EL COLEGIO DE MÉXICO
CENTRO DE ESTUDIOS ECONÓMICOS

LICENCIATURA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL TÍTULO DE LICENCIADO EN ECONOMÍA

MULTIAGENT INCENTIVE DESIGN WITH SUPERMODULAR SIGNALS

ELEAZAR LOZOYA YAÑEZ

PROMOCIÓN 2020-2024

ASESOR:

DR. RENÉ LEAL VIZCAÍNO

CIUDAD DE MÉXICO, 2025

Agradecimientos

Ante todo, agradezco a mi hermana, Mar, por estar a mi lado todos los días —aún si ya no nos vemos—, por guiarme a ser quien soy y por enseñarme que las cosas que nos pasaron no nos definen. A mi madre, Eli, por recordarme las cosas que importan, y por confiar en mí a pesar de todos los vaivenes. A Miguel, por los brincos y las alegrías.

Agradezco haber tenido la fortuna de aprender y trabajar con los profesores de El Colegio. Gracias a mi asesor y mentor, René Leal, por la amistad y los consejos, por mostrarme los tecnicismos y la dedicación que requiere hacer esto, y por enseñarme que nada es trivial y que infinitos monos pegándole al teclado no van escribir *La Ilíada*. Ojalá que este trabajo sea "*kosher*." A Edwin Muñoz, por siempre ofrecer su ayuda y sus consejos, por su motivación, sus enseñanzas y su amistad. A Diana Terrazas, por su constante apoyo, por nunca dejar de mostrar interés en sus alumnos y, ante todo, por su calidad humana. A Raymundo Campos, por su apoyo y por siempre empujarme a ser un mejor economista.

Gracias a las personas que conocí en esta ciudad, con quienes he pasado los últimos cuatro años. A Fer, Sebas y Andrea, por todos los momentos juntos: las salidas, las cantadas y las aventuras (en el sillón). Gracias por recordarme que la amistad puede ser así. A Alexa, por darme todo el apoyo posible mientras escribía este trabajo. A los *roomies* (por orden de tamaño): Alo, por compartir las vivencias y por enderezarme siempre que lo necesito; Joss, por las grandes pláticas y por recordarme la humildad; Isaac, por acompañarme desde la inmadurez y por las experiencias juntos; Alonso, por la hermandad y por nunca dejarme solo.

Finalmente, gracias a las personas me acompañaron: Jorge, Luis Fer, Max, Vicky, Ian, Majo, Vale, Uriel, Inés, Pepe, Renata, Héctor, y —desde la distancia— Lynn, Frías, Fer y Gina. Por su amistad incondicional y por lo que aprendí de ustedes.

Abstract

Most of the existing literature related to incentive design for multiple agents assumes that, even for team-assigned tasks, the principal can observe an individual signal for each agent. Such assumption is unrealistic, and some of the results that follow (such as high-powered incentive schemes) are seldom observed in practice. I therefore study the optimal incentive scheme when a task performed jointly among multiple agents does not yield individual signals, but only a group-aggregated signal observable to the principal. The principal can also assign individual tasks to each agent, although the success of such individual project is lesser valued than the team-project's success. I specifically analyse how the assumption of supermodularity between the agents' actions for the signal function of the joint project can affect the optimal team contract, and derive sufficient conditions that, in addition to supermodularity, lead the principal to assign a joint project instead of individual tasks. Finally, I explore how these results vary in a repeated setting and find that repeated interactions can diminish the principal's cost of implementing effort.

Table of Contents

Chapter 1

Introduction: Incentives *and then some*

People respond to incentives.
– Mankiw (1998)

Economic interactions are plagued by incentives. In their fundamental contribution to the theory of incentives, Laffont and Martimort (2002) point out that "economics is to a large extent a matter of incentives: incentives to work hard, to produce good quality products, to study, to invest, to save, etc."

Indeed, one cannot help but wonder why economic thought largely ignored the matter of incentives for decades. Whilst notable economic thinkers such as Marschak (1955) and Arrow (1963) analyse notions inherently associated to incentives (*free-riding* and *moral hazard*, respectively), they vaguely address the problem of incentives. In fact, the former writes:

> *This raises the problem of incentives. Organization rules can be devised in such a way that, if every member pursues his own goal, the goal of the organization is served. This is exemplified in practice by bonuses to executives and ... in theory by the (idealized) model of the laisser-faire economy. And there exist, of course, also negative incentives (punishments). I shall have to leave the problem of incentives aside.*
>
> – Marschak (1955)

Incentives were introduced to the agenda of economists when they realized that arguably their most significant development, General Equilibrium Theory (GE), was seemingly incapable of embedding the problem of information asymmetries, namely *moral hazard* and *adverse selection*. The framework of GE appeared to be virtually all-powerful: it was able of dealing with notions such as externalities, uncertainty and time; and, more importantly, it was capable of validating the fundamental *invisible hand* result under competitive markets.

Accordingly, economists and mathematicians resorted to game theoretical notions to build a theory of incentives that would deal with the questions surrounding information asymmetries. These intentions were preceded by the work of Hurwicz (1960), who insisted on relaxing the assumption of competitive markets as the unique resource allocation mechanism, and rather to focus on designing all sorts of *mechanisms* in response to desirable optimality properties and compatibility with the postulated behaviour of agents. Hurwicz introduced the concept of incentive compatibility for mechanisms, and was then followed by Groves (1973), who called for incentive-compatible mechanisms for public policy motives.

Before the development of *mechanism design* theory, Vickrey (1945) identified the fundamental problem of optimal income taxation: since an individual's income depends greatly on their effort, some degree of inequality is needed to incentivize the efficient levels of effort required for production. Hence "the ideal distribution of income becomes a matter of compromise between equality and incentives." Vickrey then formulated the social planner's problem, a maximization of the sum of individual utilities under incentive compatibility and budget-balancing constrains.[1] Vickrey did not, however, present a solution for the problem, as he deemed it unsolvable.

Further, Mirrlees (1971), in what I consider to be one of the most prominent works in economic thought, re-formulated Vickrey's problem and employed the Pontryagin principle to solve it. While not explicitly naming it, Mirrlees applied what would be known as the Revelation Principle by framing the problem in terms of the taxpayers truthfully revealing their hidden information contingent on the tax system being optimally designed. Therefore, one may point out that the Mirrlees model formalized the principal-agent problem, in which a social planner (principal) must design a mechanism to extract private information from the taxpayers (agents) while maximizing social welfare.

Indeed, Mirrlees' work provided a much needed link between the mathematician-dominated

[1] Of course, Vickrey's formulation of incentive compatibility was rather distinct to that proposed by Hurwicz. Unknowingly to Vickrey at the time, his analysis would become fundamental for the developing of a theory of incentives.

mechanism design theory and the economics-related problem of asymmetric information. Economists hence proceeded to formalize the Revelation Principle —now widely seen as a fundamental result in mechanism design— through a game-theoretic understanding of economic interactions (Green and Laffont, 1977; Maskin et al., 1979; Myerson, 1979, 1982). Such result states that any mechanism that allocates resources in a society is equivalent to an incentive-compatible mechanism that does so while guaranteeing that agents will truthfully reveal their private information and obey the principal's recommended actions.

Incentives, economists found out, were fundamental to the understanding of economic interactions. Therefore, economic thought should study the optimal design of incentives in its attempt to keep up with the tall order that information asymmetries represented.

From mechanisms to contracts

Mechanism design theory suggests that economists should be concerned with designing institutions that provide desirable incentives for economic agents. Of course, many would point out that institutions, being the "rules of the game", are somewhat analogous to incentives. As a matter of fact, I am at one with such view.

Furthermore, one can restrict their attention to much more specific questions. For example, what are the optimal incentives if one were to regulate the interactions between agents within an organization? These affairs led to more precise answers for the role of information asymmetries in economic interactions. The theoretical framework concerned with answering such questions then became known as *contract theory*, since it attempts to design optimal incentive schemes —*contracts*— that ensure desirable outcomes for relatively smaller interactions when the agents face adverse selection and/or moral hazard concerns.

Accordingly, this thesis is merely an attempt to both study the intricacies of contract theory and also add new insights to this subject. Hence the work is essentially divided in two parts. The first part is covered in chapter 2, and it concerns the main results that arise from the literature on optimal contract design. Furthermore, my contribution to the contract theory literature is related to incentive design for multiple agents, and I shall deal with such analysis in chapter 3.

In this work, I provide a thorough study on incentive design through contract theory, specifically concerning the moral hazard problem. For chapter 2, I do a brief overview of the contract theory framework for a principal that faces moral hazard when contracting on a single agent. I consider both static and dynamic settings. The objective for such *JEL*-esque synthesis is twofold: (*i*) the main results for the general framework will turn out to be useful for the further analysis in chapter 3; and (*ii*) it should be able to familiarize the reader with the contract-theoretical apparatus that has become standard for the solution of the moral hazard problem, and may thus facilitate the comprehension of my main analysis in chapter 3.

Furthermore, for the third chapter I study the contracting problem for a principal that designs incentives for multiple agents. Besides designing an optimal incentive scheme, the principal is concerned with choosing between assigning an individual task for each agent (mirroring the model of chapter 2) or delegating a joint task in which the agents must cooperate to achieve a successful collective outcome.

Of my particular interest is the role of complementarity between the agents' actions —modeled through a supermodular signal function— on the principal's preference for the joint project over individual assignments. While I observe that complementarity is not by itself sufficient to guarantee that the principal is inclined towards assigning the team project, my main result provides an additional condition that, along with supermodularity, is sufficient for the preference of the joint task. Such condition, which I label as the *backpacking condition*, essentially states that an agent's marginal effect on the likelihood of the team's success is greater than such effect when the agent produces an individual outcome, even if every other agent working on the task is shirking. The term "back-packing" is thus a result of the agent's effort being so valuable that he is able to relatively "carry the team" towards success.

The analysis presented in chapter 3 is most closely related to the influential work of Che and Yoo (2001), who study a multiagent moral hazard model in which the the agents' actions are correlated, though every agent produces an individual signal that measures their performance. Their main result suggests that a repeated long-term relationship gives rise to "implicit incentives" that allow the principal to extract more rent. In turn, I assume that, when performing a joint task, the agents produce a unique outcome dependent on their collective actions. Additionally, while the Che-Yoo analysis presents a brief exploration of the optimal incentives for teams when the agents produce a single joint signal, it assumes that the principal has the option of inducing an agent to produce the team project's outcome single-handedly. Given that such assumption appears to be extremely unrealistic, I instead assume that the principal can assign an individual task to each agent if he were not to delegate the joint project.

Finally, I derive the optimal incentive schemes for a repeated setting of the model, and study how additional incentives arise from the repeated interaction between the agents when the principal assigns a joint task. Assuming that the agents can perfectly monitor their actions, I find that the long-term interaction makes the team project more attractive, since the principal can partially delegate the incentive scheme's role as a punishment mechanism to the agents themselves.

Chapter 2

Hidden Actions and the Moral Hazard Problem

2.1 Introduction

In this chapter, I present a brief overview of the classic moral hazard problem, also known as the hidden action model. The static framework is introduced first in order to familiarize the reader with the general formulation and the main results of the model. This section is mainly built on the work of Holmström (1979), Grossman and Hart (1983) and Laffont and Martimort (2002). Furthermore, I replicate the results for a two-period dynamic model —repeated moral hazard— following the work of Rogerson (1985) and Laffont and Martimort (2002).

The generic formulation of the model goes as follows. Consider two individuals who operate in the same environment. One of these players—the **principal**—wishes to compensate the other player—the **agent**—in exchange for a desirable action. Hereafter I may refer to the principal as *her*, and to the agent as *him*. The agent's action, which cannot be observed, is desirable for the principal as it affects some outcome that provides her utility. However, the action is costly for the agent (one might think exerting effort is unpleasant for him). Then, *what action does the agent choose?* And, more importantly, *how does the principal design an optimal payment scheme to incentivize the agent to exert her most desired action?* Such action is, naturally, that which maximizes the principal's value.

The principal meets two challenges when designing the optimal payment scheme. First, as noted in the earlier question, the action's cost for the agent makes it necessary for the principal to employ an incentive scheme to induce the desired performance. One may think, of course, that the optimal incentive scheme should reward the agent if he chooses the principal-preferred action. Nonetheless, the inability of the principal to observe the agent's effort leads her to use some other measure of performance to design the optimal incentives: the realized outcome, which depends on the agent's action.

The second challenge relates to the fact that the outcome captures an imperfect measure of the agent's performance, as it also depends on random external factors. This leads the principal to offer an incentive scheme that induces risk on the agent, as his "good" performance might be rewarded with a subpar pay if the realized outcome is not as that desired by the principal. Now, if the agent is risk-averse, the payment based on a noisy outcome leads to a trade-off between risk and incentives; that is, a trade-off between motivating and insuring the agent.

The complete and formal formulation of the model shall deal with such issues. Particularly, the solution for the model must prescribe an optimal incentive scheme designed by the principal that addresses the trade-off between risk and incentives, while also maximizing her value.

2.2 Static Moral Hazard

2.2.1 Setup

There are two players: a risk-neutral principal (*her*) and a risk-averse agent (*him*). The agent privately takes an action $a \in A$, where $A \subset \mathbb{R}$ is the set of all possible actions and it is finite (compact) and non-empty.[1] The action is costly for the agent; for a given a, he incurs the cost $c(a)$. An action a, together with some random external factors ε, determine an outcome $x = x(a, \varepsilon)$. The principal obtains value from x, which I assume is a discrete random variable that can take on n possible values: $x_i \in X \subseteq \mathbb{R}$ where $X = \{x_1, x_2, ..., x_n\}$ is the set of all possible outcomes, and $x_1 < x_2 < ... < x_n$.

Note that each $a \in A$ yields a distribution over outcomes x: for a fixed a, the distribution over ε induces a distribution over x. As suggested by Mirrlees (1999) and Holmström (2017), it is

[1] This assumption shall be used to guarantee the existence of a solution.

convenient to think that the agent chooses not an action, but the distribution over x. Let $\pi : A \rightarrow \Delta(X)$ be a function denoting the probability of any of the n possible outcomes realizing given that some action a was selected, such that $\pi(x|a) = (\pi(x_1|a), ..., \pi(x_n|a))$ is the distribution over x. I may henceforth write $\pi_i(a)$ instead of $\pi(x_i|a)$, for simplicity of notation. It is assumed that both the agent and the principal know the distribution $\pi(\cdot|a)$ for every possible $a \in A$.

