MECHANISM DESIGN WITH RISK AVERSION: A THEORY OF THE FIRST-BEST

ALEX TORDJMAN

ABSTRACT. When agents are risk-averse, the allocation of both social surplus and risk matters to them. A mechanism is *first-best* if it lies on the Pareto-frontier of the set of interim certainty equivalents for the agents and the designer's net expected payoff, the latter acting as a monetary sink. This criterion is shown to be equivalent to a mechanism being *full insurance* for the agents and having an *ex-post efficient* outcome rule. Both *ex-post* budget balance and dominant-strategy incentive compatibility are generically incompatible with a mechanism being first-best. An (interim) incentive-compatible first-best mechanism is constructed, for a general social problem in which types can be multidimensional and agents can exhibit arbitrary (weakly) risk-averse preferences. However, the mechanism may fail to be (interim) individually rational when agents have non-trivial property rights. For binary-allocation problems with property rights and unidimensional types, in which agents' preferences for binary lotteries are *concave biseparable*, a necessary and sufficient condition for the existence of an incentive-compatible, individually rational, *ex-ante* budget-balanced first-best mechanism is given. The results shed light on the importance of first-order (as opposed to second-order) risk aversion for overcoming participation constraints in the classical bilateral trade problem.

1. INTRODUCTION

Risk aversion is a defining feature of individual decision-making under uncertainty. When agents participate in economic mechanisms, such as auctions, they face risky prospects, because they do not know the other agents' private information. A vast theoretical literature, initiated by Vickrey (1961), has analyzed how to design first-best mechanisms in which *risk-neutral* agents are induced to reveal their private information truthfully. Yet, even though there is ample evidence that real-world agents participating in mechanisms exhibit risk-averse preferences, the question of

Date: August 11, 2026.
Tordjman: Department of Economics, Stanford University. Email: tordjman@stanford.edu.

The author thanks Roberto Corrao, Laura Doval, Paul Milgrom, Benny Moldovanu, Ilya Segal, and Mengxi Zhang for insightful comments on this project and especially Ravi Jagadeesan for his invaluable guidance. This research was partly carried out while the author was visiting the Hausdorff Research Institute for Mathematics at the University of Bonn, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2047/2 – 390685813.

how to design incentive-compatible, first-best economic mechanisms in the presence of risk aversion remains largely open.[1]

This paper defines and studies first-best mechanisms when agents are *risk-averse*. Because agents dislike being exposed to risk, implementing the first-best requires not only choosing an *ex-post* efficient outcome rule, but also deciding how to allocate risk between the agents.

We consider the problem of a risk-neutral designer who wants to design a *first-best* mechanism in the presence of risk-averse agents with arbitrary risk preferences, in a general social choice environment with private values, costs and possibly multidimensional types, in which agents evaluate outcomes *ex-post* quasilinearly in transfers.[2] In order not to take a stance as to how agents reduce compound lotteries and whether they are dynamically consistent, we refrain from framing our notion of *first-best* mechanisms at the *ex-ante* stage.[3] We instead propose a criterion at the *interim* stage. Any mechanism induces a certainty equivalent at the *interim* stage for each type of each agent, as well as a payoff to the designer which equals the sum of expected transfers net of the cost of implementing the social outcome. A mechanism *CE-Pareto dominates* another mechanism if it gives each type of each agent a weakly higher certainty equivalent under truthful reporting as well as weakly higher expected payoff to the designer, with one of these inequalities being strict (for agents, the strict inequality has to hold for a positive-measure set of the type space). Our desideratum for a mechanism to be *first-best* is that it is not CE-Pareto dominated by another mechanism.

We develop our analysis of *first-best* mechanisms into three parts. First, we give an equivalent characterization of our *interim* criterion in terms of both an *ex-ante* criterion— the mechanism has to maximize the sum of the expected certainty equivalents of all agents and the designer's expected payoff — and an *ex-post* criterion — the mechanism has to be *ex-post* efficient and provide *full insurance* to agents. Hence, the risk-neutral designer acts not only as the sink of all payments but also as the sink of all risk. Second, we construct an incentive-compatible, first-best mechanism, which is also *ex-ante* budget-balanced. This mechanism is valid for arbitrary risk

[1]In their analysis of bidding behavior in US Forest Service timber auctions, Athey and Levin (2001) find evidence of bidder risk aversion. The reader is referred to Vasserman and Watt (2021) for a review of the relevant empirical and theoretical literature on risk aversion in auctions.

[2]This allows us to distinguish risk aversion from income effects. Agents' private information affects both the way the evaluate *ex-post* lotteries, and parameterizes their risk preferences.

[3]Indeed, this matters for risk-averse agents whose preferences cannot be represented by expected-utility.

preferences and types. Third, observing that the previous mechanism is not *interim* individually rational in the presence of property rights, we specify the environment to binary-allocation problems with unidimensional private information, in which agents exhibit *concave biseparable* preferences. We provide a necessary and sufficient condition for the existence of an incentive-compatible, individually rational, first-best, *ex-ante* budget-balanced mechanism. In bilateral trade, the impossibility result of Myerson and Satterthwaite (1983) still holds if one of the two agents has preferences exhibiting second-order risk aversion — this is in particular the case for risk-averse (smooth) expected utility beyond risk-neutrality. First-order risk aversion can help reverse the impossibility result if both agents are sufficiently risk-averse.

For the first part, we observe that, having defined the *ex-ante* welfare of a mechanism as the sum of the designer's expected payoff and the sum of the expected certainty equivalents of all agents, the difference between the *ex-ante* efficient social welfare and the welfare of an arbitrary mechanism can be decomposed into two separate losses — one due to allocative inefficiency, and one due to agents' risk-premia. This decomposition is the key to showing that the *ex-ante* welfare maximization is equivalent to both our *interim* definition of *first-best* mechanisms, and to two *ex-post* properties. The first property is that the mechanism be *ex-post* efficient, just as in the risk-neutral case. The second property has to do with agents' risk preferences. When agents are *weakly risk-averse* —i.e., when their preferences are such that their certainty equivalent for a lottery with bounded support is no greater than its expectation — this second property is called the *zero risk-premium* property, meaning that all agents' certainty equivalents of the random variable induced by the mechanism (by the randomization device and by the other agents' reports) should be equal to their mean. When agents are *strictly risk-averse*, in the sense that equality between a lottery's certainty equivalent and its expectation can only occur for degenerate lotteries, this second property is that of *full insurance*, meaning the *ex-post* utility of an agent who reports truthfully in the mechanism only depends on his type almost surely.[4]

Having defined our notion of the *first-best*, we turn to whether it is possible to design first-best mechanisms in which agents are induced to reveal their private information truthfully — that is, mechanisms that satisfy some form of incentive-compatibility.

[4]Our characterization relies on the fact that the designer is risk neutral, and faces no limited-liability constraint. Under these conditions, the designer can absorb all aggregate risk.

One may wonder whether some of the classical efficient, incentive-compatible mechanisms proposed in the literature work in this setting. The answer is negative. We first turn to investigate dominant-strategy incentive-compatible mechanisms. The mechanism of Vickrey (1961), Clarke (1971) and Groves (1973) fails to be full insurance. In fact, we show a more general impossibility result. When mechanisms are deterministic and the environment satisfies the assumptions enabling to use (Milgrom and Segal, 2002, Theorem 1), a *first-best*, dominant-strategy incentive-compatible mechanism exists if and only if the efficient surplus is additively separable in types (e.g., when it is always efficient to give the good to the same fixed agent, irrespective of types). This result is a consequence of our characterization of *first-best* mechanisms together with (Milgrom and Segal, 2002, Theorem 1). Second, we show that the mechanism of Arrow (1979) and d'Aspremont and Gérard-Varet (1979) also fails to be *first-best* in this setting, because it exposes agents to risk. Actually, a similar incompatibility exists between *ex-post* budget balance, which is a key property of the d'Aspremont and Gérard-Varet (1979) mechanism, and our criterion for being *first-best*. Again, except in the knife-edge case where the efficient surplus function is additively separable in types almost everywhere, there does not exist a *first-best*, *ex-post* budget-balanced mechanism.

These two impossibility results imply that, for general environments, the best one can hope for in terms of budget and incentives, is to construct a (Bayesian) incentive-compatible, *ex-ante* budget-balanced mechanism. We construct such a mechanism. The mechanism is valid for arbitrary weakly risk-averse preferences (e.g., represented by expected or non-expected utility) and multidimensional type spaces. This deterministic mechanism implements an *ex-post* efficient allocation rule and provides *full insurance* to every agent — hence, it is *first-best* — by making the *ex-post* utility of any agent be the difference between the *interim* efficient social surplus induced by an agent's report and the *ex-ante* efficient social surplus. This construction makes every agent fully internalize the impact of their report on expected social surplus.[5] We complete the mechanism by lump-sum side transfers which make the mechanism *ex-ante* budget-balanced. Proving that the mechanism is incentive-compatible relies on arguments similar to the ones of d'Aspremont and Gérard-Varet (1979) or Safronov (2018), in the sense that our mechanism has an expected-externality structure. However, an additional dimension, coming from the fact that an untruthful report exposes

[5]The payment rule construction is similar to that in Dughmi and Peres (2012), which studied a related problem for the case of expected utility.

an agent to risk, helps complete the argument. Our mechanism exploits both the fact that a misreport changes the outcome to one that is efficient under the report and exposes the deviator to a payoff lottery. Just like the mechanism of d'Aspremont and Gérard-Varet (1979), our mechanism can fail to be *interim* individually rational when agents have property rights (e.g., in bilateral trade).

In the last part of the paper, we therefore turn to the issue of participation in *first-best* mechanisms. Specifically, we investigate the possibility of designing *first-best*, incentive-compatible, *ex-ante* budget-balanced mechanisms which respect an *interim* individual rationality constraint. We restrict the setting to unidimensional private information, binary-allocation problems, and restrict agents' preferences on binary lotteries to be *concave biseparable* (Ghirardato et al. (2004)).[6] That is, for binary lotteries, agents are allowed to distort both probability and outcomes, the latter in a concave continuously differentiable way. Our restriction encompasses as special cases von Neumann and Morgenstern's (1944) expected utility with a smooth Bernoulli utility, Safra and Segal's (1998) constant risk-averse preferences, Quiggin's (1993) rank-dependent utility with concave utility over money, and Yaari's (1987) dual expected utility. Importantly, the class allows agents to exhibit either first or second order risk aversion, as defined by Segal and Spivak (1990). We characterize the set of first-best, incentive compatible mechanisms, thus generalizing (Gershkov et al., 2022, Proposition 3) with an arbitrary concavity in the evaluation of money on binary lotteries.[7] When agents exhibit second-order risk aversion, the usual payoff equivalence formula holds. We then give a general necessary and sufficient condition for the existence of a *first-best*, incentive-compatible, individually rational, and *ex-ante* budget-balanced mechanism, where we allow a type-dependent participation constraint in order to capture some forms of property rights.[8] In the bilateral trade problem, the impossibility result of Myerson and Satterthwaite (1983) still holds even if only one of the agents exhibits second-order risk-averse preferences. However, it can be overturned when both agents are sufficiently risk-averse and exhibit first-order risk aversion. We illustrate the latter through examples.

[6]We impose no restrictions as to how agents may evaluate more complex lotteries.

[7]Gershkov et al. (2022) did not consider the possibility of having a concave transformation over monetary payoffs, and their proof relies on the cash invariance property of Safra and Segal's (1998) constant risk-averse preferences. Our result does not impose this restriction and admits risk-averse expected-utility as a special case.

[8]Hence, the model captures public good provision, partnership dissolution, bilateral trade, and auctions as special cases.

Related Literature. Implementation of the first-best when agents are risk-neutral and information is dispersed in society is an old problem in economic theory, dating back to at least Vickrey (1961), Clarke (1971) and Groves (1973) who proposed a dominant-strategy incentive compatible, efficient mechanism. Following this work, Arrow (1979) and d'Aspremont and Gérard-Varet (1979) proposed an incentive-compatible, *ex-post* budget-balanced and efficient mechanism. The literature has then investigated whether such mechanisms could be accommodated to satisfy participation constraints. The literature features a number of (im)possibility results — in a bilateral trade setting, Myerson and Satterthwaite (1983) have shown that no individually rational, incentive-compatible, *ex-post* budget-balanced and efficient mechanism exists, while the result is flipped in the partnership dissolution model of Cramton et al. (1987) when partners initially hold sufficiently symmetric shares.[9] In the public good provision problem, Mailath and Postlewaite (1990) have shown that the probability of provision goes to zero as the number of agents grows large. All these results feature expected-utility maximizing agents who are risk-neutral.

