review of Economic Studies*, 82, 297–332.

——— (2017): "Auctions versus Negotiations: The Effects of Inefficient Renegotiation," *The RAND Journal of Economics*, 48(3), 647–672.

LAFFONT, J.-J., AND J. TIROLE (1993): *A Theory of Incentives in Procurement and Regulation.* MIT Press, Cambridge, Massachusetts.

MCAFEE, R. P., AND J. MCMILLAN (1986): "Bidding for Contracts: A Principal-Agent Analysis," *The RAND Journal of Economics*, 17(3), 326–338.

TIROLE, J. (2009): "Cognition and Incomplete Contracts," *The American Economic Review*, 99(1), 265–294.

WILLIAMSON, O. E. (1985): *The Economic Institutions of Capitalism.* Free Press, New York.

Wilson, R. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. F. Bewley, chap. 2, pp. 33–70. Cambridge Univ. Press, Cambridge.