Notice that the outcome space X, along with a distribution $\pi(x|a)$ of the family of distributions $\{\pi(x|a)\}_{a\in A}$, constitute an information structure for which the realized signal is an outcome x_i. This outcomes-as-signals interpretation will come in handy for the analysis presented in chapter 3.

Assumption 2.1. *Every distribution $\pi(a)$ in the family of distributions $\{\pi(a)\}_{a\in A}$ is bounded away from zero. That is, $\forall a \in A$ and $\forall x_i \in X$, $\pi(x_i|a) > 0$.*

Moreover, since the principal is risk-neutral, her utility is simply $x - w$. Because she obtains value from x, the principal intents to "hire" the agent to perform a costly action by offering a contract with duration of one period. A *contract*, or *incentive scheme*, is an n-tuple $\boldsymbol{w} = (w(x_1), w(x_2), ..., w(x_n))$ that prescribes an incentive contingent on the realization of each of the possible outcomes. I may henceforth write w_i instead of $w(x_i)$, and it is assumed that $w_i \in \mathbb{R} \;\; \forall x_i \in X$.

Let $U : \mathbb{R} \times A \rightarrow \mathbb{R}$ be the agent's von Neumann-Morgenstern utility. The following assumption is made.

Assumption 2.2. *The agent's von Neumann-Morgenstern utility is separable. That is, it can be written as $U(w, a) = u(w) - c(a)$, where $u : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous, strictly increasing ($u' > 0$) and strictly concave ($u'' < 0$) function over $\mathbb{R}$, and $\lim_{w\rightarrow -\infty} u(w) = -\infty$ (to rule out corner solutions); while c is a real-valued, continuous, strictly increasing ($c' > 0$) and strictly convex ($c'' > 0$) function over A. Also, let $h = u^{-1}$ denote the inverse function of u, such that h is strictly increasing ($h' > 0$) and strictly convex ($h'' > 0$).*

It is assumed that the principal knows the agent's utility function, his action set A and the distribution over outcomes $\pi(a)$. However, as noted in the generic formulation of the problem, she cannot observe a. Hence why the principal will offer a contract that rewards the agent according to the observed outcome, a realization of the random variable x. Given the offered incentive scheme $\boldsymbol{w}$, the agent will choose an action that maximizes his expected utility. The agent chooses an action a that satisfies:

$$a \in \arg\max_{a \in A} \sum_{i=1}^{n} \pi_i(a)[u(w_i) - c(a)] \tag{2.1}$$

One might think of this interaction as a game with the following timing:

1. The principal offers a contract $\boldsymbol{w} : X \to \mathbb{R}$. The incentive scheme $\boldsymbol{w}$ hence works as a function that maps each possible realization of x into a payment. One can always think of the contract as $\boldsymbol{w} = (w(x_1), w(x_2), ..., w(x_n))$.

2. The agent accepts or rejects the contract. If he rejects, the game is over. If he accepts, then he must privately choose an action $a \in A$. He chooses some a that satisfies (2.1).

Since the principal must commit to the contract $\boldsymbol{w}$, she must solve the previous game using backwards induction. That is, she must take into account the agent's problem and his election of a; then, she maximizes her own payoff. Nonetheless, one can further simplify the solution method of the game into a simple constrained maximization problem for the principal. This is done by including the agent's problem as a constraint in the principal's program. Note that a second constraint should be included, as the agent must first accept the contract before choosing his action.

Definition (Incentive compatibility). *Let $a^* \in A$ be the desired action for the principal. A contract $\boldsymbol{w}$ is said to satisfy* ***incentive compatibility*** *if the agent's expected utility under a^* is at least as high as his expected utility for any other action:*

$$\sum_{i=1}^{n} \pi_i(a^*)u(w_i) - c(a^*) \geq \sum_{i=1}^{n} \pi_i(a)u(w_i) - c(a) \quad \forall a \in A. \tag{IC}$$

Note that this is equivalent to the previously stated condition $a^* \in \arg\max_{a \in A} \sum_{i=1}^{n} \pi_i(a)[u(w_i) - c(a)]$.

Definition (Individual rationality). *Let $a^* \in A$ be the desired action for the principal. A contract $\boldsymbol{w}$ is said to satisfy* ***individual rationality*** *if the agent's expected utility under a^* is at least* *as high as his reservation utility (assumed to be zero):*

$$\sum_{i=1}^{n} \pi_i(a^*)u(w_i) - c(a^*) \geq 0. \tag{PC}$$

This is also called the ***participation constraint***.

Definition (Implementation). *A contract $\boldsymbol{w}$ is said to* ***implement*** *an action $a \in A$ if it satisfies both incentive compatibility and individual rationality for such a.*

Therefore, taking into account the previous constraints, the principal chooses an incentive scheme $\boldsymbol{w} = (w_1, w_2, ..., w_n)$ that solves the following constrained optimization problem (the **principal's problem**):

$$\max_{(w_1,...,w_n)} \quad \sum_{i=1}^{n} \pi_i(a^*)[x_i - w_i] \tag{2.2a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a^*)u(w_i) - c(a^*) \geq \sum_{i=1}^{n} \pi_i(a)u(w_i) - c(a) \quad \forall a \in A \tag{IC}$$

$$\sum_{i=1}^{n} \pi_i(a^*)u(w_i) - c(a^*) \geq 0 \tag{PC}$$

This program has a linear objective function and concave constraints, which may prove difficult to solve. Nonetheless, Grossman and Hart (1983) suggest a simple solution: replace the agent's payment w_i with the utility given by such payment, which can be obtained using the inverse utility function h (it can easily be seen that $h(u(w_i)) = w_i$). Hence one substitutes w_i with $h(v_i)$, where for easiness of notation I write v_i to represent the utility given by payment w_i, i.e. $v_i := u(w_i)$. The modified program yields a convex objective function (since h is convex), and linear constraints (since utility is given). The modified principal's problem is thus:

$$\min_{v_1,...,v_n} \quad \sum_{i=1}^{n} \pi_i(a^*)h(v_i) \tag{2.3a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a^*)v_i - c(a^*) \geq \sum_{i=1}^{n} \pi_i(a)v_i - c(a) \quad \forall a \in A \tag{IC}$$

$$\sum_{i=1}^{n} \pi_i(a^*)v_i - c(a^*) \geq 0 \tag{PC}$$

The principal's program highlights the "economics" of the game: the principal chooses an optimal incentive scheme that minimizes the cost of implementing her desired action. Such scheme should depend on (*i*) the agent actually preferring to choose the action desired by the principal, which is simply his optimization problem; and (*ii*) the agent's willingness to accept the contract given its offered incentives for each possible outcome.

Before highlighting some of the properties of the optimal contract, one may ask if a solution even exists. Even more, *what does the optimal contract actually mean?* Therefore, before proving the existence of such optimal incentive scheme, one must first define the second-best cost function and incentive scheme for the principal. I now proceed to give a formal and intuitive definition of these concepts.

Definition (Second-best cost function). *Consider any $a \in A$. Let $C_{SB}(a)$ denote the minimum expected cost at which the principal can induce the agent to choose such a. More formally, define:*

$$C_{SB}(a) = \begin{cases} \inf_{v_1,\dots,v_n} \sum_{i=1}^{n} \pi_i(a)h(v_i), & \text{if } \exists\, \boldsymbol{w} \text{ that implements } a \\ \infty, & \text{if } \nexists\, \boldsymbol{w} \text{ that implements } a \end{cases} \tag{2.4}$$

*Then $C_{SB} : A \to \mathbb{R} \cup \{\pm\infty\}$ is the **second-best cost function** for the principal.*

Note that the second-best cost function "finds" the least costly incentive scheme that the principal should offer the agent to implement some action $a \in A$, although not every action will be implementable. The second-best cost function will thus come in handy for the principal in her optimization problem.

Indeed, the incentive design problem, as studied by Grossman and Hart (1983), consists on two steps. First, the principal shall compute $C_{SB}(a)$ for each $a \in A$. That is, she will minimize her expected cost through $C_{SB}(\cdot)$. As a second step, taking into account the value of $C_{SB}(a)$ for each $a \in A$, the principal must choose which action to implement to maximize her value $\sum_{i=1}^{n} \pi_i(a)x_i -$

$C_{SB}(a)$. The second step, as Grossman-Hart point out, is not generally a convex problem, since $C_{SB}(a)$ need not be convex. The following definition formalizes the previous idea.

Definition (Second-best action and incentive scheme). *A* ***second-best optimal action*** a^* *is one which solves the second step of the incentive design problem. That is,* a^* *maximizes* $\sum_{i=1}^{n} \pi_i(a)x_i - C_{SB}(a)$*. A* ***second-best optimal incentive scheme*** $\boldsymbol{w^*} = (w_1^*, \ldots, w_n^*)$ *is one that solves the first step of the incentive design problem, namely that which implements the second-best optimal action* a^* *at the least expected cost:* $\sum_{i=1}^{n} \pi_i(a^*)w_i^* = C_{SB}(a^*)$*.*

The previous definition formalizes the notion of a solution characterization for the principal's program. The principal will set out to find a second-best optimal contract given the second-best action she has chosen. Now, the existence of such solution shall be proven.

Proposition 2.1 (Grossman & Hart, 1983). *Let Assumptions 2.1 and 2.2 hold. Then, there exists a second-best optimal action* $a^* \in A$ *and a second-best optimal contract* $w^* = (w_1^*, \ldots, w_n^*)$ *that implements it.*

Proof. The proof is made in two steps. First, I show that there exists an optimal incentive scheme $\boldsymbol{w}^*$ for some action $a \in A$, i.e. the program (2.3) has a solution. The second step is to show the existence of an action $a^* \in A$ that solves $\sup_{a \in A} \sum_{i=1}^{n} \pi_i(a)x_i - C_{SB}(a)$.

First step. For the first step, it suffices to show that the constraint set is non-empty for an implementable action $a \in A$.

Let CS denote the constraint set for the principal's problem. It is straightforward to see that $\sum_{i=1}^{n} \pi_i(a)v_i$ is bounded from below in CS. Since the variance of the v_i's tends to infinity while their mean is bounded from below (from Assumption 2.1), and h is convex and nonlinear, it follows from Bertsekas (1974) that unbounded sequences in CS make $\sum_{i=1}^{n} \pi_i(a)h(v_i) \to \infty$. Hence it has to be that the constraint set CS is (artificially) bounded from above. Since CS is closed, it follows from sequential compactness that CS is compact. The existence of v^* minimum then follows from Weierstrass' theorem.

Second step. For the second step, one must show that the second-best cost function $C_{SB}(a)$ attains a minimum. First, note that $C_{SB}(a)$ is lower semicontinuous: since A is finite, then any function defined on A is continuous and is thus lower semicontinous. Consider a minimizing sequence $a_n \in A$:

$$\lim_{n\to\infty} C_{SB}(a_n) = \inf_{a\in A} C_{SB}(a).$$

Since A is finite and thus compact, it follows that there exists a subsequence a_{n_k} that converges to some point $a^* \in A$. Moreover, by definition of lower semicontinuity,

$$lim_{n\to\infty} C_{SB}(a_n) \geq C_{SB}(a).$$

Hence,

$$C_{SB}(a^*) \leq \inf_{a\to a^*} C_{SB}(a) \leq \liminf_{k\to\infty} C_{SB}(a_{n_k}).$$

which implies that $C_{SB}(a^*) = \inf_{a\in A} C_{SB}(a)$. Thus, the second-best cost function $C_{SB}(a)$ attains its minimum at a^* and therefore $\sup_{a\in A} \sum_{i=1}^{n} \pi_i(a)x_i - C_{SB}(a)$ has a solution $a^* \in A$. ■

I now turn to the solution of the model in order to highlight some of its main results and properties.

2.2.2 Binary environment: Two actions and two outcomes

First, I describe the solution for the static model in a simple two-outcome and two-action setting, in order to clear up and highlight the main results of the game. It is thus assumed that a and x can only take two possible values: $a \in \{s, e\}$, i.e. the agent can either shirk ($a = s$) or exert effort ($a = e$); and $x \in \{x^L, x^H\}$, i.e. the outcome can be either low or high. The cost of effort is normalized for simplicity; that is, $c(a = s) = 0$ and $c(a = e) = c$. Since the outcome can only take on two values, the contract shall prescribe two possible compensations: $w(x^H)$ and $w(x^L)$. Let $\pi^e = Pr(x = x^H|a = e)$ and $\pi^s = Pr(x = x^H|a = s)$. I further assume that $\pi^e > \pi^s$, and denote $\Delta\pi = \pi^e - \pi^s > 0$.

A final and relevant assumption is made. Suppose the principal wishes to induce the agent into choosing $a = e$. That is, I will solve the first step of the Grossman-Hart analysis assuming that, in the second step, the principal would determine that inducing effort $a = e$ is optimal.

Complete information: First-Best Scenario

Assume the principal can perfectly observe the agents actions a. The solution for this full information model will serve as a benchmark and will represent the *first-best optimal contract* that the principal can achieve.

Since the principal can perfectly observe effort, she need not consider the incentive compatibility constraint. Intuitively, an agent's deviation from $a = e$ will be detected (from perfect observability) and punished through the incentive scheme $\boldsymbol{w}$. The **principal's problem** is thus:

$$\min_{v^L, v^H} \quad \pi^e h(v^H) + (1 - \pi^e) h(v^L) \tag{2.5a}$$

$$\text{s.t.} \quad \pi^e v^H + (1 - \pi^e) v^L - c \geq 0 \tag{PC-FB}$$

Let μ denote the multiplier associated to the participation constraint in a Lagrangian setting of the problem. Using the fact that $h'(\cdot) = \frac{1}{u'(\cdot)}$, the first-order conditions for the program are as follows.

$$\begin{aligned} (v^H): \quad & \pi^e h'(v^H) - \mu \pi^e = 0 \\ \iff \quad & \mu = \frac{1}{u'(w^H)} \end{aligned} \tag{2.6}$$

$$\begin{aligned} (w^L): \quad & (1 - \pi^e) h'(v^L) - \mu (1 - \pi^e) = 0 \\ \iff \quad & \mu = \frac{1}{u'(w^L)} \end{aligned} \tag{2.7}$$

Therefore, the optimization problem yields the condition $\mu = \frac{1}{u'(w^L)} = \frac{1}{u'(w^H)}$, which implies that $u'(w^L) = u'(w^H)$. Now, since u is strictly concave, it follows that $w^L = w^H = w$.