Little is known about whether the first-best is implementable when agents are risk-averse. For the case of expected-utility, Chatterjee and Samuelson (1983) show that a double auction can implement the efficient allocation in the limit when agents become infinitely risk-averse. Garratt and Pycia (2026), in the bilateral trade model of Myerson and Satterthwaite (1983), characterize when the efficient allocation can be implemented when agents are risk-averse. These papers generally take a different view on welfare, in the sense that even if agents are risk-averse, they study the implementation of efficient social surplus by a risk-neutral designer, without providing insurance to the agents.[10] On the contrary, we propose a criterion for first-best mechanisms that ensures mechanisms are both risk-efficient and surplus-efficient.[11] The conclusion that a risk-neutral party absorbs aggregate risk is standard in the theory

[9]The impossibility result of Myerson and Satterthwaite (1983) also holds if one only required *ex-ante* budget balance.

[10]In particular, when Garratt and Pycia (2026) are able to overturn the impossibility result of Myerson and Satterthwaite (1983), their mechanism exploits the risk aversion of agents by exposing them to risk.

[11]Holmström and Myerson (1983) distinguish efficiency concepts according to the informational stage at which welfare is evaluated — *ex ante*, *interim*, or *ex post* — and according to whether incentive constraints restrict the set of admissible comparisons. Our criterion is closest to their classical interim efficiency notion. Every interim type is treated as a distinct Pareto claimant, while a dominating mechanism need not itself be incentive-compatible. We augment their framework by including the risk-neutral designer as an additional Pareto claimant and by allowing types to evaluate payoff lotteries through arbitrary risk-preferences, in particular non-expected-utility certainty equivalents.

of Pareto-optimal risk sharing and appears, in different forms, in Borch (1962), Wilson (1968), Arrow and Lind (1972) as well as in the modern literature on monetary risk measures (see, for instance, Jouini et al. (2008)). We show that this risk-sharing principle follows endogenously from an interim certainty equivalent Pareto criterion in a mechanism-design environment, and study its compatibility with incentive compatibility, participation, and budget balance.

Dughmi and Peres (2012) show how to fully insure a given risk-neutral allocation rule without losing incentive compatibility. The present paper asks which allocation and risk-sharing rule is first-best in the first place, and when that first best is compatible with incentives, participation, and budget balance. We also allow for arbitrary risk preferences, not just expected utility — to the best of our knowledge, no study has investigated the implementation of the first-best with static mechanisms when agents are not expected utility maximizers.[12,13]

A recent literature has investigated static mechanism design problems with non-expected utility maximizers —e.g., optimal auctions (Gershkov et al. (2022)), insurance problems (Gershkov et al. (2023), Di Giovan Paolo and Higueras Corona (2026)) and security design (Gershkov et al. (2025)). Contrary to these papers, the present study does not aim to solve the maximization problem of a seller facing risk-averse agents, but rather asks when the first-best can be implemented. Ahn et al. (2026) and Rivera Mora and Strack (2026) show that, when agents have non-expected utility, the designer may be able to implement any allocation using *dynamic* mechanisms. In particular, full surplus extraction is feasible, so that the efficient social welfare is implementable without leaving rents to the agents — Ahn et al. (2026) show that this is feasible while providing full insurance on-path to the agents. Contrary to these papers, the focus of the present paper is on implementation of the first-best with *static* mechanisms.

Paper Outline. The rest of the paper is organized as follows. Section 2 is the model. Section 3 develops the first-best criterion and characterizes it. Section 4 shows its

[12]Indeed, while the transfer rule of the incentive-compatible, first-best mechanism we construct coincides with that in (Dughmi and Peres, 2012, Theorem 4.1) when agents have expected utility with an increasing, concave Bernoulli utility, our mechanism works for more general preferences. Our proof does not rely on the use of Jensen's inequality. The mechanism we construct is also *ex-ante* budget-balanced. Moreover, we allow the private information of agents to not only parameterize how they evaluate *ex-post* lotteries, but also their risk preferences.

[13]More distant work studies efficient implementation in bilateral trade when agents are ambiguity averse, as in Wolitzky (2016), or loss-averse, as in Benkert (2025), who explores when the first-best is implementable in the model of Kőszegi and Rabin (2007).

incompatibility with *ex-post* budget balance and dominant-strategy incentive compatibility. Section 5 constructs an incentive-compatible, *ex-ante* budget-balanced, first-best mechanism. Section 6 studies first-best implementation with participation constraints in binary allocation problems. Section 7 is a conclusion.

2. Model

2.1. **Economic environment.** Consider a social choice environment with private values. There is a finite set I of agents, and the designer is denoted by d. Each agent i has a (compact metric) space of outcomes Ω_i. The set of feasible social outcomes is a nonempty closed subset $\Omega \subseteq \times_{i\in I}\Omega_i$. A generic social outcome is denoted by $x = (x_i)_{i\in I}$. There is a continuous social cost function $c : \Omega \to \mathbb{R}$.

Agent i's type θ_i lies in a (compact metric) space Θ_i, and is distributed according to a (Borel) probability measure F_i. Types are (mutually) independent, and the space of type profiles is $\Theta = \times_{i\in I}\Theta_i$, so that $F = \bigotimes_{i\in I} F_i$ denotes the resulting probability measure on Θ. For each agent i, we use the standard notation $\Theta_{-i} = \times_{j\neq i}\Theta_j$ and $F_{-i} = \otimes_{j\neq i}F_j$. Each agent $i \in I$ has a continuous valuation function $v_i : \Omega_i \times \Theta_i \to \mathbb{R}$, mapping outcome-type pairs to the real line, so that the environment has *private values*. When an outcome $x \in \Omega$ is implemented and agent i pays a transfer $t_i \in \mathbb{R}$ to the designer, the *ex-post* monetary payoff of agent i is $v_i(x_i;\theta_i) - t_i$. The designer, acting as a monetary sink, has payoff $\sum_{i\in I} t_i - c(x)$.

We denote by $S : \Omega \times \Theta \to \mathbb{R}$ defined by $S(x,\theta) = \sum_{i\in I} v_i(x_i;\theta_i) - c(x)$ the social surplus function. Define $S^*(\theta) = \max_{x\in\Omega} S(x,\theta)$ to be the efficient surplus function and $\Omega^*(\theta) = \arg\max_{x\in\Omega} S(x,\theta)$ to be the efficient outcome correspondence.[14] A measurable selection of an efficient outcome rule exists. (The formal argument is in the appendix.)

Lemma 1. *The function $S^* : \Theta \to \mathbb{R}$ is continuous. The correspondence $\Omega^* : \Theta \rightrightarrows \Omega$ is nonempty-valued, compact-valued and upper hemicontinuous. In particular, there exists a (Borel) measurable selection $x^* : \Theta \to \Omega$ such that $x^*(\theta) \in \Omega^*(\theta)$ for all $\theta \in \Theta$.*

2.2. **Mechanisms.** All mechanisms considered in the paper are static and direct. Agents submit their reports before the mechanism's randomization is realized. The *randomization device* is the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ denotes

[14] These objects are well-defined because Ω is compact as a closed subset of a compact space, and $S(.,\theta)$ is continuous in $x \in \Omega$ for any fixed $\theta \in \Theta$.

the Lebesgue measure. Let $r \sim U([0,1])$ be the random seed, which is drawn independently of all types.[15] The joint probability space for truthful play is therefore $(\Theta \times [0,1], \mathcal{B}(\Theta) \otimes \mathcal{B}([0,1]), \mathbb{P})$ where $\mathbb{P} = F \otimes \lambda$. A *mechanism* is a pair $M = (x,t)$ consisting of a (Borel) measurable *outcome rule* $x : \Theta \times [0,1] \to \Omega$ and a (bounded, Borel) measurable *transfer rule* $t = (t_i)_{i \in I} : \Theta \times [0,1] \to \mathbb{R}^I$.[16]

2.3. **Interim payoff lotteries.** Given a mechanism $M = (x,t)$, we define the random payoff of agent i of type θ_i from reporting $\hat{\theta}_i$ as the (bounded) random variable $Y_i(\theta_i, \hat{\theta}_i) : \Theta_{-i} \times [0,1] \to \mathbb{R}$ given by

$$Y_i(\theta_i, \hat{\theta}_i)(\theta_{-i}, r) = v_i(x_i(\hat{\theta}_i, \theta_{-i}, r); \theta_i) - t_i(\hat{\theta}_i, \theta_{-i}, r).$$

Under truthful reporting, we write $Y_i^M(\theta_i) \equiv Y_i^M(\theta_i, \theta_i)$ so that

$$Y_i^M(\theta_i)(\theta_{-i}, r) = v_i(x_i(\theta, r); \theta_i) - t_i(\theta, r).[17]$$

Let $m_i^M(\theta_i, \hat{\theta}_i) = \mathbb{E}_{\tilde{\theta}_{-i}, \tilde{r}}[Y_i^M(\theta_i, \hat{\theta}_i)(\tilde{\theta}_{-i}, \tilde{r})]$ be the conditional mean payoff of agent i of type θ_i from reporting $\hat{\theta}_i$, and $m_i^M(\theta_i) \equiv m_i^M(\theta_i, \theta_i)$ be the truthful conditional mean.

2.4. **Risk preferences.** Let $\mathcal{P}(\mathbb{R})$ be the set of Borel probability measures on $\mathbb{R}$, endowed with the weak-* topology and its (Borel) σ-algebra. Let $\mathcal{P}_b(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$ denote the subset of laws with bounded support, endowed with the trace σ-algebra. Agent i's preferences over monetary payoff lotteries are represented by a (Borel) measurable *certainty equivalent* functional $\mathrm{CE}_i : \mathcal{P}_b(\mathbb{R}) \times \Theta_i \to \mathbb{R}$ such that $\mathrm{CE}_i(\delta_z, \theta_i) = z$ for any $z \in \mathbb{R}$ and $\theta_i \in \Theta_i$, where δ_z is the degenerate distribution concentrated on z.[18] For a random variable Z, we write $\mathrm{CE}_{i,\theta_i}(Z)$ as a shorthand for the certainty equivalent of agent i of type θ_i for the law of Z. Throughout, we will make use of the following monotonicity assumption.[19]

Assumption 1 (Monotonicity)**.** If μ first-order stochastically dominates ν, then $\mathrm{CE}_{i,\theta_i}(\mu) \geq \mathrm{CE}_{i,\theta_i}(\nu)$.

[15]The representation through a single uniform random variable entails no loss relative to defining a randomized mechanism as a (Borel) probability kernel over $\Omega \times \mathbb{R}^I$.

[16]We require that transfers be bounded. Because valuations and costs are bounded on their compact domains, all payoff random variables induced by mechanisms are bounded.

[17]The interpretation is as follows— at the *interim* stage, agents face a lottery on the probability space $(\Theta_{-i} \times [0,1], \mathcal{B}(\Theta_{-i}) \otimes \mathcal{B}([0,1]), F_{-i} \otimes \lambda)$ and the resolution of all uncertainty (in other agents types and in the randomization device) takes place in a single stage.

[18]If $Y : T \times \Omega_0 \to \mathbb{R}$ is bounded and jointly (Borel) measurable, and $(\Omega_0, \mathcal{A}_0, P_0)$ is a probability space, then $t \mapsto \mathrm{Law}_{P_0}(Y(t,.))$ is (Borel) measurable from T into $\mathcal{P}(\mathbb{R})$. Consequently, all the certainty-equivalent schedules considered in the paper are (Borel) measurable.

[19]The mechanism constructed in Section 5 does not require agents to respect first-order stochastic dominance.

We will work with two rather permissive definitions of risk aversion.

Definition 1. Agent i of type θ_i is *weakly risk-averse* if $CE_{i,\theta_i}(\mu) \leq \int_{\mathbb{R}} \tilde{z}\, d\mu(\tilde{z})$ for every $\mu \in \mathcal{P}(\mathbb{R})$. Agent i is *weakly risk-averse* if all his types are weakly risk-averse.

Agent i of type θ_i is *strictly risk-averse* if he is weakly risk-averse and for all $\mu \in \mathcal{P}(\mathbb{R})$, if $\mathrm{CE}_{i,\theta_i}(\mu) = \int_{\mathbb{R}} \tilde{z}\, d\mu(\tilde{z})$ then there exists $z \in \mathbb{R}$ such that $\mu = \delta_z$. Agent i is *strictly risk-averse* is all his types are strictly risk-averse.

Hence, an agent is *weakly risk-averse* if he evaluates the certainty equivalent of a random variable to be no greater than its expectation, while for a *strictly risk-averse* agent, equality can hold only for degenerate random variables.[20]

Denote the interim-certainty equivalents induced by a mechanism $M = (x,t)$ by $V_i^M(\theta_i, \hat{\theta}_i) \equiv \mathrm{CE}_{i,\theta_i}\left(Y_i^M(\theta_i, \hat{\theta}_i)\right)$, and under truthful reporting $V_i^M(\theta_i) = V_i^M(\theta_i, \theta_i)$. These functions are (Borel) measurable.