Claim. *When the principal has complete information (i.e. she can perfectly observe the agent's action), the agent is fully insured: the optimal first-best contract* $\boldsymbol{w}^{FB}$ *will prescribe the same incentive* w *independently of the realized outcome.*

Furthermore, one can find the form of the outcome-independent payment w. Note that $\mu = \frac{1}{u'(w^L)} =$

$\frac{1}{u'(w^H)} > 0$. That is, the participation constraint is binding. Then, write the PC as:

$$\pi^e v^H + (1-\pi^e)v^L - c = 0 \tag{2.8}$$

Since $u(w^L) = u(w^H) = u(w)$, or equivalently $v^L = v^H = v$, rewrite (2.8) as

$$\begin{aligned} &\pi^e v + (1-\pi^e)v - c = 0 \\ \iff \quad & v = c \\ \iff \quad & u(w^H) = u(w^L) = c \end{aligned} \tag{2.9}$$

Claim. *Under complete information, the optimal outcome-independent incentive offered to the agent is $w = h(c)$. Hence the first-best optimal contract is $\mathbf{w}^{FB} = (h(c), h(c))$.*

Second-Best Scenario

I assume once again that the principal cannot observe the agent's action. Hence the principal is now constrained by incentive compatibility. Then, the principal's program is:

$$\min_{v^L, v^H} \quad \pi^e h(v^H) + (1-\pi^e)h(v^L) \tag{2.10a}$$

$$\text{s.t.} \quad \pi^e v^H + (1-\pi^e)v^L - c \geq \pi^s v^H + (1-p^s)v^L \tag{IC-SB}$$

$$\pi^e v^H + (1-\pi^e)v^L - c \geq 0 \tag{PC-SB}$$

Consider a Lagrangian program for the previous problem, for which λ and μ denote the multipliers associated to the IC and PC, respectively. The first-order for such program yield:

$$\begin{aligned} (v^H): \quad & \pi^e h'(v^H) - \lambda(\pi^e - \pi^s) - \mu\pi^e = 0 \\ \iff \quad & \frac{1}{u'(w^H)} = \mu + \lambda\frac{\Delta\pi}{\pi^e} \end{aligned} \tag{2.11}$$

$$\begin{aligned} (v^L): \quad & (1-\pi^e)h'(v^L) + \lambda(\pi^e - \pi^s) - \mu(1-\pi^e) = 0 \\ \iff \quad & \frac{1}{u'(w^L)} = \mu - \lambda\frac{\Delta\pi}{1-\pi^e} \end{aligned} \tag{2.12}$$

Lemma 2.1. *The incentive compatibility and participation constraints are binding under the second-best optimal contract. That is $\lambda > 0$ and $\mu > 0$.*

Proof. See Appendix A1.

Intuitively, the fact that IC and PC are binding in the optimum suggests that the principal attempts to induce the agent to choose $a = e$ whilst offering the lowest payment up the limit. This mirrors the fact that the principal uses the second-best cost function $C_{SB}(a)$ to minimize the offered incentive scheme while still implementing the desired action a.

Now, I shall employ the previous statement to identify the structure of the optimal incentive scheme.

Proposition 2.2. *For the two-outcome and two-action model, the second-best optimal contract $\boldsymbol{w}^*$ that implements action $a = e$ is:*

$$\boldsymbol{w}^* = \left(\underbrace{h\left(c - \pi^e \frac{c}{\Delta\pi}\right)}_{w^L}, \underbrace{h\left(c + (1 - \pi^e)\frac{c}{\Delta\pi}\right)}_{w^H} \right).$$

Proof. See Appendix A2.

Furthermore, the structure of the previous optimal incentive scheme allows to compare the second-best scenario with the full information first-best benchmark.

Proposition 2.3. *The second-best optimal contract, $\boldsymbol{w}^*$, is inferior to the first-best optimal contract, $\boldsymbol{w}^{FB}$. Equivalently, the second-best cost that implements $a = e$ is higher than the first-best cost that does so.*

Proof. First, recall that with full information we have that $w^H = w^L = w = h(c)$. Hence the expected first-best cost for the principal is $\mathbb{E}(\boldsymbol{w}_{FB}) = h(c)$.

Now, under the second-best contract, the expected cost that implements action e for the principal is:

$$C_{SB}(e) = \mathbb{E}(\boldsymbol{w}^*) = \pi^e w^H + (1 - \pi^e) w^L \tag{2.13}$$

$$= \pi^e h\left(c + (1 - \pi^e)\frac{c}{\Delta\pi}\right) + (1 - \pi^e) h\left(c - \pi^e \frac{c}{\Delta\pi}\right) \tag{2.14}$$

Since h is strictly convex, apply Jensen's inequality to get:

$$\mathbb{E}(\boldsymbol{w}^*) = \pi^e h\left(c + (1 - \pi^e)\frac{c}{\Delta\pi}\right) + (1 - \pi^e) h\left(c - \pi^e \frac{c}{\Delta\pi}\right) \tag{2.15}$$

$$> h\left[\left((\pi^e(c + (1 - \pi^e)\frac{c}{\Delta\pi}\right) + (1 - \pi^e)\left(c - \pi^e \frac{c}{\Delta\pi}\right)\right] \tag{2.16}$$

$$= h(c) = \mathbb{E}(\boldsymbol{w}^{FB}). \tag{2.17}$$

Therefore, the expected cost for the principal is higher under the second-best incentive scheme than for the first-best contract. ■

COROLLARY 2.1. *Whenever the high outcome is realized, $x = x^H$, the second-best optimal contract $\boldsymbol{w}^*$ offers the agent a higher payment than the first-best contract $\boldsymbol{w}^{FB}$; however, it offers a lower payment when for the low outcome $x = x^L$.*

Proof. Since h is increasing in c, it is straightforward to see that:

$$w^H = h\left(c + (1 - \pi^e)\frac{c}{\Delta\pi}\right) > h(c) = \mathbb{E}(\boldsymbol{w}^{FB}),$$

$$w^L = h\left(c - \pi^e \frac{c}{\Delta\pi}\right) < h(c) = \mathbb{E}(\boldsymbol{w}^{FB}).$$

■

Intuitively, because the agent is risk-averse, the principal "compensates" the risk borne by the agent through an incentive that rewards the high outcome, i.e. $w^H > w^L$.

2.2.3 Full model solution

The binary model proved useful to highlight some of the main properties for the moral hazard framework. However, Corollary 2.1. leads to a relevant question: *does the second-best optimal* *contract always reward higher outcomes with higher incentives?* I turn back to the solution for the baseline model with a finite action space and n possible outcomes to answer this question. Specifically, it is of my interest to show the conditions under which the optimal incentive scheme is monotone in outcomes. Once again, let $x \in X = \{x_1, x_2, ..., x_n\}$ and w.l.o.g. I assume that $x_1 < x_2 < ... < x_n$.

Analogously to the two-action and two-outcome model in which I assumed that implementing effort was optimal, now suppose that the principal wishes to implement some $a \in A$. Recall that the principal's problem under this case is:

$$\min_{v_1, ..., v_n} \quad \sum_{i=1}^{n} \pi_i(a) h(v_i) \tag{2.18a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a) v_i - c(a) \geq \sum_{i=1}^{n} \pi_i(\hat{a}) v_i - c(\hat{a}) \quad \forall \hat{a} \in \mathcal{A} \tag{IC}$$

$$\sum_{i=1}^{n} \pi_i(a) v_i - c(a) \geq 0 \tag{PC}$$

Again, let $\lambda_{\hat{a}}$ and μ denote the multipliers associated to the IC and PC, respectively, in a Lagrangian setting of the problem.[2] The first-order condition for any interior v_i yields the following equality:

$$\frac{1}{u'(w_i)} = \mu + \sum_{\hat{a} \neq a} \lambda_{\hat{a}} \left(1 - \frac{\pi_i(\hat{a})}{\pi_i(a)}\right) \quad \forall i = 1, ..., n \tag{2.19}$$

The following definition will be useful for my purposes of characterizing the form of the contract.

Definition (Monotone likelihood ratio property). *The family of distributions $\{\pi(x|a)\}_{a \in A}$ is said to satisfy the* ***monotone likelihood ratio property*** *(MLRP) if for any pair $a, \hat{a} \in A$ such that $\hat{a} < a$, the ratio*

$$\frac{\pi(x_i|a)}{\pi(x_i|\hat{a})} \textit{ is non-decreasing in } x_i.$$

The intuition for the MLRP is straightforward. Consider two distributions $\pi(x|a)$ and $\pi(x|\hat{a})$ of the family of distributions $\{\pi(x|a)\}_{a \in A}$, where $a, \hat{a} \in A$, $a > \hat{a}$. If $\{\pi(x|a)\}_{a \in A}$ satisfies the MLRP,

[2] Note that there is one multiplier for each $\hat{a} \neq a$. The finiteness of A allows to do this. Without loss of generality, I work out the solution using some generic $\hat{a}$ and then present the optimality condition for every a.

then the higher the observed value for x_i, the more likely that it was drawn from distribution $\pi(x|a)$ rather than from $\pi(x|\hat{a})$. For example, consider the case in which there are two outcomes: high and low (x^H and x^L, respectively). Recall that choosing some action a is equivalent to choosing a distribution over outcomes $\pi(x|a)$). By MLRP, if the principal observes a realization of the high outcome x^H, she must think that it is more likely that the agent chose action a rather than action $\hat{a}$, since $a > \hat{a}$.

The reader may now believe that $\{\pi(x|a)\}_{a\in A}$ must satisfy MLRP to guarantee that the optimal contract is monotonic in outcomes. As the following statement shows, an additional condition is required for such result.

Proposition 2.4. *Let a^* be the second-best optimal action. Assume that the incentive compatibility constraint binds for a single action $\hat{a} \neq a^*$ (i.e. $\lambda_{\hat{a}} > 0$). If the family of distributions $\{\pi(x|a)\}_{a\in A}$ satisfies the monotone likelihood ratio property, then the second-best optimal contract $\boldsymbol{w}^* = (w^*(x_1), ..., w^*(x_n))$ is non-decreasing in x_i. That is, the incentive scheme offered by the principal is (weakly) increasing as the realized outcome increases: $w^*(x_1) \leq w^*(x_2) \leq ... \leq w^*(x_n)$.*

Proof. For clarity of exposition, I present a proof for the case in which $|A| = 2$.[3]

Consider the notation for the previously studied two-action case ($a \in \{s, e\}$). It can be seen that MLRP is equivalent to the ratio $\frac{\pi_i^e - \pi_i^s}{\pi_i^e}$ being non-decreasing in i. Also, it follows from Lemma 2.1. that incentive compatibility binds for $a = s$. One can see that the optimality condition under two possible actions (and n possible outcomes) is:

$$\frac{1}{u'(w_i)} = \mu + \lambda \frac{\pi_i^e - \pi_i^s}{\pi_i^e} \tag{2.20}$$

Without loss of generality, consider any pair $x_i, x_j \in X$ such that $x_i > x_j$. MLRP implies that:

$$\frac{\pi_i^e - \pi_i^s}{\pi_i^e} \geq \frac{\pi_j^e - \pi_j^s}{\pi_j^e}.$$

[3]For a detailed proof of the $|A| > 2$ case, see Grossman and Hart (1983).

Then, apply condition (2.20) to get the following inequality:

$$\frac{1}{u'(w_i)} \geq \frac{1}{u'(w_i)}.$$

Since u is concave, it follows that $w_i \geq w_j$. ■

The previous proposition, which may be seen as the main result of the model, guarantees that the incentive scheme is monotone in outcome levels. Intuitively, the principal will reward higher outcome realizations with higher incentives as long as her subjective probability for the realization of an outcome x_i satisfies some monotonicity criterion, namely MLRP. In other words, if the principal "believes" that higher outcomes are more likely to be drawn from higher actions, then she will reward such outcomes with a higher payment.[4]

2.3 Dynamic Model: Repeated Moral Hazard

2.3.1 Setup

Consider now the case in which the interaction between the principal and the agent is repeated over two periods, $t \in \{1, 2\}$. All the assumptions regarding the principal and agent's preferences are the same. The principal will still offer the contract before the beginning of their interaction; that is, before $t = 1$. Both players will be able to observe the realized outcome after the first period. Let x_i denote the outcome in $t = 1$, and x_j the outcome for $t = 2$. For simplicity, I assume that the agent's discount rate is $\beta = 1$, i.e. he values the future and the present the same.

In this case, the incentive scheme $\boldsymbol{w}$ will prescribe n contingent payments for each possible outcome in $t = 1$, and n^2 contingent payments for $t = 2$, since the payment in the second period is contingent on the realization of an outcome in $t = 1$ and then on $t = 2$. One therefore can say that a *dynamic contract*, or *dynamic incentive scheme*, is an $(n+n^2)$-tuple of the form $\boldsymbol{w} = (w_1, w_2)$, where $w_1 = (w_1(x_1), ..., w_1(x_n))$ is itself an n-tuple and $w_2 = (w_2(x_{11}), w_2(x_{12}), ..., w(x_{ij}), ..., w(x_{nn}))$ is an n^2- tuple. Here, x_{ij} denotes the history of realizations when the realizations are x_i in $t = 1$ and x_j in $t = 2$.

[4] It should be noted that the assumption of $\lambda_{\hat{a}} > 0$ for a single action is crucial for the monotonicity result to hold. This assumption is equivalent to what the literature has named the *spanning condition* (see Grossman and Hart, 1983).

Since the incentives for the second period are contingent on the outcome in $t = 1$, one might think that the principal can design an incentive scheme that uses the second-period incentives to reward (or punish) the agent for his first-period performance. As a matter of fact, the memory property exhibited by the dynamic contract is the main result of the repeated moral hazard model. I will now set out to prove such property.

In similar fashion to the static model, one can see the repeated interaction between principal and agent as a game in which the timing goes as follows:

1. Before $t = 1$, The principal offers a dynamic contract $\boldsymbol{w} : X \times X^2 \to \mathbb{R}$ (I abuse of notation to remark the fact that the contract offers n^2 payments for the second period).

2. In $t = 1$, the agent either accepts or rejects the contract. If the contract does not satisfy individual rationality, the agent rejects and the game is over. If it does satisfy individual rationality, he accepts and then must privately choose an action $a_1 \in A$ for the first period. He chooses some a_1 that satisfies an intertemporal incentive compatibility constraint.

3. After $t = 1$ and before $t = 2$, both the principal and agent observe the realized outcome in the first period, namely x_i.

4. Since he has pre-commited to the contract, in $t = 2$ the agent must choose an action $a_2 \in A$ that was "recommended" by the principal. Once again, the implemented a_2 must satisfy intertemporal incentive compatibility.

As the reader might have noticed, a strong assumption is made: once the agent accepts the contract and commits to the two-period relationship with the principal, he must exert the actions a_1 and a_2 that the principal implemented through the scheme. One may think that there is a court of law that enforces the contract on the agent, or some sort of sense of loyalty by the agent that constrains him to fulfill his part of the arrangement.[5]

Therefore, the principal is constrained by intertemporal incentive compatibility and individual rationality constraints. Mirroring the one-period program, the model can be reduced to a single constrained optimization program for the principal, whose solution is the optimal contract. Let a_1

[5] Interestingly enough, if one were to relax the assumption and impose an additional second-period incentive compatibility constraint on the principal, the optimal dynamic contract would yield similar results to those I shall present here (see Lambert, 1983). Hence the assumption is simply made for the sake of easiness.

and a_2 be the desired actions for the principal in the first and second periods, respectively. The principal's two-period problem is:

$$\min_{v_1, v_2} \quad \sum_{i=1}^{n} \pi_i(a_1) h(v_i) + \sum_{j=1}^{n} \pi_j(a_2) h(v_j) \tag{2.21a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a_1)[v_i - c(a_1)] + \sum_{j=1}^{n} \pi_j(a_2)[v_j - c(a_2)]$$

$$\geq \sum_{i=1}^{n} \pi_i(\hat{a}_1)[v_i - c(\hat{a}_1)] + \sum_{j=1}^{n} \pi_j(\hat{a}_2)[v_j - c(\hat{a}_2)] \quad \forall \hat{a}_1, \hat{a}_2 \in A \tag{IC-2}$$

$$\sum_{i=1}^{n} \pi_i(a_1)[v_i - c(a_1)] + \sum_{j=1}^{n} \pi_j(a_2)[v_j - c(a_2)] \geq 0 \tag{PC-2}$$

2.3.2 Solution

Let $\lambda_{\hat{a}}$ and μ be the multipliers for the intertemporal IC and PC, respectively, in a Lagrangian setting of the program. The first-order conditions with respect to v_1 and v_2 are the following:

$$(v_1): \quad \pi_i h'(v_i) - \lambda_{\hat{a}}[\pi_i(a_1) - \pi_i(\hat{a}_1)] - \mu \pi_i(a_1) = 0 \tag{2.22}$$

$$\iff \quad \frac{1}{u'(w_1(x_i))} = \mu + \lambda_{\hat{a}} \left(1 - \frac{\pi_i(\hat{a}_1)}{\pi_i(a_1)}\right) \tag{2.23}$$

$$(v_2): \quad \pi_i(a_1)\pi_j(a_2) - \lambda_{\hat{a}}[\pi_i(a_1)\pi_j(a_2) - \pi_i(\hat{a}_1)\pi_j(\hat{a}_2)] - \mu \pi_i(a_1)\pi_j(a_2) = 0 \tag{2.24}$$

$$\iff \quad \frac{1}{u'(w_2(x_{ij}))} = \mu + \lambda_{\hat{a}} \left(1 - \frac{\pi_i(\hat{a}_1)\pi_j(\hat{a}_2)}{\pi_i(a_1)\pi_j(a_2)}\right) \tag{2.25}$$

The previous conditions for the optimal dynamic incentive scheme lead to the following statement.

Proposition 2.5. *The optimal dynamic incentive scheme $w^* = (w_1^*, w_2^*)$ must satisfy the following condition:*

$$\frac{1}{u'(w_1(x_i))} = \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))} \middle| x_i\right). \tag{MGL}$$

Hence, under the optimal dynamic contract, the inverse of the agent's marginal utility is a martingale.

Proof. See Appendix A3.

Recall that the inverse utility function h can be interpreted as the cost for the principal to implement an action a. Therefore, the previous proposition can be interpreted as the principal smoothing out incentives: she chooses a dynamic incentive scheme $\boldsymbol{w}^*$ that equates her expected marginal costs across periods. As I now prove, a corollary of the previous statement is that the dynamic contract has memory, or is *history-dependent*.

Corollary 2.2 (Rogerson, 1985). *Let* $\boldsymbol{w}^* = (w_1^*, w_2^*)$ *be the optimal dynamic incentive scheme. Suppose that* $w_1^*(x_i) \neq w_1^*(x_j)$*. Then* $\exists\, x_k \in X$ *such that* $w_2^*(x_{ik}) \neq w_2^*(x_{jk})$*. That is, the optimal dynamic incentive scheme* $w^* = (w_1^*, w_2^*)$ *is history-dependent.*

The intuition for history-dependence is simple. The contract having the memory property reflects the fact that if an outcome plays any role determining current payments (w_1), then it must necessarily determine future payments (w_2).

2.3.3 Form of the optimal dynamic contract

Notice that one can extend the two-period relation into any number of periods, such that the dynamic contract still satisfies equation (MGL). Moreover, it is possible to infer how the contract will behave over time by assuming the functional form of u. In particular for these purposes, of one's interest is the convexity/concavity of the function $h'(\cdot) = \frac{1}{u'(\cdot)}$.

Proposition 2.6. *If* $h'(\cdot) = \frac{1}{u'(\cdot)}$ *is convex [concave], the optimal dynamic contract* $\boldsymbol{w}^*$ *is* ***front-loaded*** *[****back-loaded****]. That is, the expected incentives for the agent are decreasing [increasing] over time:*

$$w_1 \geq \mathbb{E}(w_2|x_i) \geq \ldots \geq \mathbb{E}(w_T|x),$$

$$or \;\; w_1 \leq \mathbb{E}(w_2|x_i) \leq \ldots \leq \mathbb{E}(w_T|x).$$

Proof. I prove the case in which $h'(\cdot) = \frac{1}{u'(\cdot)}$ is convex (the concave case is analogous). For the proof, I make use of the martingale property exhibited by the inverse marginal utility, and given by equation (MGL):

$$\frac{1}{u'(w_1(x_i))} = \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) \tag{MGL}$$

Assume $\frac{1}{u'(w)}$ is a convex function. Apply Jensen's inequality to see that

$$\frac{1}{u'(\mathbb{E}(w))} \leq \mathbb{E}\left(\frac{1}{u'(w)}\right)$$

In particular, for $w = w_2(x_{ij})$ this is

$$\frac{1}{u'(\mathbb{E}(w_2(x_{ij}))|x_i)} \leq \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij})}\middle| x_i\right)$$

Using eq. (MGL), this is:

$$\frac{1}{u'\left(\mathbb{E}(w_2(x_{ij}))|x_i\right)} \leq \frac{1}{u'\left(w_1(x_i)\right)}$$

$$\iff \quad u'(w_1(x_i)) \leq u'\left(\mathbb{E}(w_2(x_{ij}))|x_i\right)$$

Since u' is decreasing (u is concave), we get the following inequality:

$$w_1(x_i) \geq \mathbb{E}[w_2(x_{ij}))|x_i].$$

■

One can extend the previous inequality for any two periods t and $t + 1$.

2.4 Final comments

It is relevant to mention that the previous analysis does not contain any innovative results. Rather, it provides a rigorous overview of the main results that the contract theory literature has provided on the moral hazard problem. These results shall familiarize the reader with the solution methods and notation employed for the literature on incentive design, and should therefore come in handy for the analysis that follows on chapter 3.

Chapter 3

Moral Hazard, Multiple Agents, and Supermodular Signals

3.1 Introduction

Why are agents incentived to act as a group? From workplace-related tasks to competitive activities, agents find themselves induced to work as a team consistently. For many scenarios, agents are driven to perform jointly due to their fondness for some shared reward. Nevertheless, if an incentive designer were to choose between providing individual payments or giving out shared incentives, why would he lean towards the latter?

Naturally, one may think that joint projects are convenient for a principal whenever there exists some sort of complementarity between the agents that constitute the team. Perhaps the synergy exhibited by the members of the group will enlarge the likelihood of attaining the shared reward. On the other hand, a *free-riding* agent in a joint task could potentially shirk, while the rest of the team suffers from a lesser aggregated effort in their attempt to successfully achieve an assigned task.

Indeed, some of the existing literature concerning moral hazard with multiple agents has focused on studying the conditions for which the incentive schemes provided to the agents are distinct. The early literature on multiagent contracts builds on results for tournaments (Lazear and Rosen, 1981; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983) to emphasize that optimal incentive schemes should promote competition between the agents through a relative performance evaluation, provided that the agents within the team are assigned an individual task and their performance is positively correlated. These analyses, however, often find that the optimal incentives are high-powered, which is rarely seen in practice.

Further studies have drawn attention to the role of cooperation between the agents. The prominent studies of Holmström and Milgrom (1990) and Itoh (1991, 1993) show that, by coordinating the agents' actions, the principal can often secure a higher payoff. In such case, the principal implements an incentive scheme that rewards the agents for a "good" performance of other project members. Such optimal contract therefore consists on shared incentives that are contingent on the collective actions by the agents.

In this chapter, I study a moral hazard problem in which a principal faces multiple agents. The principal can either assign an individual task, for which an individual incentive scheme is provided for each agent, or delegate a joint project whose contract is collective (i.e. every agent receives the same payment) and depends on a team-generated outcome. I am mainly concerned with providing the conditions for which the principal prefers to assign the team project instead of offering to each agent the individual incentive scheme studied in chapter 2.

Of my particular interest is the role of strategic complementarity between the agents' actions when they are assigned the joint task. I model such complementarity through supermodularity of the joint outcome's signal function, which essentially indicates that the marginal effect of an agent's action on the likelihood of success is higher whenever the rest of the team is putting on a greater effort. I show that such condition is not sufficient to guarantee that the principal prefers the joint project. Nonetheless, I prove that supermodularity of the signal function with an additional assumption, which I call the *backpacking condition*, suffices to ensure that the principal assigns the team task. Intuitively, the backpacking condition states that an agent's marginal effect on the likelihood of the team's success —even when every other agent is shirking— is greater than such effect when the agent produces an individual outcome. Hence the term "backpacking": the agent's effort in a team is so valuable that he is more able to "carry the team" towards success (though the rest of the agents are shirking) than he is able of achieving individual success.

Moreover, I derive the optimal incentive schemes for a repeated setting of the model, and study how additional incentives arise from the repeated interaction between the agents when the principal assigns a joint task. Assuming that the agents can perfectly monitor their actions, I find that this long-term interaction makes the team project more attractive, since the principal can partially delegate the incentive scheme's role as a punishment mechanism to the agents themselves.

For the analysis, I assume risk-neutrality from both the principal and the agents. This assumption makes the model differ from the one studied in the previous chapter, in which I assume that the agent is risk-averse. While risk-aversion implies a trade-off between insurance and incentives, I instead impose a limited liability restriction on the offered payments. This constraint limits the extent to which the principal can punish a subpar performance by the agents, and thus inducing higher actions requires giving out a "limited liability rent". Hence the trade-off is no longer between insurance and incentives, but between rent extraction and incentives. The complexities of limited liability notwithstanding, risk-neutrality greatly simplifies the inference of the optimal incentive schemes for both the joint and individual tasks.

This analysis is most closely related to the influential work of Che and Yoo (2001). Such paper presents a multiagent moral hazard model in which the the agents' actions are correlated, though every agent produces an individual signal for their performance. Its main result suggests that a repeated long-term relationship gives rise to "implicit incentives" that allow the principal to extract more rent. The present paper mainly differs on the assumption of individual signals; instead, I assume that, when performing a joint task, the agents produce a unique outcome dependent on their collective actions. Moreover, I am mainly concerned with the role of supermodularity to ensure the predominance of the joint task, while the Che-Yoo analysis simply derives the optimal contracts under a static and repeated setting. Finally, while the Che-Yoo analysis presents a brief exploration of the optimal incentives for teams when the agents produce a single joint signal, it assumes that the principal has the option of inducing an agent to produce the team project's outcome single-handedly. Of course, such assumption appears to be extremely unrealistic, and I instead assume that the principal can assign an individual task to each agent if he were not to delegate the joint project.

This chapter is structured as follows. In Section 3.2, I provide an overview of the literature most closely related to this work. Section 3.3 then presents the model and introduces some definitions concerning supermodularity. Then, in Section 3.4 I show the main results of the model, and discuss how the joint task contract defines a normal-form game. Section 3.5 analyses the optimal contracts in a repeated setting. Finally, I offer a brief conclusion and some thoughts on possible extensions in section 3.6.

3.2 Related Literature

The first formal analysis for incentives in teams is due to Marschak and Radner (1972) and Groves (1973), who study free-riding behaviour in teams and formulate a primitive structure of optimal incentive schemes for groups, with the goal of ensuring that the agents coordinate towards a common objective.

Specifically for the "contractual" approach, the literature for multiagent moral hazard goes back to Holmström (1982), whose *non-budget-balancing* result shows that there does not exist a contract that can allocate a team's outcome to every member of the project while implementing the optimal effort. Moreover, Holmström suggests that the principal can achieve an outcome close to the full-information benchmark through an incentive scheme that arbitrarily-severely punishes the agents if the realized joint outcome is relatively low.

Other early contributions to this literature (Lazear and Rosen, 1981; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983) consider risk-averse agents who produce individual signals and draw attention to the role of correlated performances for the design of an optimal incentive scheme. This *rank-order tournament* literature suggests that the principal should employ relative performance evaluation (i.e. ranking the agents by their performance relative to their peers) for the optimal contract.

Moreover, Mookherjee (1984) studies the multiple agent moral hazard problem in a model analogous to the seminal single-agent analysis by Grossman and Hart (1983). He finds that rank-order contracts suffer from vulnerability to collusion among agents and are prone to multiple equilibria arising. Ma et al. (1988) further explore the issue of multiple equilibria and, assuming the agents can monitor each other, propose a subtle yet costless mechanism that gets rid of the undesirable equilibria.

The work by Holmström and Milgrom (1990) suggests that, for a principal that faces agents who work on independent projects, a joint incentive scheme leads the agents to monitor each other. Therefore, the principal can rely on team incentives to a certain extent. In a similar fashion, Itoh (1991, 1993) considers agents that produce individual outcomes and shows that a collective contract that rewards the whole team for the success of an agent's project can motivate the agents to "help each other out". Itoh's analysis, however, relies on a rather strong assumption: an agent is motivated to help if his marginal disutility of effort is null.