Remark 1. The model features agents who have quasilinear preferences at the *ex-post* stage, expressed in monetary terms. Hence, there are no income effects at the *ex-post* stage. However, agents are risk-averse in that they can apply a non-linear transformation to their *ex-post* random monetary payoffs at the *interim* stage. This enables to disentangle risk aversion from income effects, as in Greenwood et al. (1988) and the subsequent macroeconomic literature. Private information for the agents enters both in their evaluation of *ex-post* payoffs, and in their risk-preferences. For the main results in this paper, the designer doesn't have to know the exact risk-preferences of agents.[21]

For the main results in the paper, we will be concerned with Bayesian (interim) incentive-compatibility.

Definition 2. A mechanism $M = (x,t)$ is *incentive-compatible (IC)* if for all agents $i \in I$, for all types $\theta_i \in \Theta_i$ and all reports $\hat{\theta}_i \in \Theta_i$, $V_i^M(\theta_i) \geq V_i^M(\theta_i, \hat{\theta}_i)$.

[20]When agents' preferences can be described by expected-utility, this notion of weak risk aversion is equivalent to the usual requirement that an agent prefers a lottery over any mean preserving spread of that lottery. However, these two notions do not coincide beyond the expected-utility case.

[21]The mapping from types into risk preferences is common knowledge, but the realized type, and therefore the realized risk preference, is private. This is conceptually different from the designer not even knowing the mapping from types into certainty equivalents. In fact, the mechanism we construct in Theorem 2 is robust enough to handle this second possibility as well, provided the designer knows that every possible type is weakly risk-averse.

The designer is risk-neutral. Her utility from a mechanism $M = (x, t)$ is

$$U_d^M = \mathbb{E}_{\tilde{\theta},\tilde{r}}\left[\sum_{i\in I} t_i(\tilde{\theta},\tilde{r}) - c(x(\tilde{\theta},\tilde{r}))\right]$$

We end this section with an accounting identity which holds for all mechanisms and will prove helpful in our subsequent analyses.

Lemma 2. *For any mechanism $M = (x, t)$, the following accounting identity holds*

$$U_d^M + \sum_{i\in I} \mathbb{E}_{\tilde{\theta}_i}[m_i^M(\tilde{\theta}_i)] = \mathbb{E}_{\tilde{\theta},\tilde{r}}[S(x(\tilde{\theta},\tilde{r}),\tilde{\theta})] \tag{1}$$

Proof. For a mechanism $M = (x, t)$, the identity follows from

$$\begin{aligned} U_d^M + \sum_{i\in I} \mathbb{E}_{\tilde{\theta}_i}[m_i^M(\tilde{\theta}_i)] &= \mathbb{E}_{\tilde{\theta},\tilde{r}}\left[\sum_{i\in I} t_i(\tilde{\theta},\tilde{r}) - c(x(\tilde{\theta},\tilde{r}))\right] + \mathbb{E}_{\tilde{\theta},\tilde{r}}\left[\sum_{i\in I} Y_i^M(\tilde{\theta}_i)(\tilde{\theta}_{-i},\tilde{r})\right] \\ &= \mathbb{E}_{\tilde{\theta},\tilde{r}}\left[\sum_{i\in I} v_i(x_i(\tilde{\theta},\tilde{r});\tilde{\theta}_i) - c(x(\tilde{\theta},\tilde{r}))\right] \\ &= \mathbb{E}_{\tilde{\theta},\tilde{r}}[S(x(\tilde{\theta},\tilde{r}),\tilde{\theta})], \end{aligned}$$

where the first equality is by Fubini's theorem and the definition of U_d^M, the second equality is by linearity of the integral and definition of $Y_i^M(\theta_i)$, and the last equality by definition of $S(x(\theta, r), \theta)$. □

3. First-Best Mechanisms with risk aversion

When agents are risk-averse, the allocation of both social surplus and risk between them matters. In this section, a criterion for a mechanism to be *first-best* in the presence of risk aversion is developed and characterized.

3.1. **Welfare criterion under risk aversion.**

Defining efficiency in the presence of risk aversion as simply implementing an *ex-post* efficient allocation rule seems paternalistic. Indeed, the agents also care about risk-sharing. Rather than imposing an *ad hoc*, *ex-post* criterion, our notion of *first-best* mechanisms is defined at the *interim* stage.

Definition 3. A mechanism M' *CE-Pareto dominates* M if $U_d^{M'} \geq U_d^M$ and for every agent $i \in I$, $V_i^{M'}(\theta_i) \geq V_i^M(\theta_i)$ for F_i-almost every θ_i, and either $U_d^{M'} > U_d^M$ or there exists an agent $i \in I$ for whom $F_i\left(\left\{\theta_i \in \Theta_i \,|\, V_i^{M'}(\theta_i) > V_i^M(\theta_i)\right\}\right) > 0$.[22]

[22]With nonatomic type distributions, literal pointwise Pareto efficiency is a vacuous concept— any zero-probability type can be given a strictly larger transfer at no expected cost to the designer. We therefore identify *interim* utility profiles modulo F_i-null sets. For finite full-support type spaces, this definition coincides with ordinary pointwise Pareto efficiency.

Definition 4. A mechanism M is *first-best (FB)* if no mechanism M' CE-Pareto dominates M.

Remark 2. One might wonder whether it is possible to give a reasonable definition of first-best mechanisms, for arbitrary risk-averse preferences, at the *ex-ante* stage, meaning prior to the realization of all uncertainty, including that in the realization of one's own type. Although Theorem 1 below shows that our *interim* criterion has natural *ex-post* and *ex-ante* equivalents, we argue this is not a well-defined exercise. First, it is unclear what the aggregation function of *interim* certainty equivalents into an *ex-ante* objective should be (e.g., when agents have heterogeneous Yaari (1987) preferences). Second, one would have to take a stance as to the exact *informational preferences* of agents, that is how they treat the reduction of compound lotteries and whether they are dynamically consistent.[23] As shown by Ahn et al. (2026), recursive-utility models and multi-self models of Yaari (1987) agents with the same risk preferences can lead to different predictions as to how agents treat compound lotteries. The criterion proposed in this paper is agnostic as to the exact informational preferences of agents, and the resolution of uncertainty between the *interim* and the *ex-post* stages is treated as a single-round lottery (see Footnote 17).

3.2. **Characterization of first-best mechanisms.** We characterize our *interim* criterion in terms of both the maximization of some *ex-ante* welfare function, and two *ex-post* properties. First, we formally define the two *ex-post* properties. The first property is efficiency of the *ex-post* outcome rule.

Definition 5. A mechanism $M = (x, t)$ is *ex-post efficient (EPEFF)* if $x(\theta, r) \in \arg\max_{x \in \Omega} S(x, \theta)$ for $\mathbb{P}$-almost every (θ, r).

Equivalently, a mechanism $M = (x, t)$ is EPEFF if $S(x(\theta, r), \theta) = S^*(\theta)$ for $\mathbb{P}$-almost every (θ, r). This is equivalent to requiring that, for F-almost every θ, $x(\theta, r) \in \Omega^*(\theta)$ for λ-almost every r. The second property is an insurance property, which takes a different form whether agents are assumed to be weakly or strictly risk-averse.

Definition 6. A mechanism $M = (x, t)$ satisfies the *zero risk-premium property (ZRP)* if for every agent $i \in I$, $V_i^M(\theta_i) = m_i^M(\theta_i)$ for F_i-almost every θ_i.

[23] Indeed, if one wants to construct a welfare objective at the *ex-ante* stage but then evaluates incentives and participation at the interim stage, it is as if the mechanism had two stages in the resolution of uncertainty — a first stage where one's own value is revealed to oneself, and a second stage in which the final outcome and transfers are realized as a function of the other agents' private information and the randomization device.

Definition 7. A mechanism $M = (x, t)$ is *full insurance (FI)* if for every agent $i \in I$, there exists a (bounded, Borel) measurable function $w_i^M : \Theta_i \to \mathbb{R}$ such that, for F_i-almost every θ_i, $Y_i^M(\theta_i)(\theta_{-i}, r) = w_i^M(\theta_i)$ for $F_{-i} \otimes \lambda$-almost every (θ_{-i}, r).

When a mechanism is FI, a truthful agent $i \in I$ bears neither the risk associated with the other agent's unknown types nor the risk generated by the mechanism's randomization device.

Second, for the *ex-ante* characterization, for any mechanism M, define $\mathcal{W}(M) = U_d^M + \sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^M(\tilde{\theta}_i)]$, which corresponds to the sum of the designer's monetary payoff and the agents' expected *interim* certainty equivalents.[24] The welfare loss of any mechanism, under this criterion, admits a natural decomposition — Lemma 3 gives a decomposition of the welfare loss of a mechanism in terms of the expected allocative loss and the aggregate risk premium loss borne by the agents.

Lemma 3. *Suppose all agents are weakly risk-averse, then for any mechanism M,*

$$\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] - \mathcal{W}(M) = \underbrace{\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] - \mathbb{E}_{\tilde{\theta}}[S(x(\tilde{\theta}, \tilde{r}), \tilde{\theta})]}_{\text{expected allocative loss}} + \underbrace{\sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[m_i^M(\tilde{\theta}_i) - V_i^M(\tilde{\theta}_i)]}_{\text{aggregate risk-premium loss}} \tag{2}$$

and the two terms on the right-hand side are non-negative.

Proof. The decomposition in (2) follows by definition of $\mathcal{W}(M)$ and the accounting identity in (1). The first-term in the right-hand side of (2) is non-negative by definition of $S^*(\theta)$, while the second term is non-negative because all agents are weakly risk-averse. □

We can now state the main result of this section, which characterizes *first-best* mechanisms as exactly the ones that maximize the above defined measure of *ex-ante* welfare, and are ZRP (or FI, when agents are strictly risk-averse) and EPEFF.

Theorem 1. *Suppose all agents are weakly risk-averse (resp. strictly risk-averse). Then, the following are equivalent:*

(i) M is FB,
(ii) M maximizes $\mathcal{W}$ over all mechanisms,
(iii) M is EPEFF and ZRP (resp. FI).

The formal proof is in the appendix. From the *ex-post* perspective, if a mechanism is FB and agents are strictly risk-averse, the (risk-neutral) designer fully insures all

[24] The *ex-ante* criterion is not what the agents care about at the *ex-ante* stage, see Remark 2.

agents against the risk in the mechanism — first-best mechanisms are both surplus-efficient and risk-efficient.[25] From the *ex-ante* perspective, all FB mechanisms lie on the hyperplane given by

$$\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] = U_d^M + \sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^M(\tilde{\theta}_i)]$$

but they may redistribute the available expected social surplus in any way between the designer and the agents. In particular, a mechanism in which the outcome rule is given by an efficient selector x^* of Lemma 1 and in which payments are such that all agents' *ex-post* utilities are constant equal to zero lie on this hyperplane and give utility $\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$ to the designer. However, this mechanism is not really interesting. In Section 5, an *ex-ante* budget-balanced, incentive-compatible, FB mechanism is constructed — hence, we show that within this hyperplane, we can find a mechanism that is incentive-compatible while leaving zero payoff to the designer. When agents are weakly risk-averse, we can restrict attention to FI mechanisms without loss when looking at attainable first-best expected utility profiles, even though not every first-best mechanism is itself FI. (The formal argument is in the appendix.)

Lemma 4. *Assume all agents are weakly risk-averse, then any FB mechanism M is CE-Pareto equivalent to a FB, FI mechanism.*[26]

Remark 3. When agents' preferences are described by expected-utility and agents are risk-neutral, the risk-premium term in (2) vanishes. Consequently, $\mathcal{W}(M) = \mathbb{E}_{\tilde{\theta},\tilde{r}}[S(x(\tilde{\theta},\tilde{r}),\tilde{\theta})]$. Hence, Theorem 1 recovers that a mechanism is FB if and only if it maximizes expected social surplus, which in turn is equivalent to the mechanism being EPEFF.

4. Impossibility results for first-best mechanisms

In this section, we provide characterizations of when first-best risk sharing is compatible with two canonical *ex-post* requirements. In smooth private-value environments, a deterministic dominant-strategy incentive-compatible first-best mechanism exists if and only if efficient surplus is additively separable across types. Even without

[25] As discussed in the Introduction, the idea that a risk-neutral party should provide full insurance to the risk-averse parties in risk-sharing problems is not new, and dates back to at least Wilson (1968). Theorem 1 shows that this desideratum emerges as equivalent to a natural interim welfare criterion for mechanism design when agents are risk-averse and the designer is risk-neutral.

[26] Formally, two mechanisms M and M' are CE-Pareto equivalent if $U_d^M = U_d^{M'}$ and for each agent $i \in I$, $V_i^M(\theta_i) = V_i^{M'}(\theta_i)$ for F_i-almost every θ_i.

imposing incentive compatibility, an *ex-post* budget-balanced first-best mechanism exists if and only if efficient surplus is essentially additively separable.

4.1. **Incompatibility of first-best mechanisms with dominant-strategy incentive compatibility.** One might wonder whether the Vickrey-Clarke-Groves (VCG) mechanism might be a FB mechanism in this setting. While it is still *dominant-strategy incentive-compatible*, it fails to be FI.[27]

Definition 8. A deterministic mechanism M is *dominant-strategy incentive compatible (DSIC)* if for all agents $i \in I$, for all type profiles $\theta \in \Theta$ and reports $\hat{\theta}_i \in \Theta_i$, $Y_i^M(\theta_i)(\theta_{-i}) \geq Y_i(\theta_i, \hat{\theta}_i)(\theta_{-i})$.

Actually, a more general impossibility result holds. Indeed, for deterministic mechanisms, when the environment is such that payoff equivalence holds, DSIC and FB are generically incompatible.

Assumption 2 (Envelope theorem)**.** For each agent $i \in I$, type spaces Θ_i are convex subsets of $\mathbb{R}^{d_i}$ with non-empty interior, F_i has a strictly positive density $f_i > 0$, $\theta_i \mapsto v_i(x_i; \theta_i)$ is continuously differentiable on an open neighborhood of Θ_i for all $x_i \in \Omega_i$, and $\partial_{\theta_i} v_i$ is continuous.

Definition 9. The surplus function S^* is *additively separable* if for each agent $i \in I$, there exists a continuous function $w_i : \Theta_i \to \mathbb{R}$ such that $S^*(\theta) = \sum_{i \in I} w_i(\theta_i)$ for all $\theta \in \Theta$.

Proposition 1. *Suppose all agents are strictly risk-averse and Assumption 2 holds, then there exists a deterministic mechanism M which is DSIC and FB if and only if S^* is additively separable.*

The formal argument is in the appendix. Proposition 1 should really be seen as an impossibility result. Indeed, only in knife-edge cases will the surplus function be additively separable (e.g., when it is always optimal to give a private good to the same agent). Thus, one has to resort to IC (and not DSIC) when investigating the truthful revelation of private information with FB mechanisms.[28]

[27] In this subsection only, we restrict attention to deterministic mechanisms. In deterministic mechanisms, the outcome and the transfer rules do not depend on r.

[28] In (Dughmi and Peres, 2012, Theorem 3.1), insurance removes only the mechanism's randomization conditional on an agent's opponents' reports, so their transformation is not FI in the sense of the present paper.

4.2. **Incompatibility of first-best mechanisms with ex-post budget balance.** The failure of the VCG mechanism, and more generally, of DSIC mechanisms to be FB in the presence of risk aversion, naturally raises the question of whether the Arrow-d'Aspremont–Gérard-Varet (AGV) mechanism, which relaxes DSIC to IC and obtains *ex-post* budget balance, is FB beyond the risk-neutral case.

Definition 10. A mechanism $M = (x, t)$ is *ex-post budget-balanced (EPBB)* if for all type profiles $\theta \in \Theta$, $\sum_{i \in I} t_i(\theta, r) = c(x(\theta, r))$ for λ-almost every r.[29]

It turns out that the answer is again negative, because this mechanism is not FI. Actually, a more general impossibility result between EPBB and FB holds in this setting (contrary to Proposition 1, this result does not specify the model beyond the general setup of Section 2).

Definition 9′. The surplus function S^* is *essentially additively separable* if for each agent $i \in I$, there exists a (bounded, Borel) measurable function $w_i : \Theta_i \to \mathbb{R}$ such that $S^*(\theta) = \sum_{i \in I} w_i(\theta_i)$ for F-almost every θ.[30]

Proposition 2. *Suppose all agents are strictly risk-averse, then there exists a mechanism M which is FB and EPBB if and only if S^* is essentially additively separable.*

The complete proof is in the appendix. With independent types and risk-neutral Bayesian agents, Börgers and Norman (2009) show an equivalence between *ex-ante* and *ex-post* budget balance by allocating the *ex-post* budget risk among agents. That argument cannot preserve the first-best with strictly risk-averse agents, because EPBB forces someone to absorb state-dependent surplus risk. Therefore, when one investigates the existence of IC, FB mechanisms under a budget balance constraint, that constraint can only be imposed at the *ex-ante* stage.

5. An Incentive-Compatible First-Best Mechanism

Motivated by the failure of VCG — or more generally, of any DSIC deterministic mechanism — and AGV — or more generally, of any EPBB mechanism — in being

[29] Alternatively, one could say that a mechanism is EPBB if $\sum_{i \in I} t_i(\theta, r) = c(x(\theta, r))$ for $\mathbb{P}$-almost every (θ, r). The results are not sensitive to this change in the definition, but the more traditional definition which requires budget balance to hold pointwise in type reports is preferred.

[30] Since S^* is bounded, this definition is equivalent to the same definition with arbitrary real-valued (Borel) components w_i. Indeed, if $S^*(\theta) = \sum_{i \in I} w_i(\theta_i)$ for F-almost every θ, by Fubini's theorem, one can fix $\bar{\theta}_{-i}$ such that $w_i(\theta_i) = S^*(\theta_i, \bar{\theta}_{-i}) - \sum_{j \neq i} w_j(\bar{\theta}_j)$, so that w_i is essentially bounded and admits a bounded (Borel) modification.

FB in arbitrary environments, an (interim) incentive-compatible, *ex-ante* budget-balanced, first-best mechanism is constructed. It turns out that this mechanism may fail (interim) individual rationality.

Definition 11. A mechanism $M = (x, t)$ is *ex-ante budget-balanced (EABB)* if $U_d^M = 0$.

5.1. **Construction of the mechanism.** We henceforth study the following deterministic mechanism, where for ease of notation, we omit the randomization device. The (deterministic) outcome rule is given by the measurable selection $x^*(\hat{\theta})$ of Lemma 1, for a vector of reports $\hat{\theta}$. Define a (deterministic) payment rule for each agent i by

$$\begin{aligned} t_i^*(\hat{\theta}) &= v_i(x_i^*(\hat{\theta}); \hat{\theta}_i) - \mathbb{E}_{\tilde{\theta}_{-i}}\left[v_i(x_i^*(\hat{\theta}_i, \tilde{\theta}_{-i}); \hat{\theta}_i) + \sum_{j \neq i} v_j(x_j^*(\hat{\theta}_i, \tilde{\theta}_{-i}); \tilde{\theta}_j) - c(x^*(\hat{\theta}_i, \tilde{\theta}_{-i})) \right] \\ &\quad + \mathbb{E}_{\tilde{\theta}}\left[\sum_{j \in I} v_j(x^*(\tilde{\theta}); \tilde{\theta}_j) - c(x^*(\tilde{\theta})) \right] - \alpha_i \\ &= \underbrace{v_i(x_i^*(\hat{\theta}); \hat{\theta}_i)}_{\textit{ex-post}\text{ valuation}} - \underbrace{\left[\mathbb{E}_{\tilde{\theta}_{-i}}\left[S^*(\hat{\theta}_i, \tilde{\theta}_{-i}) \right] - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta}) \right] \right]}_{\text{expected externality from reporting } \hat{\theta}_i} - \alpha_i, \end{aligned}$$

where $\alpha_i \in \mathbb{R}$ are constants such that $\sum_{i \in I} \alpha_i = \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta}) \right]$.

Theorem 2. *When all agents are weakly risk-averse, the mechanism $M^* = (x^*, t^*)$ is FB, IC, and EABB.*[31]

We lay out the intuition for the proof of this result. EPEFF and EABB hold by construction. The mechanism gives a truthful agent i a deterministic payoff at the *ex-post* stage of

$$V_i^{M^*}(\theta_i) = \mathbb{E}_{\tilde{\theta}_{-i}}\left[S^*(\theta_i, \tilde{\theta}_{-i}) \right] - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta}) \right] + \alpha_i,$$

and therefore satisfies FI. (A truthful type receives his conditional expected contribution to efficient surplus, which is deterministic conditional on his type.) To get some intuition as to why the mechanism is IC, note that a deviation for agent i of type θ_i does two things. First, it changes the outcome for one that is efficient for the reported type $\hat{\theta}_i$, hence lowering true social surplus. Second, it exposes the deviator to a payoff lottery over θ_{-i}. (While under risk-neutrality, incentive-compatibility is ensured only using the first effect — as in the AGV mechanism — here, risk aversion

[31] This result does not require agents to respect first-order stochastic dominance, as Assumption 1 is used nowhere in its proof.

strengthens the argument because the deviating payoff has a certainty equivalent no larger than its expectation.)

Proof of Theorem 2. To show that $M^* = (x^*, t^*)$ is FB, it is sufficient to show that it is EPEFF and ZRP by Theorem 1. Since $x^*(\theta) \in \arg\max_{x\in\Omega} S(x,\theta)$, the mechanism is EPEFF by construction. The *ex-post* utility of agent i of type θ_i when all agents report truthfully is

$$Y_i^{M^*}(\theta_i)(\theta_{-i}) = v_i(x_i^*(\theta);\theta_i) - t_i^*(\theta) = \mathbb{E}_{\tilde{\theta}_{-i}}\left[S^*(\theta_i,\tilde{\theta}_{-i})\right] - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta})\right] + \alpha_i$$

which is independent of θ_{-i}, hence proving that (x^*, t^*) is FI (thus, it is ZRP). To prove that the mechanism is IC, suppose that bidder i reports type $\hat{\theta}_i$ while his true type is θ_i. His *ex-post* payoff is

$$Y_i^{M^*}(\theta_i,\hat{\theta}_i)(\theta_{-i}) = v_i(x_i^*(\hat{\theta}_i,\theta_{-i});\theta_i) - v_i(x_i^*(\hat{\theta}_i,\theta_{-i});\hat{\theta}_i) + \mathbb{E}_{\tilde{\theta}_{-i}}\left[S^*(\hat{\theta}_i,\tilde{\theta}_{-i})\right] - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta})\right] + \alpha_i$$

Taking expectations over θ_{-i}, this yields

$$\begin{aligned} m_i^{M^*}(\theta_i,\hat{\theta}_i) = \mathbb{E}_{\tilde{\theta}_{-i}}\left[v_i(x_i^*(\hat{\theta}_i,\theta_{-i});\theta_i) + \sum_{j\neq i} v_j(x_j^*(\hat{\theta}_i,\tilde{\theta}_{-i}),\tilde{\theta}_j) - c(x^*(\hat{\theta}_i,\tilde{\theta}_{-i}))\right] \\ - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta})\right] + \alpha_i \end{aligned}$$

By definition of $x^*(\theta) \in \arg\max_{x\in\Omega} S(x,\theta)$, we have $S^*(\theta) \geq v_i(x_i^*(\hat{\theta}_i,\theta_{-i});\theta_i) + \sum_{j\neq i} v_j(x_j^*(\hat{\theta}_i,\theta_{-i}),\theta_j) - c(x^*(\hat{\theta}_i,\theta_{-i}))$ for any $\theta_{-i} \in \times_{j\neq i}\Theta_j$. Hence, by monotonicity of the expectation operator, we have

$$m_i^{M^*}(\theta_i,\hat{\theta}_i) \leq \mathbb{E}_{\tilde{\theta}_{-i}}\left[S^*(\theta_i,\tilde{\theta}_{-i})\right] - \mathbb{E}_{\tilde{\theta}}\left[S^*(\tilde{\theta})\right] + \alpha_i$$

The right-hand side of this inequality is exactly $V_i^{M^*}(\theta_i) = m_i^{M^*}(\theta_i)$ by FI. Since all agents are weakly risk-averse, we have

$$V_i^{M^*}(\theta_i,\hat{\theta}_i) \leq m_i^{M^*}(\theta_i,\hat{\theta}_i) \leq m_i^{M^*}(\theta_i) = V_i^{M^*}(\theta_i),$$

so that the mechanism is IC. Finally, we have

$$\mathbb{E}_{\tilde{\theta}}\left[\sum_{i\in I} t_i^*(\tilde{\theta})\right] = \mathbb{E}_{\tilde{\theta}}\left[\sum_{i\in I} v_i(x^*(\tilde{\theta});\tilde{\theta}_i)\right] - \sum_{i\in I}\alpha_i = \mathbb{E}_{\tilde{\theta}}[c(x^*(\tilde{\theta}))],$$

where the last equality follows from the fact that the constants are chosen so that $\sum_{i\in I}\alpha_i = \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$. Hence, $U_d^{M^*} = 0$, and the mechanism is EABB.

□

5.2. **Failure of interim individual rationality with property rights.** So far, we have abstracted away from participation constraints to focus on the revelation of

private information. However, it is natural to ask whether the constructed mechanism is also interim individually rational, in the sense that agents prefer to participate in the mechanism at the *interim* stage rather than walking away.

Definition 12. A mechanism $M = (x, t)$ is *individually-rational (IR)* if for all agents $i \in I$, for all types $\theta_i \in \Theta_i$, $V_i^M(\theta_i) \geq 0$.

When agents have non-trivial property rights, it turns out that this mechanism may fail IR.