Furthermore, Fleckinger (2012) studies a model with multiple agents who produce individual signals and argues that the interdependency between agents is threefold, as the agents are potentially correlated through technology, information, and strategies. He then shows that incentive schemes should exhibit an "optimal mix" of competitive and joint incentives. More recently, Halac et al. (2021) study a complex model in which agents are ranked according to their performance, though their incentives are contingent on team success. Interestingly enough, the principal in this model has an additional role as an information designer, since she can induce beliefs on the agent rankings.[1]

The possibility of free-riding arising in the repeated interaction between multiple agents has been mainly studied in principal-less scenarios for which incentives emerge merely due to the agents' strategic interactions (see Matsushima, 1989; Fudenberg et al., 1994). As mentioned previously, Che and Yoo (2001) do consider the repeated interaction between a principal and multiple agents, and show that the long-term relationship can lure the principal into providing joint incentives.

Finally, the foundations for lattice theory and supermodularity are mostly due to the seminal contributions of Topkis (1978), Vives (1990), Milgrom and Roberts (1990), and Milgrom and Shannon (1994). Further, Holmström and Milgrom (1994) first consider supermodularity in a contracting problem with multiple agents, albeit their analysis strongly differs from the one here presented.

3.3 Setup

Consider a risk-neutral principal (*her*) who faces n risk-neutral agents, indexed by $i \in N$, where $N = \{1, \ldots, n\}$ is the set of all agents. Agents are assumed to be symmetric, and each agent (*him*) possesses the ability to perform an action desirable by the principal. The agent i can thus privately take an action $a_i \in A_i$, where $A_i = [\underline{a}, \overline{a}] \subset \mathbb{R}$ is a finite set with all of his possible actions. As in the classic moral hazard model by Holmström (1979), each action may be interpreted as the level of effort that i exerts in behalf of the principal, and which the principal cannot observe. Performing

[1] See Fleckinger et al. (2024) for an excellent survey on the theory of incentive design for multiple agents.

an action is costly to each agent i. Let $c(a_i)$ denote the cost for the agent as a result of performing action $a_i \in A_i$.

Define $A_{-i} = \bigtimes_{j \neq i} A_j$ the set of all possible combinations of actions executed by agents other than i, and $A = A_i \times A_{-i}$ such set also including i's action. Then $\boldsymbol{a}_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$ is an $(n-1)$-sized vector which indicates the action chosen by each agent excluding i, whilst $\boldsymbol{a} = (a_1, .., a_n)$ is an action profile with the chosen action by every agent. Naturally, $\boldsymbol{a}_{-i} \in A_{-i} \subset \mathbb{R}^{n-1}$ and $\boldsymbol{a} \in A \subset \mathbb{R}^n$. The agents' actions, together with some external random shock, will determine an outcome realization that is valued by the principal. As I will now explain, it is assumed that the principal is endowed with a pair of information structures whose realizations are in fact the desired outcomes. However, the principal can rely only on one of these signals to infer something about the agents' performance; that is, she has to choose which signal to condition the contract on.

The principal can observe a signal x from an information structure $\Xi_x = \langle X, \pi(\cdot|a_i) \rangle$, where X is the *signal space* (the set of possible realizations of x) and $\pi : A_i \rightarrow [0, 1]$ is the *signal function* for x. That is, $\pi(x|a_i)$ states the probability of outcome x realizing given that action a_i was selected. The family of distributions $\{\pi(\cdot|a_i)\}_{a_i \in A}$ then contains a possibly different distribution $\pi(\cdot|a_i)$ for each of i's actions. Note that the signal function $\pi(\cdot)$ depends exclusively on a single action. In fact, a realization of the signal x is called an *individual outcome*,[2] since I assume that x can be used exclusively as a measure of individual performance. If the principal assigns the individual task to agent i, then outcome x is determined by a_i and some random external factor ε. For simplicity, I will assume that x may only take on two possible realizations: success (s) or failure (f). Hence the signal space is $X = \{f, s\}$.

On the other hand, the principal can also observe a signal y from the information structure $\Xi_y = \langle Y, \sigma(\cdot|a_i, \boldsymbol{a}_{-i}) \rangle$. Correspondingly, Y and $\sigma(\cdot|a_i, \boldsymbol{a}_{-i})$ are the *signal space* and *signal function* for y, respectively. In this case, the signal function $\sigma : A_i \times A_{-i} \rightarrow [0, 1]$ states the probability of a realization y conditional on two elements: a_i and $\boldsymbol{a}_{-i}$. Indeed, a realization of signal y is a *joint outcome*, and it measures the performance of a task run by multiple agents. The signal y can be employed by the principal whenever she contracts on multiple agents to jointly carry on a project. Similarly to the individual signal, I will assume that $Y = \{F, S\}$, i.e. the joint outcome can either

[2] One can see that the individual outcome is simply the one studied in chapter 2, in which the principal designs incentives for a single agent.

be failure (F) or success (S).

Therefore, the principal can either draw a contract with each agent individually and assign him the task of producing some x, or she can assign a team project run by multiple agents, for which she employs the signal y to measure their performance. The principal's problem is thus not only that of designing an optimal incentive scheme, but also choosing the information structure on which she will draw the contract.[3] *How does the principal make such decision?* As it will be seen, she designs the optimal contract —namely, the minimum cost of providing incentives— for both the individual and team contributions. Then, she will simply compare her benefits when implementing individual tasks (using x) to those when running a team project (employing y).

Since the principal and the n agents are risk-neutral, their utility functions are as follows. The principal's expected utility is $V(z, w) = \mathbb{E}(z - w)$, where z could be the either x or y, depending on the signal on which the principal draws the contract. Agent i's von Neumann-Morgenstern utility is $U(w, a_i) = w - c(a_i)$, where $c(\cdot)$ is real-valued and strictly increasing in a_i.

Some further assumptions are made. Naturally, both the principal and the n agents know the signal functions $\pi(\cdot|a_i)$ and $\sigma(\cdot|a_i, \boldsymbol{a_{-i}})$. The success of the joint project is strictly more valuable to the principal than the the individual outcome's success: $S > s$. The extent to which S is larger, however, will be part of a later analysis. Moreover, consider any two actions $a, a' \in A_i$, where $a' < a$. Define $\boldsymbol{a}'_{-i} = (a', \ldots, a')$ as the $n-1$ vector in which every agent except for i choose a'. It is then assumed that $\pi(s|a) > \pi(s|a')$, while $\sigma(S|a, \boldsymbol{a}_{-i}) > \sigma(S|a', \boldsymbol{a}_{-i})$ and $\sigma(S|a, \boldsymbol{a'_{-i}}) > \sigma(S|a', \boldsymbol{a'_{-i}})$. That is, the probability of success for both tasks is greater when agent i increases his effort; for the joint task, such property holds notwithstanding the other agents' actions.

Finally, I assume that the individual signal functions are conditionally independent, i.e. $\pi(x|a_i, a_j) = \pi(x|a_i)$. Hence an agent j's action shall not affect the probability of an outcome realization for i's individual task. Although this assumption was implicitly made before, it is expressed for purposes of clarity.

[3]Notice that an implicit assumption is being made: the principal does not necessarily deem any of the information structures more informative in the sense of Blackwell. If she did so, one may wonder why would she even consider employing the Blackwell-dominated signal for the contract. Therefore, her comparison between tasks simply relies on her expected gains from using the signals, assuming she deems both of these "sufficient" in the sense of Holmström (1979).

3.3.1 Lattices and supermodularity: Some definitions

It is of this paper's interest to study why the principal could be leaning towards assigning the joint project over n individual tasks. Aside from the greater value that the team project's success provides, one might think that some sort of complementarity between the agents actions arises when they work together on the joint task. With that in mind, it is first necessary to formalize such idea of complementarity for the later analysis. The following definitions will do so. Throughout, I will refer to the terms increasing (decreasing) and higher (lower) in the weak sense.

Definition (Lattice). *A partially ordered set L is said to be a* ***lattice*** *if for any two elements $l_1,\ l_2 \in L$,*

$$l_1 \vee l_2 := \inf\{k \in L | k \geq l_1, k \geq l_2\} \in L$$
$$l_1 \wedge l_2 := \sup\{k \in L | k \leq l_1, k \leq l_2\} \in L.$$

The operators $\vee$ and $\wedge$ are called the join and meet, respectively. Therefore, a lattice is a poset with a specific structure that allows any two elements of the set to have both a greatest lower bound (infimum) and a least upper bound (supremum).

Definition (Sublattice). *Consider a lattice L. A poset $M \subseteq L$ is said to be a* ***sublattice*** *if for any two elements $m_1,\ m_2 \in M$, $m_1 \vee m_2 \in M$ and $m_1 \wedge m_2 \in M$.*

Claim. *$\mathbb{R}^n$, endowed with the component-wise order, is a lattice. The space $A_i \times A_{-i}$ is a sublattice.*

Proof. Consider any two vectors $\alpha, \beta \in \mathbb{R}^n$. The component-wise order states that: $(\alpha_1, \ldots, \alpha_n) \geq (\beta_1, \ldots, \beta_n) \iff \alpha_i \geq \beta_i\ \forall i$. It can easily be seen that $\mathbb{R}^n$ is a lattice where:

$$\alpha \wedge \beta = (\max\{\alpha_1, \beta_1\}, ..., \max\{\alpha_n, \beta_n\}) \in \mathbb{R}^n$$
$$\alpha \vee \beta = (\min\{\alpha_1, \beta_1\}, ..., \min\{\alpha_n, \beta_n\}) \in \mathbb{R}^n.$$

It follows directly that $A_i \times A_{-i} \subset \mathbb{R}^n$ is a sublattice. Specifically, the join and meet for $A_i \times A_{-i}$

sublattice are:

$$\boldsymbol{a} \wedge \boldsymbol{a'} = (\max\{a_1, a'_1\}, ..., \max\{a_n, a'_n\}) \in \mathbb{R}^n$$
$$\boldsymbol{a} \vee \boldsymbol{a'} = (\min\{a_1, a'_1\}, ..., \min\{a_n, a'_n\}) \in \mathbb{R}^n.$$

■

Lattice theory allows to formalize the idea of complementarity in terms of functions. The following definition establishes such notion of complementarity and lays the foundations for the later analysis.

Definition (Supermodular function). *Let L be a lattice. A function $f : L \longrightarrow \mathbb{R}$ is said to be **supermodular** if for any $l_1, l_2 \in L$,*

$$f(l_1 \wedge l_2) + f(l_1 \vee l_2) \geq f(l_1) + f(l_2).$$

In particular for the setup of this model, supermodularity goes as follows. Since $A_i \times A_{-i}$ sublattice can be ordered component-wise, supermodularity captures the notion of complementarity between A_i and A_{-i}. Consider two complete action profiles $\boldsymbol{a} = (a_i, \boldsymbol{a}_{-i})$ and $\boldsymbol{a'} = (a'_i, \boldsymbol{a'}_{-i})$, with $a'_i \geq a_i$ and $\boldsymbol{a}_{-i} \geq \boldsymbol{a'}_{-i}$. Then, the *joint signal function for success S, $\sigma(S|\cdot)$, is **supermodular*** in $(a_i, \boldsymbol{a}_{-i})$ if:

$$\sigma(S|a'_i, \boldsymbol{a}_{-i}) + \sigma(S|a_i, \boldsymbol{a'}_{-i}) \geq \sigma(S|a_i, \boldsymbol{a}_{-i}) + \sigma(S|a'_i, \boldsymbol{a'}_{-i}). \tag{SM}$$

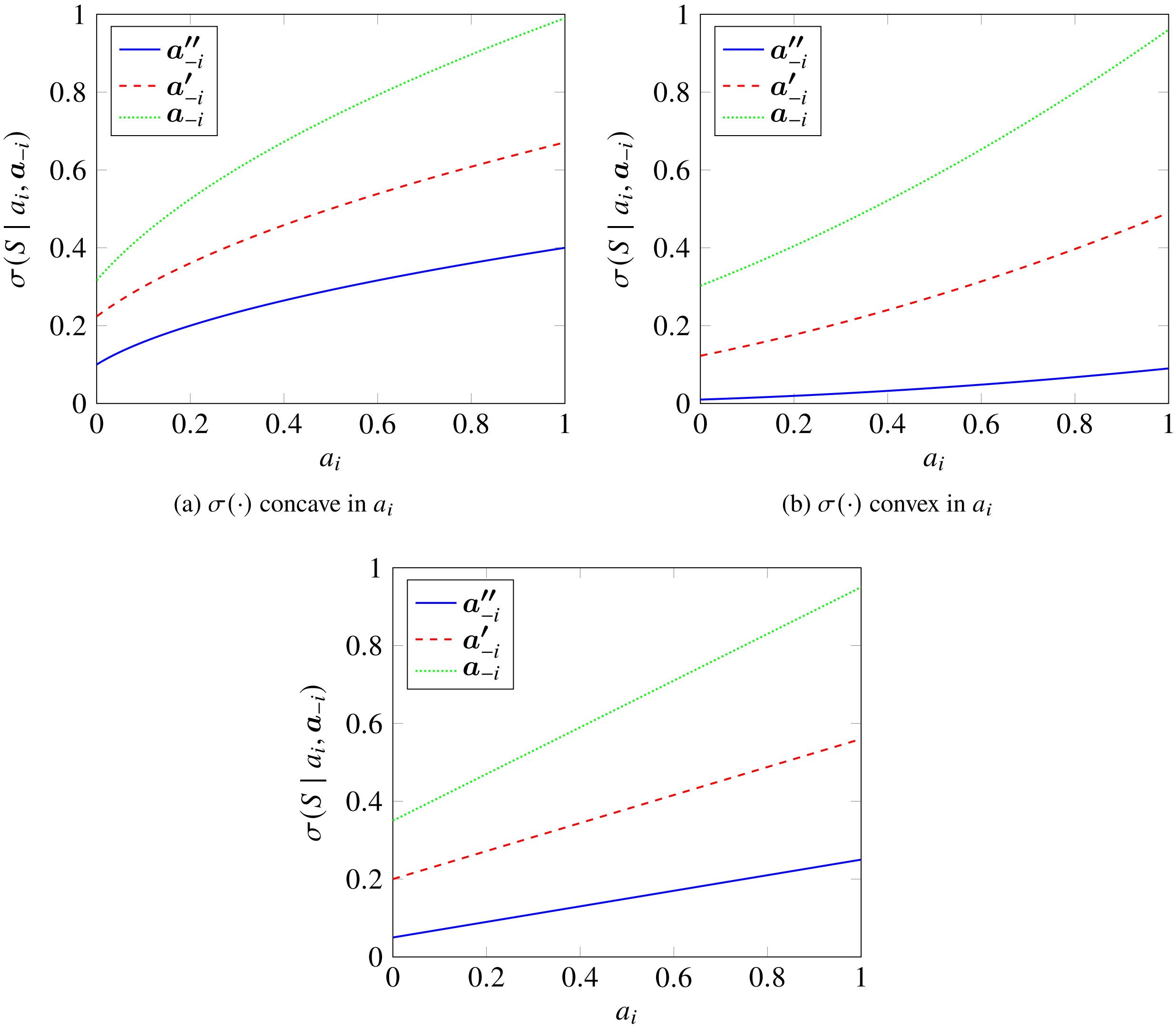
For the sake of clarity, one can rewrite (SM) in the following way:

$$\sigma(S|a'_i, \boldsymbol{a}_{-i}) - \sigma(S|a_i, \boldsymbol{a}_{-i}) \geq \sigma(S|a'_i, \boldsymbol{a'}_{-i}) - \sigma(S|a_i, \boldsymbol{a'}_{-i}) \tag{ID}$$

The equation (ID), which is known as the *increasing-differences property* and is equivalent to supermodularity for real-valued functions, perfectly captures the idea of complementarity between the agents' actions in the joint signal function. It indicates that the marginal effect of increasing agent i's action from a_i to a'_i on the likelihood of success is higher whenever one increases the non-i action profile from $\boldsymbol{a'}_{-i}$ to $\boldsymbol{a}_{-i}$.[4] Therefore, agent i's higher effort is more valuable (in a probabilistic

[4]It can be easily seen that whenever $a_i \geq a'_i$ and $\boldsymbol{a}_{-i} \geq \boldsymbol{a'}_{-i}$, then supermodularity holds trivially with equality.

sense) when every other agent is taking a higher action, namely by increasing $\boldsymbol{a}_{-i}$.