Example 1. We illustrate the failure of the M^* mechanism to be IR by a simple bilateral trade example. There are two agents, a buyer and a seller, whose value and cost are uniformly and independently distributed on $[0, 1]$. We have

$$\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] = \int_0^1 z(1-z)\,dz = \frac{1}{6}.$$

The M^* mechanism gives interim (and *ex-post*, by FI) utilities

$$V_B^M(\theta_B) = \frac{\theta_B^2}{2} - \frac{1}{6} + \alpha_B,$$

$$V_S^M(\theta_S) = \frac{(1-\theta_S)^2}{2} - \frac{1}{6} + \alpha_S,$$

where $\alpha_B + \alpha_S = \frac{1}{6}$. The IR constraints require that $V_B^M(0) \geq 0$, $V_S^M(1) \geq 0$, which is equivalent to $\alpha_B \geq \frac{1}{6}$, $\alpha_S \geq \frac{1}{6}$. Hence, no EABB normalization of M^* is IR in this environment. (The failure of the M^* mechanism is independent of the exact risk preferences of agents. In particular, their preferences could be described by risk-averse expected utility or by risk-averse non-expected utility.)

5.3. **Relation to existing mechanisms.** The mechanism M^* bears similarity to mechanisms that have been proposed in the literature. First, it is the analog of the VCG or AGV mechanisms for the risk-averse setting. Indeed, in the risk-neutral case, these two constructions implement the first-best with DSIC and IC mechanisms, respectively. Its payment rule, except for the lump-sum EABB normalization, coincides with Dughmi and Peres's (2012) construction for the case of risk-averse expected utility maximizers, applied to the AGV mechanism.[32] The mechanism constructed here allows agents to exhibit arbitrary weakly risk-averse preferences, and its incentive and

[32]Relative to that result, Theorem 2 extends the incentive compatibility argument beyond concave expected utility to any type of weak risk aversion, and also does not assume that the designer knows the risk-preferences of agents. Esö and Futo (1999) proposed construction related to that in Dughmi and Peres (2012), in which risk-neutral agents insure a risk-averse auctioneer.

budget properties are the best one can hope for in a general environment. Indeed, by Proposition 1 and Proposition 2, the properties of DSIC and EPBB are generically incompatible with first-best implementation when agents are risk-averse. Similarly to the AGV mechanism, it is IC and EPEFF, but trades-off EPBB against FI and EABB. The proposed mechanism also bears similarity with the coalition-proof version of the AGV mechanism proposed by Safronov (2018). Indeed, the transfer of agent i is exactly the net transfer he would make if he were placed first in Safronov's (2018) ordered externality mechanism.[33] The mechanism M^*, just like the AGV mechanism, may fail to be interim individually rational.[34]

6. Participation constraints in binary-allocation problems

As shown by Example 1, the mechanism $M^* = (x^*, t^*)$ may fail to be individually rational. Therefore, it is natural to investigate whether, in some classically studied environments with property rights, it is possible to find a FB, IC, IR, EABB mechanism, that is, a mechanism that retains the properties of M^* but also satisfies participation constraints. We thereafter specify the environment to binary-allocation problems, which include auctions, public-good provision problems, bilateral trade, and partnership dissolution as landmark applications. Formally, agents' outcome spaces are $\Omega_i = \{0, 1\}$ and the space of feasible social outcomes is a non-empty subset $\Omega \subseteq \times_{i \in I} \Omega_i$. Agent i's type space is $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$, and each F_i admits a strictly positive density $f_i > 0$.[35] Valuations are $v_i(x_i; \theta_i) = \theta_i x_i$. In this section, we allow for some form of property rights, in the sense that agents have a continuous certainty equivalent $V_i^0 : \Theta_i \to \mathbb{R}$ from not participating in a given mechanism. This requires us to redefine the right-hand side of the IR constraint as being type-dependent.

Definition 12′. A mechanism M is *individually rational (IR)* if for each agent $i \in I$, $V_i^M(\theta_i) \geq V_i^0(\theta_i)$, for all $\theta_i \in \Theta_i$.

By Theorem 1, when agents are strictly risk-averse, every FB mechanism is EPEFF and FI. In the present binary-allocation environment, the efficient outcome is unique

[33] In Safronov's (2018) mechanism, the first report changes the expected payoffs of all other agents, and later reports trigger compensating payments that telescope against the first agent's own realized payoff — the resulting net transfer has precisely the form above. Safronov's (2018) mechanism is also EPBB, which is incompatible with FB by Proposition 2. The reader is referred to Jagadeesan et al. (2025) for further use of Safronov's (2018) mechanism.

[34] The fact that the AGV mechanism may fail interim individual rationality is a direct consequence of the impossibility theorem of Myerson and Satterthwaite (1983).

[35] A property holding Lebesgue almost everywhere is equivalent to it holding F_i-almost everywhere, and we do not make the distinction thereafter.

almost everywhere.[36] We let $X_i^*(\theta_i)$ denote the resulting (non-decreasing) interim efficient allocation probability.

6.1. **Risk preferences and orders of risk aversion.** We are interested in the existence of a FB, IC, IR, EABB mechanism in this setting. When mechanisms are FI, a deviation by a player makes him face a binary lottery, so only the preferences of agents on binary lotteries matter. In order to be able to carry the analysis through, we assume that the behavior of agents when faced with binary lotteries can be described by *concave biseparable* preferences (Ghirardato et al. (2004)).

Assumption 3 (Concave biseparable preferences on binary lotteries)**.** For each agent i, the restriction of his certainty equivalent functional to binary lotteries does not depend on his type and is *concave biseparable*, that is there exists a continuously differentiable, concave, strictly increasing Bernoulli utility $\varphi_i : \mathbb{R} \to \mathbb{R}$ and a continuous, increasing, probability weighting function $g_i : [0,1] \to [0,1]$, with $g_i(0) = 0$, $g_i(1) = 1$, such that, for every $z \geq y$ and $p \in [0,1]$, $g_i(p) \leq p$ and

$$\varphi_i(\mathrm{CE}_i(z,p;y,1-p)) = g_i(p)\varphi_i(z) + (1-g_i(p))\varphi_i(y), \tag{3}$$

where $(z,p;y,1-p)$ is the binary lottery giving z with probability p and y with probability $1-p$.

We do not restrict the risk preferences of agents outside of binary lotteries, beyond Assumption 1 and risk aversion in the sense of Definition 1.[37] Assumption 3 contains notable special cases, such as when agents' risk preferences can be described by von Neumann and Morgenstern's (1944) risk-averse expected utility with a smooth Bernoulli utility, Quiggin's (1993) rank-dependent utility with a smooth concave utility over outcomes and Safra and Segal's (1998) constant risk aversion preferences.[38,39] Let B_p denote a Bernoulli random variable with probability of success p, and denote by $Z_p = B_p - p$.

[36] Since $\Omega \subseteq \{0,1\}^I$ is finite, the efficient outcome is unique F-almost everywhere. We fix one measurable efficient selector x^* and use this same selector in all canonical constructions below. All integrated objects are invariant to this choice.

[37] Allowing privately known risk-attitudes on binary lotteries would turn the analysis of this section into a genuinely multidimensional screening problem, which is beyond the scope of the present paper.

[38] Safra and Segal's (1998) constant risk aversion preferences admit as special cases Gul's (1991) disappointment-averse preferences with linear utility over outcomes, Kőszegi and Rabin's (2007) loss-averse preferences with linear utility over outcomes and Yaari's (1987) dual expected utility. For further special cases, see Gershkov et al. (2022) and Ahn et al. (2026).

[39] The preferences of the agents could also coincide with any of the above on binary lotteries, and differ elsewhere.

Under Assumption 3, if φ_i is twice-continuously differentiable,

$$\mathrm{CE}_i(w+\varepsilon Z_p) \underset{\varepsilon\to 0}{=} w+\varepsilon(g_i(p)-p)-\frac{1}{2}A_i(w)g_i(p)(1-g_i(p))\varepsilon^2+o(\varepsilon^2),$$

where $A_i(w)=-\frac{\varphi_i''(w)}{\varphi_i'(w)}$. When $g_i(p)=p$ and $\varphi_i''(w)<0$, the first-order term vanishes and agents' risk preferences reduce to risk-averse expected utility on binary lotteries at wealth w, with a twice continuously differentiable, concave Bernoulli utility, so that the agent exhibits second-order risk aversion in the sense of Segal and Spivak (1990).[40] When $g_i(p)<p$ for all $p\in(0,1)$, the preferences compatible with Assumption 3 exhibit first-order risk aversion in the sense of Segal and Spivak (1990).

6.2. **Necessary and sufficient condition for existence.** We first characterize the truthful certainty-equivalent schedules generated by FB, IC mechanisms. (The formal argument is in the appendix.)

Lemma 5. *For each agent $i\in I$, let $V_i:\Theta_i\to\mathbb{R}$ be a bounded function. Under Assumption 3 and strict risk-aversion, the following are equivalent:*

(i) there exists a FB, IC mechanism M such that for each agent $i\in I$, $V_i(\theta_i)=V_i^M(\theta_i)$ for F_i-almost all θ_i,

(ii) for each agent $i\in I$, V_i admits an absolutely continuous version and

$$g_i(X_i^*(\theta_i))\le V_i'(\theta_i)\le 1-g_i(1-X_i^*(\theta_i)), \tag{4}$$

for F_i-almost all θ_i.

Moreover, whenever (ii) holds, a (canonical, deterministic) implementing mechanism is obtained by setting

$$t_i(\theta)=\theta_i x_i^*(\theta)-V_i(\theta_i). \tag{5}$$

Lemma 5 generalizes (Gershkov et al., 2022, Proposition 3), who derived a similar result for the class of preferences exhibiting constant risk aversion. Contrary to Gershkov et al. (2022), we do not assume that preferences satisfy cash-invariance — indeed, Lemma 5 shows that the same characterization survives with an arbitrary concavity in φ_i on binary lotteries. In particular, when $g_i(p)=p$ and φ_i is an arbitrary concave smooth function, Lemma 5 reduces to the same payoff-equivalence formula as under risk-neutrality even though agents exhibit risk-averse expected utility preferences on binary lotteries. Towards our necessary and sufficient condition for

[40] Smoothness of the derivative of the Bernoulli utility φ_i is essential to get second-order risk aversion, as discussed by Segal and Spivak (1990).

the existence of a FB, IC, IR, EABB mechanism, define, for an agent $i \in I$ with type θ_i who reports $\hat{\theta}_i$,

$$D_i(\theta_i, \hat{\theta}_i) = \begin{cases} \int_{\hat{\theta}_i}^{\theta_i} g_i(X_i^*(s))\, ds & \text{if } \hat{\theta}_i \leq \theta_i, \\ -\int_{\theta_i}^{\hat{\theta}_i} (1 - g_i(1 - X_i^*(s))\, ds & \text{otherwise.} \end{cases}$$

In light of Lemma 5, $D_i(\theta_i, \hat{\theta}_i)$ is the smallest payoff at θ_i implied by giving a payoff of zero at $\hat{\theta}_i$, given the bounds in (4). Define agent i's minimal IC, IR payoff schedule by

$$V_i^{\min}(\theta_i) = \sup_{\hat{\theta}_i \in \Theta_i} (V_i^0(\hat{\theta}_i) + D_i(\theta_i, \hat{\theta}_i)).$$

(Because Θ_i is compact, V_i^0 is continuous and D_i is continuous, Berge's maximum theorem implies that the supremum is finite and attained and that $V_i^{\min}$ is continuous.) We thereafter derive the main result in this section, namely a necessary and sufficient condition, under Assumption 3, for the existence of a FB, IC, IR, EABB mechanism.

Theorem 3. *Suppose all agents are strictly risk-averse, under Assumption 3, there exists a FB, IC, IR, EABB mechanism if and only if*

$$\sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^{min}(\theta_i)] \leq \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] \tag{6}$$

Moreover, the set of attainable expected interim certainty equivalent profiles in such a mechanism is the set of $(U_i)_{i \in I} \in \mathbb{R}^I$ such that $U_i \geq \mathbb{E}_{\tilde{\theta}_i}[V_i^{min}(\theta_i)]$ for each agent i, and $\sum_{i \in I} U_i = \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$.

The complete argument is in the appendix. The proof relies on first showing that $V_i^{\min}$ is the pointwise least IC, IR payoff schedule compatible with the efficient reduced form X^*. Necessity follows straightforwardly from this, while sufficiency is proven by constructing a FB mechanism giving interim (and *ex-post*, by FI) utility $V_i^{\min}$ to each agent, up to a constant tailored to make EABB hold. With type-dependent outside options, the participation constraint need not bind at the lowest type, and the lower IC slope bound need not bind everywhere.

6.3. **Application to bilateral trade.** We now discuss the implications of Theorem 3 for the canonical bilateral trade problem. In bilateral trade, there are two agents, a buyer and a seller whose value and cost are respectively distributed according to F_B and F_S on $[0,1]$.[41] The model is described by $\Omega = \{(0,0),(1,1)\}$, where $v_B(x_B; \theta_B) =$

[41]The type of the seller, which multiplies the allocation, is interpreted as the negative of his cost of providing the good. The case of overlapping but non-identical supports yields the same intuition, so we focus on the present model for notational convenience.

$\theta_B x_B$, $v_S(x_S;\theta_S) = \theta_S x_S$, and $V_B^0(\theta_B) = 0$, $V^0(\theta_S) = 0$. One efficient allocation rule is for example $x_B^*(\theta_B,\theta_S) = x_S^*(\theta_B,\theta_S) = 1(\theta_B \geq -\theta_S)$. Both agents are strictly risk-averse. We denote the minimal rents in an IC, IR, FB, EABB mechanism by

$$R_B = \int_0^1 (1 - F_B(z)) g_B(F_S(z))\, dz,$$

$$R_S = \int_0^1 F_S(z) g_S(1 - F_B(z))\, dz.$$

The *ex-ante* gains from trade are

$$\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] = \int_0^1 F_S(z)(1 - F_B(z))\, dz.$$

When agents are risk-neutral, Myerson and Satterthwaite (1983) have shown that no EPEFF (i.e., FB under risk-neutrality), IC, IR, EABB mechanism exists. The same exercise can be investigated in the present setting, where FI (or ZRP) is added to EPEFF by the characterization of our welfare criterion from Theorem 1.

Corollary 1. *In the bilateral trade model, under Assumption 3, there exists a FB, IC, IR, EABB mechanism if and only if* $R_B + R_S \leq \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$.

Corollary 1 is a direct consequence of Theorem 3, and its proof is omitted. There, the IR constraint binds for the lowest types (lowest buyer's value and highest seller's cost) while every agent's certainty equivalent's derivative is brought to the lower bound in (4).

Corollary 2. *Under Assumption 3, if* $g_B = id_{[0,1]}$ *or* $g_S = id_{[0,1]}$*, then no mechanism is FB, IC, IR and EABB in this setting.*

Proof. Assume $g_B(p) = p$ for all $p \in [0,1]$ without loss, then we have $R_B = \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$, so that $R_B + R_S \leq \mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})]$ can never hold. Indeed, $R_S > 0$ because the distributions have full-support, g_S is continuous and $g_S(1) = 1$. □

Corollary 2 shows that, even with an arbitrary concavity in φ_i, second-order risk aversion prevents the existence of a FB, IC, IR, EABB mechanism. Therefore, the impossibility result of Myerson and Satterthwaite (1983) extends in this setting to risk-averse expected utility maximizers, with arbitrary smooth concave Bernoulli utility, beyond the risk-neutral case.[42,43]

[42] It also extends to any risk-averse preferences that coincide with smooth, risk-averse expected utility only on binary lotteries, but are unrestricted elsewhere.

[43] Garratt and Pycia (2026) study the implementation of the *ex-post* efficient allocation rule in bilateral trade when parties have risk-averse expected-utility. Their notion of welfare does not

Second, when agents exhibit first-order risk aversion, the impossibility still holds if both agents are sufficiently close to being risk-neutral — see Examples 2 and 3 below for more on this point — because the inequality in Corollary 1 cannot hold. Third, a pointwise decrease in g_i weakly lowers that agent's minimal rent and therefore weakly relaxes condition (6).

We next present two examples to understand some of the implications of Corollary 1. In all examples, the buyer's value and the seller's cost are independently and identically distributed according to a uniform distribution on $[0,1]$. Hence, $\mathbb{E}_{\tilde{\theta}}[S^*(\tilde{\theta})] = \int_0^1 z(1-z)\,dz = \frac{1}{6}$. In both examples, the agents exhibit first-order risk aversion.

Example 2. Consider first-order risk preferences in the power family,

$$g_B(p) = g_S(p) = p^\alpha,$$

parameterized by a real number $\alpha > 1$, with arbitrary smooth, concave utility over money φ_B, φ_S. Then,

$$R_B + R_S = \int_0^1 [(1-z)z^\alpha + z(1-z)^\alpha]\,dz = \frac{2}{(\alpha+1)(\alpha+2)}.$$

The existence of a FB, IC, IR, EABB mechanism requires that $\frac{2}{(\alpha+1)(\alpha+2)} \leq \frac{1}{6}$, which is equivalent to $\alpha \geq 2$. Hence, the agents need to be sufficiently risk-averse in order to be able to overturn the impossibility result of Myerson and Satterthwaite (1983). In the limit case where $\alpha = 2$, the mechanism of Theorem 3 gives interim (and *ex-post*, by FI) utilities

$$V_B(\theta_B) = \int_0^{\theta_B} z^2\,dz = \frac{\theta_B^3}{3},$$
$$V_S(\theta_S) = \int_{\theta_S}^1 (1-z)^2\,dz = \frac{(1-\theta_S)^3}{3},$$

under truthful reporting.

Example 3. Consider risk preferences in the linear-quadratic family,

$$g_S(p) = g_B(p) = (2-\lambda)p + (\lambda-1)p^2,$$

require their risk-neutral designer to insure agents. Therefore, they are able to find budget-balanced, incentive-compatible, individually rational mechanisms which implement an *ex-post* efficient rule by exposing parties to risky prospects.

with $\lambda \in (1,2]$, and $\varphi_B(z) = \varphi_S(z) = z$, which correspond to Kőszegi and Rabin's (2007) loss-averse preferences with linear utility over outcomes. Then,

$$R_B + R_S = \int_0^1 2[(1-z)((2-\lambda)z + (\lambda-1)z^2)]\,dz = \frac{1}{2} - \frac{\lambda}{6}.$$

Hence, only in the knife-edge case where $\lambda = 2$ will the inequality $R_B + R_S \leq \frac{1}{6}$ be satisfied, and in this case the mechanism corresponds to that of Example 2 for $\alpha = 2$. In all other cases, it is impossible for the designer to find a FB, IC, IR, EABB mechanism when agents exhibit Kőszegi and Rabin's (2007) loss-averse preferences with linear utility over outcomes.

Corollary 2 shows that second-order risk-aversion can never help overturn the result of Myerson and Satterthwaite (1983). Example 3 shows that first-order risk aversion is not by itself enough to ensure existence of a FB, IC, IR, EABB mechanism in bilateral trade. Example 2 shows that sufficiently (first-order) risk-averse agents can help overturn Myerson and Satterthwaite's (1983) impossibility result.

7. Conclusion

In this paper, a theory of the first-best in the presence of private information and of a risk-neutral uninformed designer has been proposed. A general incentive-compatible, first-best mechanism has been constructed. The possibility of implementing the first-best in an incentive-compatible mechanism which also respects agents' participation constraints and the designer's budget constraint has been studied. This last analysis sheds light on the differences between first-order and second-order risk aversion when it comes to economic design, in particular when one tries to design full insurance, efficient mechanisms. A more systematic analysis of the differences between these forms of risk aversion in mechanism design problems is surely interesting and is left for future research.

Appendix A. Omitted Proofs

A.1. **Proof of Lemma 1.** Since I is finite and each v_i and c is continuous, the function S is continuous on the compact metric space $\Omega \times \Theta$. The Weierstrass theorem (Aliprantis and Border, 2006, Theorem 2.34) implies that $\Omega^*(\theta)$ is nonempty and compact for each $\theta \in \Theta$. By Berge's maximum theorem, S^* is continuous and Ω^* is upper hemicontinuous. A compact-valued upper hemi-continuous correspondence from a compact metric space into a compact metric space is (Borel) measurable. The Kuratowski–Ryll-Nardzewski measurable-selection theorem (Aliprantis and Border, 2006, Theorem 18.13) therefore yields a Borel measurable selector x^*.

A.2. **Proof of Theorem 1.** Assume that all agents are weakly risk-averse. By Lemma 3, $\mathcal{W}(M) \leq \int_\Theta S^*(\theta)\, dF(\theta)$ for all mechanisms M. Since both loss terms in the right-hand side of (2) are non-negative, equality holds if and only if $S^*(\theta) = S(x(\theta, r), \theta)$ for $\mathbb{P}$-almost every (θ, r) and, for each agent $i \in I$, $m_i^M(\theta_i) = V_i^M(\theta_i)$ for F_i-almost every θ_i, that is, if and only if M is EPEFF and ZRP. This proves that (ii) is equivalent to (iii).[44]

If a mechanism M is not FB, then there exists a mechanism M' which CE-Pareto dominates M, and hence $\mathcal{W}(M') > \mathcal{W}(M)$ by the definition of $\mathcal{W}$, so that M cannot maximize $\mathcal{W}$ over all mechanisms. This proves that (ii) implies (i).

Assume that $M = (x, t)$ is not EPEFF or not ZRP. Define a deterministic mechanism $M' = (x^*, t')$ where x^* is an efficient-selector from Lemma 1, by

$$t_i'(\theta) = v_i(x_i^*(\theta); \theta_i) - m_i^M(\theta_i).$$

Under M', each truthful type obtains a deterministic payoff $m_i^M(\theta_i)$, so that

$$V_i^{M'}(\theta_i) = m_i^M(\theta_i) \geq V_i^M(\theta_i). \tag{A.1}$$

Moreover, by Lemma 2, we have

$$U_d^{M'} - U_d^M = \mathbb{E}_{\tilde{\theta}, r}[S^*(\tilde{\theta}) - S(x(\theta, r), \theta)] \geq 0.$$

If M is not EPEFF, then $U_d^{M'} > U_d^M$. If M is not ZRP, then (A.1) holds strictly for a positive measure set under F_i. In any case, M' CE-Pareto dominates M, so that M is not FB. This proves that (i) implies (iii).

[44]Note that the formulation in (ii) is correct. Indeed, the maximum is attained by considering the mechanism (x^*, t) where x^* is a measurable selection from Lemma 1 and for all agents i, $t_i(\theta) = v_i(x_i^*(\theta); \theta_i)$ so that all agents' *ex-post* utilities under truthful reporting are zero.

Finally, note that under risk aversion, equality between the certainty equivalent and the mean is possible only for degenerate lotteries, so that ZRP becomes FI.

A.3. **Proof of Lemma 4.** Let $M = (x, t)$ be a FB mechanism. By Theorem 1, M is ZRP and EPEFF. Consider the following mechanism $M' = (x, t')$ with the same outcome rule as M and transfer rule defined by

$$t'_i(\theta, r) = v_i(x_i(\theta, r); \theta_i) - m_i^M(\theta_i).$$

This mechanism is EPEFF by definition, and FI as $Y_i^{M'}(\theta_i)(\theta_{-i}, r) = m_i^M(\theta_i)$. Since M is ZRP, the certainty equivalents under truthful reporting in M' are the same as those in M, and the designer's utility is unchanged by Lemma 2. Hence, the two mechanisms are CE-Pareto equivalent.

A.4. **Proof of Proposition 1.** Suppose S^* is additively separable, that is for each agent i, there exists a continuous function $w_i : \Theta_i \to \mathbb{R}$ such that $S^*(\theta) = \sum_{i \in I} w_i(\theta_i)$ for all $\theta \in \Theta$. Consider the following deterministic mechanism, $M = (x^*, t)$ where x^* is a selection from Lemma 1 and $t_i(\theta) = v_i(x^*(\theta); \theta_i) - w_i(\theta_i)$. The mechanism is EPEFF and FI by construction, so it is FB by Theorem 1. We have, for any agent i of type θ_i and any report $\hat{\theta}_i$,

$$\begin{aligned} S^*(\theta_i, \theta_{-i}) &\geq v_i(x_j^*(\hat{\theta}_i, \theta_{-i}); \theta_i) + \sum_{j \neq i} v_j(x_i^*(\hat{\theta}_i, \theta_{-i}); \theta_j) - c(x^*(\hat{\theta}_i, \theta_{-i})) \\ &= S^*(\hat{\theta}_i, \theta_{-i}) + v_i(x_i^*(\hat{\theta}_i, \theta_{-i}); \theta_i) - v_i(x_i^*(\hat{\theta}_i, \theta_{-i}); \hat{\theta}_i), \end{aligned}$$

where both the inequality and the equality are by definition of the efficient surplus function. Note that $S^*(\theta_i, \theta_{-i}) - S^*(\hat{\theta}_i, \theta_{-i}) = w_i(\theta_i) - w_i(\hat{\theta}_i)$ by additive separability. Hence, we have

$$w_i(\theta_i) \geq w_i(\hat{\theta}_i) + v_i(x_i^*(\hat{\theta}_i, \theta_{-i}); \theta_i) - v_i(x_i^*(\hat{\theta}_i, \theta_{-i}); \hat{\theta}_i).$$

The left-hand side is the *ex-post* utility from truthful reporting while the right-hand is the *ex-post* payoff for agent i of type θ_i to report $\hat{\theta}_i$. Therefore, M is DSIC.