(a) $\sigma(\cdot)$ concave in a_i

(b) $\sigma(\cdot)$ convex in a_i

(c) $\sigma(\cdot)$ linear in a_i

Figure 1: Supermodularity in $(a_i, \boldsymbol{a}_{-i})$ of signal function $\sigma(S|\cdot)$

Figure 1 captures the idea of supermodularity graphically. For this figure, I fix the value of $\boldsymbol{a}_{-i}$ and observe how $\sigma(S|\cdot)$ behaves strictly as a function of a_i. For easiness of exposition, let the action space be $A_i = [0, 1]$.[5] Naturally, $\sigma(S|\cdot)$ is bounded from above by 1. It is assumed that $\boldsymbol{a}_{-i} > \boldsymbol{a}'_{-i} > \boldsymbol{a}''_{-i}$. Note that supermodularity poses no restrictions on the form of $\sigma(S|\cdot)$. Indeed, Figure 1a shows supermodularity for $\sigma(S|\cdot)$ concave in a_i; although the function is marginally

[5]This assumption is strictly made for the purpose of exposition of this figure. For the rest of this paper, I will again assume that $A_i = [\underline{a}, \overline{a}]$.

decreasing in a_i (by concavity), such effect decreases "more slowly" when the other agents' action profile $\boldsymbol{a}_{-i}$ is higher. Similarly, Figure 1b assumes that $\sigma(S|\cdot)$ is convex in a_i: the marginal effect of i's effort is increasing, and such marginal effect is "more accelerated" whenever $\boldsymbol{a}_{-i}$ increases. Finally, Figure 1c displays a linear signal function that exhibits supermodularity. In this case, the signal function has a greater slope for higher values of $\boldsymbol{a}_{-i}$. All three figures satisfy supermodularity: for higher fixed values of $\boldsymbol{a}_{-i}$, the marginal effect of increasing a_i is higher (with respect to its marginal effect if $\boldsymbol{a}_{-i}$ were lower).

Therefore, a joint signal function that exhibits supermodularity in the agents' actions would appear to be a desirable condition for the principal if she wishes to implement the joint task. However, as I will further study, supermodularity does not suffice to guarantee that the principal prefers the team project.

3.4 Optimal incentive schemes and main results

As mentioned earlier, the principal faces a dilemma: she can either make use of her information structure Ξ_x and assign individual tasks to each of the agents, or she can delegate a joint project in which the agents work together, thus employing her information structure Ξ_y. In a setting where each agent produces an individual signal, Fleckinger (2012) shows that whenever individual signals are conditionally independent, then the optimal incentive scheme is a piece-rate contract that exhibits independent performance evaluation. One can easily see that such result extends to this model's case in which the principal assigns individual tasks through Ξ_x, whose signal function is conditionally independent.

Therefore, the principal's decision essentially boils down to whether she offers an individual incentive scheme or a joint incentive scheme. For such resolution, she shall work out the optimal incentive scheme for each of the two cases and then compare her expected benefits in order to lean towards a decision on her task assignment. It is of this paper's particular interest to study the role of supermodularity of the joint signal function for such comparison.

3.4.1 Static setting

Individual task

For the following analysis, I will assume that the principal wishes to implement the highest possible action $\overline{a}$ for both the individual and team projects. More explicitly, I am assuming that the principal has solved the second step of the Grossman-Hart analysis explored in chapter 2, which concerned the selection of an action to be implemented. Of course, one may consider this assumption to simply mirror the fact that the choice of $\overline{a}$ comes after the principal has minimized the cost of implementation. Furthermore, I assume that the likelihoods of success between the tasks are equal whenever every agent involved in the project is disobedient.[6] Specifically, define $\hat{\boldsymbol{a}}_{-i} = (\hat{a}, \hat{a}, \dots, \hat{a})$ as an $n-1$ vector in which every agent except for i choose an action $\hat{a} < \overline{a}$. It is then assumed that $\pi(s|\hat{a}) = \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})$.

The principal's program if she were to assign an individual task to each agent i mirrors the one studied in chapter 2, albeit with the assumptions of risk-neutral agents and limited liability. Denote w_x the payment offered by the principal whenever the realized outcome is $x \in \{f, s\}$. Note that, since risk-aversion implies that $u(w_x) = w_x$, the Individual Rationality constraint (or Participation constraint) holds trivially if one imposes a Limited Liability constraint:

$$w_x \geq 0 \quad x \in \{f, s\}. \tag{LL-x}$$

Therefore, the principal's program considers the Limited Liability constraint as well as the Incentive Compatibility constraint (IC). Considering the fact that the individual task implementation involves the use of the information structure $\Xi_x = \langle X, \pi(\cdot|a_i)\rangle$, then incentive compatibility is written as follows:

$$\pi(s|\overline{a})w_s + (1 - \pi(s|\overline{a}))w_f - c(\overline{a}) \geq \pi(s|\hat{a})w_s + (1 - \pi(s|\hat{a}))w_f - c(\hat{a}) \quad \forall \hat{a} \in A_i. \tag{IC-x}$$

[6] I refer to "disobedience" in the sense of Myerson (1982), who defines an agent as *disobedient* if he does not choose the principal's recommended action.

Taking (IC-x) and (LL-x) into account, the principal's program is then:

$$\min_{w_f, w_s} \quad \pi(s|\overline{a})w_s + (1-\pi(s|\overline{a}))w_f \tag{3.1a}$$

$$\text{s.t.} \quad \pi(s|\overline{a})w_s + (1-\pi(s|\overline{a}))w_f - c(\overline{a})$$

$$\geq \pi(s|\hat{a})w_s + (1-\pi(s|\hat{a}))w_f - c(\hat{a}) \quad \forall \hat{a} \in A_i \tag{IC-x}$$

$$w_x \geq 0, \quad x \in \{f, s\}. \tag{LL-x}$$

Needless to say, the solution for the principal's program is a pair $\boldsymbol{w}^x = (w_f, w_s)$, which I shall refer to as the *optimal individual-task incentive scheme*. The ensuing statements show the actual form of such contract and use it to obtain the principal's expected utility.

Proposition 3.1. *Let* Ξ_x *be the information structure on which the contract is conditioned, such that the principal assigns the individual task for each* $i \in N$*. Then, the optimal individual-task incentive scheme that implements the action* $\overline{a}$ *is:*

$$\boldsymbol{w}^x = \left(0, \ \frac{c(\overline{a}) - c(\hat{a})}{\pi(x|\overline{a}) - \pi(x|\hat{a})}\right).$$

Proof. It is well-known from the incentive design literature that the limited liability constraint binds for the "bad state". In other words, $w_f = 0$, such that it is of the principal concern to simply determine the optimal incentive if $x = s$.[7] Since the program is linear, the principal can, for instance, choose an individual incentive w_s that makes the agent's incentive compatibility constraint binding, taking into account the fact that $w_f = 0$. Hence the principal solves:

$$\pi(s|\overline{a})w_s - c(\overline{a}) = \pi(s|\hat{a})w_s - c(\hat{a}).$$

Some simple algebra yields the optimal incentive for $x = s$:

$$w_s = \frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})}.$$

■

[7] See Laffont and Martimort (2002), chapter 4, for a textbook treatment of limited liability with risk-neutrality.

Corollary 3.1. *Let Ξ_x be the information structure on which the contract is conditioned, such that the principal assigns the individual task for each $i \in N$, and assume that the principal derives no utility from the individual project's failure: $f = 0$.*[8] *Then, the principal's expected utility, V^x, is:*

$$V^x = \pi(s|\overline{a}) \left[ns - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})} \right) \right]. \tag{3.2}$$

Proof. Recall that the principal's expected utility is $V(x, \boldsymbol{w}) = \mathbb{E}(x - \boldsymbol{w})$. It follows directly from imputing n times both the successful outcome s and the optimal incentive scheme $\boldsymbol{w}^x$ in the principal's expected utility, since each of the n agents is offered such contract to produce s. ∎

Joint task

Now, I assume that the principal draws the contract conditioning the incentives on the information structure $\Xi_y = \langle Y, \sigma(\cdot) \rangle$, which implies that she will assign a joint task. Naturally, a similar analysis to the one made for the individual task is necessary in order to determine the *optimal joint-task incentive scheme*, which is a pair $\boldsymbol{w}^y = (w_F, w_S)$, as well as the principal's payoff under these assumption.

As I stated previously, it is assumed that, for the joint task, the principal once again deems optimal to implement the action $\overline{a}$. Moreover, although the contract's payments are conditioned on the joint output y, the principal is still concerned with providing individual incentives to each agent, such that $a_i = \overline{a} \;\; \forall i \in N$.

Therefore, I derive the principal's program conditional on every non-i agent choosing the optimal action a, i.e. $\boldsymbol{a}_{-i} = \overline{\boldsymbol{a}}_{-i} = \underbrace{(\overline{a}, ..., \overline{a})}_{n-1}$. The problem closely mirrors the one studied for the individual task:

$$\min_{w_F, w_S} \quad \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) w_S + (1 - \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i})) w_F \tag{3.3a}$$

$$\text{s.t.} \quad \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) w_S + (1 - \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i})) w_F - c(\overline{a})$$

$$\geq \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) w_S + (1 - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})) w_F - c(\hat{a}) \quad \forall \hat{a} \in A_i \tag{IC-y}$$

[8] This assumption is without loss of generality as long as it is assumed that the joint task's failure is also valueless ($F = 0$).

$$w_y \geq 0, \quad y \in \{F, S\}. \tag{LL-y}$$

Needless to say, the implementation of $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ requires that the previous program's constraints be satisfied for each agent $i \in N$. The solution for the principal's joint-task program is derived in the next statement.

PROPOSITION 3.2. *Let Ξ_y be the information structure on which the contract is conditioned, such that the principal assigns the joint task in which the n agents collaborate. Then, the optimal joint-task incentive scheme that implements the action profile $\overline{\boldsymbol{a}} = (\overline{a}, \overline{\boldsymbol{a}}_{-i})$ is:*

$$\boldsymbol{w}^y = \left(0, \ \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \right).$$

Proof. The proof is analogous to that of the optimal individual-task incentive scheme. ■

COROLLARY 3.2. *Let Ξ_y be the information structure on which the contract is conditioned, such that the principal assigns the joint task in which the n agents collaborate, and assume that the principal derives no utility from the joint project's failure: $F = 0$. Then, the principal's expected utility, V^y, is:*

$$V^y = \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \left[S - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \right) \right]. \tag{3.4}$$

Proof. Once again, simply plug in n times the optimal contract $\boldsymbol{w}^y$ in the principal's expected utility $V(y, \boldsymbol{w}) = \mathbb{E}(y - \boldsymbol{w})$. ■

Note that the principal's expected utility for the joint task does not consider n times the successful outcome (S), while the individual-task utility does. This is due to the fact that, for the individual assignment, the principal "employs" *each* agent to produce s and thus expects to receive a total output of ns. On the other hand, the principal incentives the agents to cooperatively produce a shared total outcome S in the case of the joint project. The relationship between the total outcomes S and ns shall be key for the analysis that follows.

Comparison between potential tasks

The comparison of the principal's expected utilities obtained in the previous exploration of the optimal incentive schemes allows to determine on which signal should the principal condition the contract; or, more intuitively, whether the principal desires to assign a joint or individual project.

Although such comparison could seem straightforward and painless, a quick inspection of the principal's expected utilities V^x and V^y shall allow the reader to see that supermodularity in $(a_i, \boldsymbol{a}_{-i})$ of the joint signal function is not a sufficient condition to guarantee that the principal prefers to implement the joint project.

Therefore, further conditions are required to ensure the principal prefers to delegate any two of the assignments. The following analysis hence focuses on providing somewhat reasonable sufficient conditions that guarantee a preference for the joint task —though, of course, the conditions that favour the individual tasks are implicitly derived.