Conversely, let M be a deterministic DSIC, FB mechanism. Under Assumption 2, F_i is equivalent to the Lebesgue measure on Θ_i. By Theorem 1, the mechanism is EPEFF and FI, and the properties hold almost everywhere on Θ. By DSIC, we have

$$Y_i^M(\theta_i)(\theta_{-i}) = \max_{\hat{\theta}_i \in \Theta_i} Y_i^M(\theta_i, \hat{\theta}_i)(\theta_{-i}), \tag{A.2}$$

for all $\theta_{-i} \in \Theta_{-i}$. By Assumption 2, $Y_i^M(.)(\theta_{-i})$ is a Lipschitz function, since $\partial_{\theta_i} v_i$ is uniformly bounded. By Rademacher's Theorem, $Y_i^M(.)(\theta_{-i})$ is differentiable almost

everywhere. For the same reasons, $S^*(.,\theta_{-i})$ is differentiable almost everywhere. We use the standard multidimensional consequence of (Milgrom and Segal, 2002, Theorem 1) envelope theorem obtained by applying their scalar theorem along line segments. At every differentiability point of a Lipschitz value function, its gradient equals the type gradient of the objective evaluated at any optimizer. Applying this to (A.2) yields

$$\partial_{\theta_i} Y_i^M(\theta_i)(\theta_{-i}) = \partial_{\theta_i} v_i(x_i(\theta_i,\theta_{-i});\theta_i).$$

Similarly, applying the same results to S^* at points at which x is *ex-post* efficient yields

$$\partial_{\theta_i} v_i(x_i(\theta_i,\theta_{-i});\theta_i) = \partial_{\theta_i} S^*(\theta_i,\theta_{-i}),$$

almost everywhere in Θ. Since M is EPEFF, the two identities hold simultaneously almost everywhere. By Fubini's theorem, the Lipschitz function $S^*(.,\theta_{-i}) - Y_i^M(.)(\theta_{-i})$ has gradient zero almost everywhere on the convex set Θ_i. By absolute continuity on lines, it is constant. Combined with FI, we get that there exist Borel measurable functions $w_i : \Theta_i \to \mathbb{R}$ and $h_i : \Theta_{-i} \to \mathbb{R}$ such that,

$$S^*(\theta) = w_i(\theta_i) + h_i(\theta_{-i}) \quad \text{for } F\text{-almost every } \theta.$$

For $i \neq j$, define the rectangle difference

$$\begin{aligned} D_{ij}S^*(\theta_i,\theta_j,\theta_i',\theta_j',\theta_{-ij}) = {} & S^*(\theta_i,\theta_j,\theta_{-ij}) - S^*(\theta_i',\theta_j,\theta_{-ij}) \\ & - S^*(\theta_i,\theta_j',\theta_{-ij}) + S^*(\theta_i',\theta_j',\theta_{-ij}), \end{aligned}$$

we have $D_{ij}S^*(\theta_i,\theta_j,\theta_i',\theta_j',\theta_{-ij}) = 0$ for $F_i \otimes F_i \otimes F_j \otimes F_j \otimes F_{-ij}$-almost every $(\theta_i,\theta_j,\theta_i',\theta_j',\theta_{-ij})$. Since S^* is continuous and the relevant product measure has full support, $D_{ij}S^* = 0$ identically. Fix a type profile $\bar{\theta} \in \Theta$, and define $\tilde{w}_i(\theta_i) = S^*(\theta_i,\bar{\theta}_{-i}) - S^*(\bar{\theta})$. The vanishing of all mixed rectangle differences implies that the effect of changing coordinate i is independent of all other coordinates. Changing one coordinate at a time gives

$$S^*(\theta) = S^*(\bar{\theta}) + \sum_{i \in I} \tilde{w}_i(\theta_i),$$

so that $S^*$is additively separable.

A.5. **Proof of Proposition 2.** Suppose the surplus function is essentially additively separable, $S^*(\theta) = \sum_{i\in I} w_i(\theta_i)$ for F-almost every θ. Denote by N the (Borel) F-null set on which essential additive separability does not hold. Consider the (deterministic) mechanism $M = (x^*, t)$ where x^* is an efficient selection from Lemma 1. For $\theta \notin N$, define $t_i(\theta) = v_i(x^*(\theta);\theta_i) - w_i(\theta_i)$. For $\theta \in N$, fix an arbitrary agent i_0 and define

$t_{i_0}(\theta) = c(x^*(\theta))$ and $t_i(\theta) = 0$ for $i \neq i_o$. These transfers are (Borel) measurable and bounded. The mechanism is EPEFF and FI by construction. Observe that for $\theta \notin N$,

$$\sum_{i\in I} t_i(\theta) - c(x^*(\theta)) = \sum_{i\in I} v_i(x^*(\theta);\theta_i) - \sum_{i\in I} w_i(\theta_i) - c(x^*(\theta)) = S^*(\theta) - S^*(\theta) = 0,$$

while for $\theta \in N$, $\sum_{i\in I} t_i(\theta) = c(x^*(\theta))$, so that the mechanism is also EPBB.

Conversely, let $M = (x,t)$ be an EPEFF, FI and EPBB mechanism. For each agent $i \in I$, there exists a (Borel) measurable function $w_i : \Theta_i \to \mathbb{R}$ such that $v_i(x_i(\theta,r);\theta_i) - t_i(\theta,r) = w_i(\theta_i)$ for $\mathbb{P}$-almost all (θ,r). Therefore, we have

$$\begin{aligned}
\sum_{i\in I} w_i(\theta_i) &= \sum_{i\in I} v_i(x_i(\theta,r);\theta_i) - \sum_{i\in I} t_i(\theta,r) \quad \text{for } \mathbb{P}\text{-almost all } (\theta,r),\\
&= \sum_{i\in I} v_i(x_i(\theta,r);\theta_i) - c(x(\theta,r)) \quad \text{for } \mathbb{P}\text{-almost all } (\theta,r),\\
&= S(x(\theta,r),\theta) \quad \text{for } \mathbb{P}\text{-almost all } (\theta,r),\\
&= S^*(\theta) \quad \text{for } F\text{-almost all } \theta,
\end{aligned}$$

where the second equality is by EPBB, the third by definition and the fourth by EPEFF. Hence, the surplus function is essentially additively separable.

A.6. **Proof of Lemma 5.** Assume Assumption 3 holds. Let $M = (x,t)$ be a FB, IC mechanism. By Theorem 1, M is EPEFF and FI. Hence, for each $i \in I$, there exists a bounded measurable function $w_i : \Theta_i \to \mathbb{R}$ such that $\theta_i x_i(\theta,r) - t_i(\theta,r) = w_i(\theta_i)$ for $\mathbb{P}$-almost every (θ,r). By EPEFF, uniqueness of the efficient outcome almost everywhere, and Fubini's theorem, there exists a set $T_i \subseteq \Theta_i$ with $F_i(T_i) = 1$, such that, for every $\theta_i \in T_i$,

$$\theta_i x_i(\theta,r) - t_i(\theta,r) = w_i(\theta_i), \quad x_i(\theta_i,\theta_{-i},r) = x_i^*(\theta_i,\theta_{-i}),$$

for $F_{-i} \otimes \lambda$-almost every (θ_{-i},r). Fix $\theta_i, \hat{\theta}_i \in T_i$ such that $\theta_i > \hat{\theta}_i$. If agent i of type θ_i reports $\hat{\theta}_i$, we have

$$Y_i^M(\theta_i,\hat{\theta}_i)(\theta_{-i}) = w_i(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)x_i^*(\hat{\theta}_i,\theta_{-i}).$$

By Assumption 1 and IC, we have

$$w_i(\theta_i) \geq w_i(\hat{\theta}_i).$$

Conversely, if true type $\hat{\theta}_i$ reports θ_i, his payoff is

$$Y_i^M(\hat{\theta}_i,\theta_i)(\theta_{-i}) = w_i(\theta_i) + (\hat{\theta}_i - \theta_i)x_i^*(\theta_i,\theta_{-i}),$$

which first-order stochastically dominates $w_i(\theta_i) - (\theta_i - \hat{\theta}_i)$, so that IC gives

$$w_i(\hat{\theta}_i) \geq w_i(\theta_i) - (\theta_i - \hat{\theta}_i).$$

Combining the above, we get

$$0 \leq w_i(\theta_i) - w_i(\hat{\theta}_i) \leq \theta_i - \hat{\theta}_i,$$

so that w_i is Lipschitz on the dense set T_i (because F_i admits a strictly positive density). Hence, it admits a unique Lipschitz extension to Θ_i, which we denote by V_i. This extension is absolutely continuous and agrees with the original truthful certainty-equivalent schedule except possibly on a F_i-null set.

Fix a point $\theta_i \in T_i$ at which V_i is differentiable and X_i^* is continuous.[45] Choose $\varepsilon_n > 0$ such that $\theta_i + \varepsilon_n \in T_i$ and $\varepsilon_n \underset{n\to\infty}{\to} 0$. Type $\theta_i + \varepsilon_n$ can report θ_i, and we have

$$V_i(\theta_i + \varepsilon_n) \geq \mathrm{CE}_i(V_i(\theta_i) + \varepsilon_n x_i^*(\theta_i, \theta_{-i})) \underset{n\to\infty}{=} V_i(\theta_i) + \varepsilon_n g_i(X_i^*(\theta_i)) + o(\varepsilon_n),$$

where the first inequality is by IC, and the displayed expansion is by Assumption 3.[46] Dividing by $\varepsilon_n > 0$ and letting it $n \to \infty$ shows $(V_i)'(\theta_i) \geq g_i(X_i^*(\theta_i))$. For the reverse deviation, type θ_i can report to be $\theta_i + \varepsilon_n$. By IC and Assumption 3, we have

$$V_i(\theta_i) \geq \mathrm{CE}_i(V_i(\theta_i+\varepsilon_n) - \varepsilon_n x_i^*(\theta_i+\varepsilon_n, \theta_{-i})) \underset{n\to\infty}{=} V_i(\theta_i+\varepsilon_n) - \varepsilon_n(1 - g_i(1 - X_i^*(\theta_i+\varepsilon_n))) + o(\varepsilon_n),$$

Dividing by $\varepsilon_n > 0$ and letting $n \to \infty$, and using continuity of X_i^* at θ_i, we get

$$V_i'(\theta_i) \leq 1 - g_i(1 - X_i^*(\theta_i)),$$

so that (4) holds almost everywhere. This proves that (i) implies (ii).

Conversely, suppose V_i is absolutely continuous and satisfies (4). Choose any efficient selector x^* from Lemma 1, and consider the canonical deterministic mechanism $M = (x^*, t)$ where t is given by (5). By Theorem 1, this mechanism being FI, EPEFF, with $V_i(\theta_i) = V_i^M(\theta_i)$, it is FB. It remains to prove that it is IC. Fix $\hat{\theta}_i < \theta_i$. If type θ_i reports $\hat{\theta}_i$, his *ex-post* payoff is

$$Y_i^M(\theta_i, \hat{\theta}_i)(\theta_{-i}) = V_i^M(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i) x_i^*(\hat{\theta}_i, \theta_{-i})$$

[45] Such points have full measure because V_i is absolutely continuous and X_i^* is monotone.

[46] Indeed, Assumption 3 implies that $\mathrm{CE}_i(w + \varepsilon B_p) \underset{\varepsilon\to 0}{=} w + g_i(p)\varepsilon + o(\varepsilon)$ and $\mathrm{CE}_i(w - \varepsilon B_p) \underset{\varepsilon\to 0}{=} w - (1 - g_i(1-p))\varepsilon + o(\varepsilon)$. This is due to the fact that $\varphi_i' > 0$ cancels out from a first-order Taylor expansion of (3).

The lower bound of the derivative of V_i^M and monotonicity of X_i^* and g_i give

$$V_i^M(\theta_i) - V_i^M(\hat{\theta}_i) = \int_{\hat{\theta}_i}^{\theta_i} (V_i^M)'(s)\, ds \geq \int_{\hat{\theta}_i}^{\theta_i} g_i(X_i^*(s))\, ds \geq (\theta_i - \hat{\theta}_i) g_i(X_i^*(\hat{\theta}_i)).$$

Furthermore, we have

$$\begin{aligned}
\varphi_i(V_i^M(\theta_i)) &\geq \varphi_i(V_i^M(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i) g_i(X_i^*(\hat{\theta}_i)) \\
&\geq g_i(X_i^*(\hat{\theta}_i))\varphi_i(V_i^M(\hat{\theta}_i) + \theta_i - \hat{\theta}_i)) + (1 - g_i(X_i^*(\hat{\theta}_i))\varphi_i(V_i^M(\hat{\theta}_i)) \\
&= \varphi_i(V_i^M(\theta_i, \hat{\theta}_i))
\end{aligned}$$

where the first inequality is by monotonicity of φ_i, the second by concavity, and the equality by Assumption 3. Hence, because $\varphi_i' > 0$, we have $V_i^M(\theta_i) \geq V_i^M(\theta_i, \hat{\theta}_i)$, so that downward deviations are unprofitable. The argument for upward deviations is similar. We have

$$Y_i^M(\hat{\theta}_i, \theta_i)(\theta_{-i}) = V_i^M(\theta_i) + (\hat{\theta}_i - \theta_i) x_i^*(\theta_i, \theta_{-i}).$$

Then,

$$V_i^M(\theta_i) - V_i^M(\hat{\theta}_i) \leq \int_{\hat{\theta}_i}^{\theta_i} (1 - g_i(1 - X_i^*(s)))\, ds \leq (\theta_i - \hat{\theta}_i)(1 - g_i(1 - X_i^*(\theta_i))).$$

Furthermore, by an argument similar to the one above,

$$\begin{aligned}
\varphi_i(V_i^M(\hat{\theta}_i, \theta_i)) &= g_i(1 - X_i^*(\theta_i))\varphi_i(V_i^M(\theta_i)) + (1 - g_i(1 - X_i^*(\theta_i)))\varphi_i(V_i^M(\theta_i) - \theta_i + \hat{\theta}_i) \\
&\leq \varphi_i(V_i^M(\theta_i) - (\theta_i - \hat{\theta}_i)(1 - g_i(1 - X_i^*(\theta_i)))) \\
&\leq \varphi_i(V_i^M(\hat{\theta}_i)),
\end{aligned}$$

so that the reverse deviation is also unprofitable. Hence, the mechanism is IC. This proves that (ii) implies (i).