Definition (Backpacking condition). *Consider any action $\hat{a}$ such that $\hat{a} < \overline{a}$. The joint signal function $\sigma(S|\cdot)$ is said to satisfy the* ***backpacking condition*** *with respect to the individual signal function $\pi(s|\cdot)$ if the following inequality holds:*

$$\frac{\sigma(S|\, \overline{a},\ \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\, \hat{a},\ \hat{\boldsymbol{a}}_{-i})}{\pi(s|\overline{a}) - \pi(s|\hat{a})} \geq 1. \tag{BPC}$$

The intuition for the backpacking condition (BPC) goes as follows. First, consider a joint-task incentive compatibility constraint for the principal if she desires to implement action $\overline{a}$ from agent i conditioned on every other agent being disobedient, i.e. choosing $\hat{a}$:

$$\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\overline{a}) \geq \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\hat{a}).$$[9]

Note that, if one rearranges the previous incentive compatibility inequality, the coefficient on w_S is $\sigma(S|\, \overline{a},\ \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\, \hat{a},\ \hat{\boldsymbol{a}}_{-i})$. Hence such coefficient may be interpreted as the marginal incentive of agent i exerting $\overline{a}$ whenever everybody else is disobedient (not executing $\overline{a}$). Correspondingly, $\pi(s|\overline{a}) - \pi(s|\hat{a})$ is the marginal incentive on agent i choosing $\overline{a}$ for the individual task. For better

[9] Of course, the previous inequality does not consider $y = F$ since I have previously shown that, in optimality, $w_F = 0$.

exposition, rewrite (BPC) as:

$$\underbrace{\sigma(S|\,\overline{a},\ \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\,\hat{a},\ \hat{\boldsymbol{a}}_{-i})}_{\text{Marginal incentive to exert } \overline{a} \text{ given } \hat{\boldsymbol{a}}_{-i}} \quad \geq \quad \underbrace{\pi(s|\overline{a}) - \pi(s|\hat{a})}_{\text{Marginal incentive to exert } \overline{a} \text{ for indiv. task}} \tag{BPC'}$$

Therefore, the BPC is a measure of how the agent's marginal incentive to choose $\overline{a}$ (given that every other agent is disobedient) relates to his marginal incentive to choose such action for an individual project. BPC states that, conditional on every other agent underperforming, the marginal effect of i increasing his action from $\hat{a}$ to $\overline{a}$ on the likelihood of success S is (weakly) greater than such marginal effect for an individual task.

The term *backpacking* comes from the fact that, although every other agent is not exerting enough effort, agent i's action is so influential on the team project's outcome that his effect on the likelihood of joint success is at least as great as his effect on the individual task's success (which solely depends on his effort). One might therefore say that agent i is "carrying the team on his back".

The following statement, which is the main result of this paper, shows that, whenever the total outputs resulting from the success of both projects are equivalent, then the backpacking condition and supermodularity are sufficient for the principal's preference for the joint task.

Proposition 3.3. *Let $S = ns$, such that the total outcome yielded by the joint project equals that of the n individual tasks. If $\sigma(S|\cdot)$ satisfies **supermodularity** in $(a_i, \boldsymbol{a}_{-i})$ and the **backpacking condition** with respect to $\pi(s|\cdot)$, then the principal will condition the contract on Ξ_y. That is, the principal (weakly) prefers to assign the joint task.*

Proof. From the previous analysis, see that $V^y \geq V^x$ iff

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \left[S - n\left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})}\right)\right] \geq \pi(s|\overline{a}) \left[ns - n\left(\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})}\right)\right] \tag{3.5}$$

Since $S = ns$, one can easily see it suffices to show that, under BPC and supermodularity, the following two inequalities hold:

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}), \tag{i}$$

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}) - \pi(s|\hat{a}) \tag{ii}$$

I first prove inequality (*ii*). The proof is by contradiction. Hence assume that BPC and supermodularity hold but (*ii*) does not, i.e.:

$$\begin{aligned}
&\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) - \pi(s|\hat{a}) \\
\iff \quad & \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) + \underbrace{\pi(s|\hat{a})}_{= \sigma(S|\hat{a},\ \hat{\boldsymbol{a}}_{-i})} < \pi(s|\overline{a}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) \\
\iff \quad & \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a},\ \hat{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}).
\end{aligned} \tag{3.6}$$

Since $\sigma(\cdot)$ is supermodular in $(a_i, \boldsymbol{a}_{-i})$, we know that the following inequality holds:

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a},\ \hat{\boldsymbol{a}}_{-i}) \geq \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) \tag{3.7}$$

Then, combine the equations (3.6) and (3.7) to see that:

$$\begin{aligned}
&\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) \\
\iff \quad & \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) \\
\iff \quad & \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) - \underbrace{\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})}_{= \pi(s|\hat{a})} \\
\iff \quad & \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i}) < \pi(s|\overline{a}) - \pi(s|\hat{a}),
\end{aligned} \tag{3.8}$$

which contradicts BPC. Therefore, it must be that if BPC and supermodularity hold, then (*ii*) is satisfied.

As for inequality (*i*):

$$(i) \quad \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}).$$

Note that, from BPC and using $\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i}) = \pi(s|\hat{a})$, one can infer that:

$$\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}).$$

The inequality (i) then follows directly from the fact that

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \geq \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}).$$

BPC thus implies that (i) holds. ■

The previous result compares the expected utilities of the principal assuming that the successful outcomes for the projects are equivalent, such that the principal's preference between the tasks depends solely on the likelihoods of success resulting from the signal functions $\pi(s|\cdot)$ and $\sigma(S|\cdot)$; that is, it provides a direct comparison of the tasks through the principal's expected utilities if he were to condition the contract on any two of the signals.

Intuitively, Proposition 3.3 states that the principal perceives the predominance of the joint task if two things that concern a_i happen simultaneously. Not only should the marginal effect of agent i's full effort be higher whenever the rest of the team's action profile is higher (i.e. supermodularity), but the principal also demands that the marginal effect of $a_i = \overline{a}$ on the likelihood of joint success (even if the rest of the team shirks) be greater than such effect when i performs individually. Therefore, paying w_S to agent i to incentivize him to exert the optimal action $\overline{a}$ seems like a bargain: his action has an extremely influential marginal effect on the probability of team success, even if the rest of the agents are disobedient.

Naturally, one should think twice before assuming that $\sigma(S|\cdot)$ is supermodular, let alone assuming that the backpacking condition holds. Both conditions —particularly BPC— seem a bit demanding. Consequently, these conditions may lead us to think that, in most cases, the principal will be endowed with a joint signal y that is less attractive to employ than the individual signal x, in which case assigning n individual tasks is preferred by the principal.

Nevertheless, recall that the previous statement results from the assumption that total outputs are equal ($S = ns$). Certainly, it is also of my interest to study the conditions for which assigning the team project is desirable when the success of the tasks is not equally-valued. Accordingly, for the following statement I drop the assumption that $S = ns$ and obtain moderately reasonable conditions through which the team assignment prevails.

Proposition 3.4. *If $\sigma(S|\cdot)$ satisfies the backpacking condition (BPC) with respect to $\pi(s|\cdot)$, and the condition*

$$\frac{S}{n} - s \geq w_S - w_s$$

holds, then the principal will condition the contract on Ξ_y. That is, the principal (weakly) prefers to assign the joint task. Moreover, drop the BPC and instead assume that $\sigma(S|\,\overline{a},\, \overline{\boldsymbol{a}}_{-i}) = \pi(s|\overline{a})$. Then, the principal (weakly) prefers to assign the joint task if and only if the stated condition holds.

Proof. See Appendix B1.

For the intuition behind the result, think of the case in which $\sigma(S|\,\overline{a},\, \overline{\boldsymbol{a}}_{-i}) = \pi(s|\overline{a})$, such that the main condition is necessary and sufficient for the predominance of the joint task. That is,

$$V^y \geq V^x \quad \Longleftrightarrow \quad \underbrace{\frac{S}{n} - s}_{\text{Value gain } \textit{per-agent} \text{ for success of joint task}} \geq \underbrace{w_S - w_s}_{\text{Additional cost } \textit{per-agent} \text{ for success of joint task}} .$$

Indeed, the inequality simply contends that, for the joint task's expected utility to be greater than that of the individual project, the difference between the values for the successful outcomes *per agent* are to be greater than the difference between the costs per agent that the tasks' success entails.

Alternatively, rewrite the condition as:

$$\underbrace{\frac{S}{n} - w_S}_{\text{Utility } \textit{per-agent} \text{ for success of joint task}} \geq \underbrace{s - w_s}_{\text{Utility } \textit{per-agent} \text{ for success of individual task}} .$$

In other words, if one assumes that the likelihoods of success given full effort by the agents are equivalent (i.e. $\sigma(S|\,\overline{a},\, \overline{\boldsymbol{a}}_{-i}) = \pi(s|\overline{a})$), then the principal shall simply compare her expected utility *per-agent* if the tasks are successful, namely the difference between the per-agent value of the successful outcome and the payment given to the agent for success. Of course, if such difference is greater for the joint task, then the principal may wish to implement a team project rather than n individual tasks. Therefore, assigning team tasks is optimal if and only if the joint outcome's success S is sufficiently valuable (with respect to the success of the individual task), while the incentive given to the agent to jointly achieve success w_S is sufficiently small.

Multiplicity of equilibria in the *joint-task game*

The fact that the joint task incentive scheme considers multiple agents is somewhat powerful: whenever the principal implements a joint project, the incentive scheme essentially defines a normal-form game played between the n agents, in which the agents' actions are cooperative. For this *joint-task game*, incentive compatibility allows to see that the strategy profile in which every agent "cooperates" with the implemented action, $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$, is a Nash equilibrium. Nonetheless, there is a downside, as nothing guarantees that $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ is a *unique* Nash equilibrium. Indeed, the following statement shows that full non-cooperation with the principal also composes a Nash equilibrium.

PROPOSITION 3.5. *Assume the signal function $\sigma(S|\cdot)$ is supermodular in $(a_i, \boldsymbol{a}_{-i})$, and consider any action profile $\hat{\boldsymbol{a}}_{-i} = (\hat{a}, ..., \hat{a})$, where $\hat{a} < \overline{a}$, in which every agent (except for i) chooses an action lower than the implemented action $\overline{a}$. Then, it is optimal for agent i to also choose $\hat{a}$. That is, the strategy profile $(\hat{a}, \hat{\boldsymbol{a}}_{-i})$ constitutes a Nash equilibrium for the joint-task game.*

Proof. For the proof, it suffices to show that

$$\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\hat{a}) \geq \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\overline{a}). \tag{3.9}$$

The equation (3.9) can be rewritten as

$$w_S \leq \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})}. \tag{3.10}$$

Replacing w_S in (3.10) with its derived optimal value yields the ensuing condition

$$\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \leq \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})}$$

$$\iff \quad \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) \geq \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})$$

$$\iff \quad \sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i}) \geq \sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) + \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}),$$

which is the expression for supermodularity of $\sigma(\cdot)$. ■

The existence of multiple equilibria deals a huge blow to the signal y's chances of being employed to draw the contract. Indeed, even if the sufficient conditions for the principal to derive a higher expected utility through the joint task hold, the multiplicity of equilibria may lead the principal to lack confidence on the agents coordinating on her preferred equilibrium $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$. For example, consider the case in which $n = 2$. One can think that, if for some exogenous reason the agent j disobeys the implemented action $\overline{a}$, then agent i's best response given the optimal scheme $\boldsymbol{w}^y$ is to also be disobedient. The principal may, therefore, opt for a task in which there is no risk of the agents being disobedient.

3.5 Repeated Interaction

The previous analysis has shown that, due to the strong conditions required for the joint project to be desirable, as well as the complications that arise from multiple equilibria, the principal should not be too convinced to implement a team project. Nonetheless, in a similar study, Che and Yoo (2001) show that the repeated interaction between the agents makes the joint assignment more attractive for the principal. In what follows, I aim to replicate such results for the conditions of this model. Hence I shall prove that, analogously to the Che-Yoo analysis, the repeated interactions between not only each agent $i \in N$ and the principal, but also between the n agents themselves, yield softer conditions for the predominance of the joint project. Even more, the repetition of the joint-task game may enforce cooperative behaviour between the agents.

Accordingly, suppose that the interaction between the principal and the n agents is infinitely repeated.[10] In chapter 2, it was shown that, in a setting with repeated moral hazard, the optimal incentive scheme may exhibit history-dependence. However, given the already complex environment that arises from the study of incentive design for multiple agents, I will assume that the principal commits to a time-invariant contract from the beginning of the interaction in $t = 1$. The stationary contract thus covers an infinite number of periods.

A key assumption must be made. Define a_i^t the action that agent i chooses in period t. It is assumed that, **only** when performing a team task, the agents' actions are mutually observable. That is, after actions have been chosen in period t, and before $t + 1$ starts, any agent $j \in N,\ j \neq i$, can

[10] The results presented in this section should hold for the case in which the interaction is repeated for a finite number of periods. I merely assume infinite periods to simplify the analysis.

perfectly observe a_i^t, provided that i and j were assigned to a joint project. The principal, of course, is still assumed to be dependent on the information structures Ξ_x and Ξ_y, since she cannot observe a_i^t.

Since the first-period contract is infinitely repeated, it is straightforward to see that the *optimal repeated individual-task incentive scheme*, $\{\boldsymbol{w}^x\}^\infty$, is the one obtained in Proposition 3.1 repeated over infinite periods, namely:

$$\{\boldsymbol{w}_t^x\}_{t=1}^\infty = \left\{\left(0,\ \frac{c(\overline{a}) - c(\hat{a})}{\pi(x|\overline{a}) - \pi(x|\hat{a})}\right)\right\}_{t=1}^\infty .$$

The joint task contract, on the other hand, need not be the same as for the static setting. Such observation comes from the fact that the repeated team contract essentially defines the payoffs of a repeated game. Naturally, the stage game for this repeated interaction is the normal-form game studied in the previous section. Consequently, the long-term strategic interaction between the players may lead to a divergence between the optimal joint-task contract derived for the stage game and the repeated team contract that shall be obtained.

If the principal is clever enough —and she is indeed—, she should notice that it is optimal to adopt a "*let them play*" strategy in her design of the contract. The principal observes that the incentive scheme she will offer shall determine the payoffs of a repeated game played between the n agents. Since the principal read Friedman (1971), she knows that, for their repeated interaction, the agents can adopt a strategy that leads every $i \in N$ to play a Pareto-dominant strategy in each period. One can easily see that the strategy profile $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ Pareto-dominates any other possible strategy profile, since it maximizes the agents' likelihood on achieving the successful outcome S on their joint task and thus of receiving the payment w_S (which has not been derived yet for the repeated setting) in each period.

More explicitly, the principal should provide an incentive scheme that induces a strategy profile in which agent i is punished if he does not exert the maximal effort $\overline{a}$. Such punishment does not come from an incentive offered by the principal, but from the agents themselves —hence the *let them play* expression.