A.7. **Proof of Theorem 3.** Suppose that $M = (x, t)$ is FB, IC, IR and EABB and that Assumption 3 holds. Because all agents are risk-averse, Theorem 1 implies that M is FI and EPEFF. By Lemma 5, for each agent $i \in I$, the truthful certainty equivalent V_i^M admits an absolutely continuous version, denoted V_i, satisfying (4) for F_i-almost every θ_i and such that $V_i(\theta_i) = V_i^M(\theta_i)$ for F_i-almost every θ_i. Because M is IR, $V_i(\theta_i) \geq V_i^0(\theta_i)$, for F_i-almost every θ_i. Actually, the inequality holds pointwise in θ_i, because V_i and V_i^0 are continuous, the former by absolute continuity. (Indeed, if the inequality failed at some type, continuity would make it fail on a nonempty interval, contradicting the almost-everywhere inequality.) As an intermediate step,

we prove that $V_i^{\min}$ is the pointwise least IC and IR payoff schedule compatible with EPEFF.

Lemma 6. *Under Assumption 3, when agents are strictly risk-averse, the function V_i^{min} is absolutely continuous and satisfies (4) for F_i-almost every θ_i, as well as $V_i^{min}(\theta_i) \geq V_i^0(\theta_i)$ for all $\theta_i \in \Theta_i$. Moreover, if $V_i : \Theta_i \to \mathbb{R}$ is any absolutely continuous function satisfying (4) for F_i-almost all θ_i, and $V_i(\theta_i) \geq V_i^0(\theta_i)$ for all $\theta_i \in \Theta_i$, then $V_i(\theta_i) \geq V_i^{min}(\theta_i)$ for all $\theta_i \in \Theta_i$.*

Proof. By definition, $V_i^{\min}(\theta_i) \geq V_i^0(\theta_i)$ for all $\theta_i \in \Theta_i$. Fix $\theta_i < \theta_i'$. A direct examination of the cases $\hat{\theta}_i \leq \theta_i$, $\theta_i < \hat{\theta}_i < \theta_i'$ and $\hat{\theta}_i \geq \theta_i'$ yields

$$\int_{\theta_i}^{\theta_i'} g_i(X_i^*(s))\,ds \leq D_i(\theta_i', \hat{\theta}_i) - D_i(\theta_i, \hat{\theta}_i) \leq \int_{\theta_i}^{\theta_i'} (1 - g_i(1 - X_i^*(s))\,ds,$$

for every $\hat{\theta}_i \in \Theta_i$, where we use the fact that $g_i(p) \leq p \leq 1 - g_i(1-p)$ for $p \in [0,1]$. Therefore,

$$\begin{aligned} V_i^{\min}(\theta_i') &= \sup_{\hat{\theta}_i \in \Theta_i} \left(V_i^0(\hat{\theta}_i) + D_i(\theta_i', \hat{\theta}_i)\right) \\ &\geq \sup_{\hat{\theta}_i \in \Theta_i} \left(V_i^0(\hat{\theta}_i) + D_i(\theta_i, \hat{\theta}_i)\right) + \int_{\theta_i}^{\theta_i'} g_i(X_i^*(s))\,ds \\ &= V_i^{\min}(\theta_i) + \int_{\theta_i}^{\theta_i'} g_i(X_i^*(s))\,ds, \end{aligned}$$

and similarly,

$$V_i^{\min}(\theta_i') \leq V_i^{\min}(\theta_i) + \int_{\theta_i}^{\theta_i'} (1 - g_i(1 - X_i^*(s))\,ds.$$

Hence, $V_i^{\min}$ is Lipschitz, and hence absolutely continuous, and differentiating at a point where it is differentiable, which holds for F_i-almost all θ_i, yields (4) for $V_i^{\min}$, as desired.

Let $V_i : \Theta_i \to \mathbb{R}$ be any absolutely continuous function satisfying the lemma's assumptions, then by Lemma 5,

$$g_i(X_i^*(\theta_i)) \leq V_i'(\theta_i) \leq 1 - g_i(1 - X_i^*(\theta_i)),$$

for F_i-almost every θ_i. If $\hat{\theta}_i \leq \theta_i$,

$$V_i(\theta_i) \geq V_i(\hat{\theta}_i) + \int_{\hat{\theta}_i}^{\theta_i} g_i(X_i^*(s))\,ds \geq V_i^0(\hat{\theta}_i) + D_i(\theta_i, \hat{\theta}_i),$$

while if $\hat{\theta}_i > \theta_i$,

$$V_i(\theta_i) \geq V_i(\hat{\theta}_i) - \int_{\theta_i}^{\hat{\theta}_i} (1 - g_i(1 - X_i^*(s))\, ds \geq V_i^0(\hat{\theta}_i) + D_i(\theta_i, \hat{\theta}_i).$$

Hence, by taking the supremum, $V_i(\theta_i) \geq V_i^{\min}(\theta_i)$ for all $\theta_i \in \Theta_i$. □

By Lemma 6, we have $V_i(\theta_i) \geq V_i^{\min}(\theta_i)$ for all $\theta_i \in \Theta_i$, so that taking expectations $\mathbb{E}_{\tilde{\theta}_i}[V_i^{\min}(\tilde{\theta}_i)] \leq \mathbb{E}_{\tilde{\theta}_i}[V_i^M(\theta_i)] = \mathbb{E}_{\tilde{\theta}_i}[V_i(\theta_i)]$. Because M is EPEFF and FI, the accounting identity of Lemma 2, with $U_d^M = 0$ by EABB, yields

$$\sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^M(\tilde{\theta}_i)] = E_{\tilde{\theta}}[S^*(\tilde{\theta})].$$

Combining these, we get (6) as desired.

Conversely, suppose (6) holds. We define the mechanism $M = (x^*, t)$ by

$$t_i(\theta) = \theta_i x_i^*(\theta) - V_i^{\min}(\theta_i) - \kappa_i,$$

where $\kappa_i \geq 0$ are constants such that $\sum_{i \in I} \kappa_i = E_{\tilde{\theta}}[S^*(\tilde{\theta})] - \sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^{\min}(\theta_i)]$. The mechanism is EPEFF and FI by construction, so it is FB by Theorem 1. Adding a constant preserves the inequalities in Lemma 5, so it is IC. Moreover, $V_i^M(\theta_i) \geq V_i^{\min}(\theta_i) \geq V_i^0(\theta_i)$ for all $\theta_i \in \Theta_i$, so it is IR. We have,

$$U_d^M = E_{\tilde{\theta}}[S^*(\tilde{\theta})] - \sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^M(\theta_i)] = E_{\tilde{\theta}}[S^*(\tilde{\theta})] - \sum_{i \in I} \mathbb{E}_{\tilde{\theta}_i}[V_i^{\min}(\theta_i)] - \sum_{i \in I} \kappa_i = 0,$$

so that M is EABB.

Finally, every feasible expected certainty equivalent payoff profile must satisfy $U_i \geq \mathbb{E}_{\tilde{\theta}_i}[V_i^{\min}(\theta_i)]$, while EABB and FB require $\sum_{i \in I} U_i = E_{\tilde{\theta}}[S^*(\tilde{\theta})]$. Conversely, every payoff profile satisfying these conditions is implemented by choosing $\kappa_i = U_i - \mathbb{E}_{\tilde{\theta}_i}[V_i^{\min}(\theta_i)]$.

References

Ahn, D., M. Zhang, and M. Zhang (2026). Mechanism design beyond expected utility: Dynamic (in)consistency and full surplus extraction. Working paper.

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

Arrow, K. J. (1979). The property rights doctrine and demand revelation under incomplete information. In *Economics and Human Welfare*, pp. 23–39. Elsevier.

Arrow, K. J. and R. C. Lind (1972). Uncertainty and the evaluation of public investment decisions: Reply. *American Economic Review 62*(1).

Athey, S. and J. Levin (2001). Information and competition in us forest service timber auctions. *Journal of Political Economy 109*(2), 375–417.

Benkert, J.-M. (2025). Bilateral trade with loss-averse agents. *Economic Theory 79*(2), 519–560.

Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica: Journal of the Econometric Society*, 424–444.

Börgers, T. and P. Norman (2009). A note on budget balance under interim participation constraints: the case of independent types. *Economic Theory 39*(3), 477–489.

Chatterjee, K. and W. Samuelson (1983). Bargaining under incomplete information. *Operations Research 31*(5), 835–851.

Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, 17–33.

Cramton, P., R. Gibbons, and P. Klemperer (1987). Dissolving a partnership efficiently. *Econometrica: Journal of the Econometric Society*, 615–632.

d'Aspremont, C. and L.-A. Gérard-Varet (1979). Incentives and incomplete information. *Journal of Public Economics 11*(1), 25–45.

Di Giovan Paolo, A. and J. Higueras Corona (2026). Price impact of insurance. Working paper.

Dughmi, S. and Y. Peres (2012). Mechanisms for risk averse agents, without loss. Working paper.

Esö, P. and G. Futo (1999). Auction design with a risk averse seller. *Economics Letters 65*(1), 71–74.

Garratt, R. and M. Pycia (2026). Efficient bilateral trade. Working paper.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2022). Optimal auctions: Non-expected utility and constant risk aversion. *The Review of Economic Studies 89*(5), 2630–2662.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2023). Optimal insurance: Dual utility, random losses, and adverse selection. *American Economic Review 113*(10), 2581–2614.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2025). Optimal security design for risk-averse investors. *American Economic Review 115*(6), 2050–2092.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory 118*(2), 133–173.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988). Investment, capacity utilization, and the real business cycle. *American Economic Review*, 402–417.

Groves, T. (1973). Incentives in teams. *Econometrica: Journal of the Econometric Society*, 617–631.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, 667–686.

Holmström, B. and R. B. Myerson (1983). Efficient and durable decision rules with incomplete information. *Econometrica: Journal of the Econometric Society*, 1799–1819.

Jagadeesan, R., I. R. Segal, and A. Tordjman (2025). The limits of auctions under ex-ante collusion. Working paper.

Jouini, E., W. Schachermayer, and N. Touzi (2008). Optimal risk sharing for law invariant monetary utility functions. *Mathematical Finance: An International*

Journal of Mathematics, Statistics and Financial Economics 18(2), 269–292.
Kőszegi, B. and M. Rabin (2007). Reference-dependent risk attitudes. *American Economic Review 97*(4), 1047–1073.
Mailath, G. J. and A. Postlewaite (1990). Asymmetric information bargaining problems with many agents. *The Review of Economic Studies 57*(3), 351–367.
Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica: Journal of the Econometric Society 70*(2), 583–601.
Myerson, R. B. and M. A. Satterthwaite (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory 29*(2), 265–281.
Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*. Springer.
Rivera Mora, E. and P. Strack (2026). Information without rents: Mechanism design without expected utility. Working paper.
Safra, Z. and U. Segal (1998). Constant risk aversion. *Journal of Economic Theory 83*(1), 19–42.
Safronov, M. (2018). Coalition-proof full efficient implementation. *Journal of Economic Theory 177*, 659–677.
Segal, U. and A. Spivak (1990). First order versus second order risk aversion. *Journal of Economic Theory 51*(1), 111–125.
Vasserman, S. and M. Watt (2021). Risk aversion and auction design: Theoretical and empirical evidence. *International Journal of Industrial Organization 79*, 102758.
Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance 16*(1), 8–37.
von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
Wilson, R. (1968). The theory of syndicates. *Econometrica: Journal of the Econometric Society*, 119–132.
Wolitzky, A. (2016). Mechanism design with maxmin agents: Theory and an application to bilateral trade. *Theoretical Economics 11*(3), 971–1004.
Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, 95–115.