Consider the following strategy. In each period t, each agent will play $\overline{a}$ provided that he observed the strategy profile $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ in period $t-1$. However, the agent will threaten with reverting to the lowest possible action $\underline{a} < \overline{a}$ if he does not observe $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ after $t-1$. Needless to say, such threat is only credible if the strategy profile $(\underline{a}, \underline{\boldsymbol{a}}_{-i})$ constitutes a Nash equilibrium for the stage game. Though I have already proven this for the static joint incentive scheme through Proposition 3.5, one must first derive the optimal repeated incentive scheme to conclude such claim for this setting. The following statement thus displays the *optimal repeated joint-task contract*.

Proposition 3.6. *Consider any $\hat{a} < \overline{a}$, and assume that agent i's disobedience in $t-1$ ($a_i^{t-1} \neq \overline{a}$) is punished by every other agent choosing $\hat{a}$ infinitely from t. Then, the optimal repeated joint-task incentive scheme is:*

$$\{\boldsymbol{w}_t^y\}_{t=1}^{\infty} = \left\{ \left(0, \ \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - (1-\delta)\sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) - \delta\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})} \right) \right\}_{t=1}^{\infty}. \tag{3.11}$$

Proof. See Appendix B2.

A statement arises from the previous proposition. It confirms that, as a matter of fact, the optimal repeated incentive scheme satisfies the conditions required for the agents to credibly threaten each other through the action $\underline{a}$.

Lemma 3.1. *Assume the signal function $\sigma(S|\cdot)$ is supermodular in $(a_i, \boldsymbol{a}_{-i})$, and consider any action $\hat{a} < \overline{a}$. If $\boldsymbol{a}_{-i} = \hat{\boldsymbol{a}}_{-i}$ in the stage game of the repeated joint-task game, then it is optimal for agent i to also choose $\hat{a}$. That is, the strategy profile $(\hat{a}, \hat{\boldsymbol{a}}_{-i})$ constitutes a Nash equilibrium for the stage game of the repeated joint-task game.*

Proof. The proof is essentially the same as that for Proposition 3.5. For such proof, it was determined that w_S should satisfy:

$$w_S \leq \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \hat{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})}. \tag{3.12}$$

Simply plug-in w_S from the optimal repeated joint-task contract (3.11) and see that condition (3.12) holds. ■

This result confirms that the profile $(\underline{a}, \underline{\boldsymbol{a}}_{-i})$ is a credible threat for the agents' strategic interaction in the repeated joint-task game. Using such information, one can see that the repeated interaction has a convenient *subgame-perfect equilibrium*.

Proposition 3.7. *Assume the signal function $\sigma(S|\cdot)$ is supermodular in $(a_i, \boldsymbol{a}_{-i})$, and let $\{\boldsymbol{w}_t^y\}_{t=1}^{\infty}$ be the optimal repeated joint-task contract. For δ sufficiently close to 1, the strategy profile $\{(\overline{a}^t, \overline{\boldsymbol{a}}_{-i}^t)\}_{t=1}^{\infty}$ constitutes a subgame-perfect equilibrium for the repeated joint-task game.*

Proof. The existence of a minmax strategy profile for the stage game, namely $(\underline{a}, \underline{\boldsymbol{a}}_{-i})$, follows from Lemma 3.1. Moreover, the optimal incentive scheme guarantees that the profile $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ Pareto-dominates any other strategy profile. It then follows from the Folk Theorem that $\{(\overline{a}^t, \overline{\boldsymbol{a}}_{-i}^t)\}_{t=1}^{\infty}$ is a subgame-perfect equilibrium. ■

Therefore, the principal's plan was right all along. Exercising her role as a *mechanism designer* for the repeated joint-task game, she designed a time-invariant contract that would align the agents' interests with her own, since $(\overline{a}, \overline{\boldsymbol{a}}_{-i})$ Pareto-dominates any other profile for every player of this game. The agents, credibly threatening each other with reverting to the minmax profile if anyone deviates from $\overline{a}$, enforce themselves into cooperative relationship in which every $i \in N$ chooses the principal-preferred action. Therefore, when the stage game is repeated infinitely many times, the principal is no longer concerned with the possibility of multiple equilibria arising, since the agents enforce themselves into her preferred equilibrium.

Remark. The interaction between the agents is extremely convenient for the principal when it is infinitely repeated. For any agent i, both "continuation" profiles $\{(\underline{a}, \underline{\boldsymbol{a}}_{-i})\}_{t+1}^{\infty}$ and $\{(\overline{a}, \overline{\boldsymbol{a}}_{-i})\}_{t+1}^{\infty}$ are self-enforcing, since his action in any period t, a_i^t, will essentially implement the rest-of-the-game strategy profile. The optimal repeated incentive scheme is therefore a *self-enforcing contract*: the contract need not be enforced by some sort of higher institution or court of law, since the agents themselves can discipline each other through their credible threats. Furthermore, recall that for the repeated single-agent model studied in chapter 2, it was assumed that the principal need not consider a second-period incentive compatibility constraint, since some sort of higher law shall force the agent to choose the action to which he committed before the interaction began —or, alternatively, the agent is "loyal". One can see that the repeated joint-task incentive scheme does not require such assumption: its self-enforcing property guarantees that, because the agents regulate themselves, the principal can comfortably sleep at night knowing that the agents will always respect their commitment to the contract, independently of their loyalty to the principal.

Moreover, the reader may have noticed from Proposition 3.6 that the optimal repeated incentive for success ($y = S$) appears to be lower than its static counterpart. The following statement formalizes such claim.

Proposition 3.8. *Let the principal assign the joint task in the repeated interaction, such that the optimal repeated joint-task incentive scheme is offered to each agent, and assume that $\delta > 0$. Then, the principal's expected utility for the stage game of the repeated joint-task game, V_R^y, is (strictly) greater than the principal's expected utility for the static setting, V^y. That is, $V_R^y > V^y$.*

Proof. See Appendix B3.

The previous result suggests that the conditions required for the principal to prefer the assignment of the joint task (repeatedly) should be less restrictive than those found for the static setting. The intuition behind this result actually has its roots on Proposition 3.7. Given that the agents can mutually sanction each other through the minmax profile, the principal can allow herself to partially delegate her incentivizing duties to the agents themselves. Of course, if the agents do not care about the future ($\delta = 0$), their sanctioning power will vanish and we are back to the one-shot game studied in section 3.3.

3.6 Conclusion and possible extensions

Why should firms create teams among their workers and assign them joint projects? Even if the value of a successful team outcome is equivalent to that of n successful individual outcomes, could a principal still lean towards offering a joint incentive scheme in which the agents cooperate for a common goal?

In the previous analysis, I provide sufficient conditions that allow to answer these questions with an affirmative. Specifically, I show that an agent i's action should satisfy two conditions: (i) the marginal effect of increasing a_i on the likelihood of team success is greater when the remaining

$n-1$ agents provide higher efforts (supermodularity); and (*ii*), the marginal effect of increasing a_i on the likelihood of team success is greater than its marginal effect on the probability of individual success, even if the remaining $n-1$ agents are shirking. The latter assumption, which I termed the *backpacking condition*, essentially expresses that a single agent is able to "carry the team on his back" by providing the optimal effort.

Regardless of the previous conditions seeming to be a bit daunting, the fact that joint tasks are commonly seen in practice leads one to wonder if perhaps such conditions are in fact satisfied. Even more, the results concerning repeated interactions may offer some help, as I proved that repeated joint tasks are slightly more attractive than their static counterpart. Additionally, the dynamic joint contract is self-enforcing on the team equilibrium for which every agent chooses the Pareto-efficient strategy, i.e. the principal can rely on the agents to address the issue with multiple equilibria.

Of course, the real-life observation of team projects being commonly assigned may very well be due to the fact that team outcomes are much more valuable than the sum of individual projects, as Proposition 3.4 suggests.

I consider that the following extensions to the model should produce somewhat interesting results. First, while the assumption of limited liability and agent risk-neutrality allows to understand the trade-off between incentives and rent extraction, an analysis that instead derives the trade-off between insurance and incentives may lead to less restrictive conditions for the joint task to prevail.

Furthermore, the repeated joint-task game exploration relied on the assumption of the incentive scheme being stationary. In chapter 2, I showed that the dynamic contract for a single agent exhibits history-dependence. This property should be explored for dynamic incentive schemes designed for multiple agents.

Finally, one should not overlook the assumption regarding the principal deeming both information structures to be sufficient. For example, consider the case in which the team outcome's success is strictly more valuable than the aggregated individual outputs, but the information structure for the joint outcome is less informative in the sense of Blackwell than that for the individual outcome. Should the principal still assign a joint task, albeit knowing that the contract would be conditioned on an unreliable signal?

Concluding remarks

The exploration of the intricacies of the contract theory framework, particularly that for the moral hazard problem, should provide enough evidence to convince the reader that the task of designing incentives for agents is extremely complex. Naturally, I hope this work provides enough insights to illuminate such intricacies carefully. Even more, I expect that the analysis presented in chapter 3 —which constitutes my main contribution to the literature on the theory of incentives— can provide clarity on the relevance of cooperation when one must design incentives for multiple agents that work together.

Moreover, the complexity of this framework should lead one to reconsider the role of incentives in most interactions. As I suggested in chapter 1, economic thought ran into a wall when information asymmetries were to be embedded in classical frameworks, such as General Equilibrium Theory. Ironically, I believe that economists should be glad for such obstacle, since it allowed the profession not only to pivot to the mathematically-elegant mechanism design theory, but also to reconsider the importance of incentives for even the most trivial situations. Indeed, "people respond to incentives," and our development of a theory of incentives —be it through mechanism design or contract theory— is fundamental for the understanding of economics.

Appendix A

A1. ***Proof of Lemma 2.1.***

I first show that the PC is binding. The derived first-order conditions are:

$$(v^H): \quad \frac{1}{u'(w^H)} = \mu + \lambda \frac{\Delta\pi}{\pi^e} \tag{3.13}$$

$$(v^L): \quad \frac{1}{u'(w^L)} = \mu - \lambda \frac{\Delta\pi}{1-\pi^e} \tag{3.14}$$

Write (3.13) and (3.14) as:

$$\pi^e \mu = \frac{\pi^e}{u'(w^H)} - \lambda\Delta\pi \tag{3.15}$$

$$(1-\pi^e)\mu = \frac{(1-\pi^e)}{u'(w^L)} + \lambda\Delta\pi \tag{3.16}$$

Adding up (3.15) and (3.16) yields:

$$\mu = \frac{\pi^e}{u'(w^H)} + \frac{1-\pi^e}{u'(w^L)} > 0 \tag{3.17}$$

Therefore, the participation constraint is binding.

As for the IC binding, replace (3.17) in (3.13). This yields:

$$\begin{aligned} \frac{1}{u'(w^H)} &= \frac{\pi^e}{u'(w^H)} + \frac{1-\pi^e}{u'(w^L)} + \lambda\frac{\Delta\pi}{\pi^e} \\ \iff \quad & \lambda\Delta\pi = \pi^e(1-\pi^e)\left(\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)}\right) \end{aligned}$$

$$\iff \quad \lambda = \underbrace{\frac{\pi^e(1-\pi^e)}{\Delta\pi}}_{>0}\underbrace{\left(\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)}\right)}_{>0} > 0, \tag{3.18}$$

where $\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)} > 0$ follows from the fact that $v^H > v^L$, which is itself a result of IC. Then, the concavity of u implies that $u'(w^H) < u'(w^L)$, and thus $\frac{1}{u'(w^H)} > \frac{1}{u'(w^L)}$.

Therefore, the incentive compatibility constraint is also binding. ■

A2. ***Proof of Proposition 2.2.***

From Lemma 2.1, we know that IC and PC are binding. Therefore, write (IC-SB) as:

$$\begin{aligned} &\pi^e v^H + (1-\pi^e)v^L - c = \pi^s v^H + (1-\pi^s)v^L \\ \iff \quad &\Delta\pi u(w^H) - c = \Delta\pi u(w^L) \\ \iff \quad &u(w^L) = u(w^H) - \frac{c}{\Delta\pi} \end{aligned} \tag{3.19}$$

Replacing (3.19) in binding PC and substituting $u(w^i)$ for v^i, we get:

$$\begin{aligned} &\pi^e u(w^H) + (1-\pi^e)\left(u(w^H) - \frac{c}{\Delta\pi}\right) - c = 0 \\ \iff \quad &u(w^H) = c + (1-\pi^e)\frac{c}{\Delta\pi} \\ \implies \quad &u(w^L) = u(w^H) - \frac{c}{\Delta\pi} = c - \pi^e\frac{c}{\Delta\pi} \end{aligned} \tag{3.20}$$

The utilities associated to the optimal incentives are thus given by:

$$u(w^H) = c + (1-\pi^e)\frac{c}{\Delta\pi}, \tag{3.21}$$

$$u(w^L) = c - \pi^e\frac{c}{\Delta\pi} \tag{3.22}$$

Using $h(u(w)) = u^{-1}(u(w)) = w$, write the optimal incentives for each outcome as:

$$w^H = h\left(c + (1-\pi^e)\frac{c}{\Delta\pi}\right), \tag{3.23}$$

$$w^L = h\left(c - \pi^e \frac{c}{\Delta\pi}\right). \tag{3.24}$$

■

A3. ***Proof of Proposition 2.5.***

The first-order conditions for the problem are:

$$(v_1): \quad \frac{1}{u'(w_1(x_i))} = \mu + \lambda_{\hat{a}}\left(1 - \frac{\pi_i(\hat{a}_1)}{\pi_i(a_1)}\right) \tag{3.25}$$

$$(v_2): \quad \frac{1}{u'(w_2(x_{ij}))} = \mu + \lambda_{\hat{a}}\left(1 - \frac{\pi_i(\hat{a}_1)\pi_j(\hat{a}_2)}{\pi_i(a_1)\pi_j(a_2)}\right) \tag{3.26}$$

From (3.25), it is easy to see that the first-period incentive's optimality condition mirrors that of the static model. Furthermore, condition the expected value of the second-period condition (3.26) on the realization of an outcome x_i to get:

$$\mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \mathbb{E}\left[\mu + \lambda\left(1 - \frac{\pi_i(\hat{a}_1)\pi_j(\hat{a}_2)}{\pi_i(a_1)\pi_j(a_2)}\right)\middle| x_i\right]$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \sum_{j=1}^{n}\pi_j(a_2)\left[\mu + \lambda\left(1 - \frac{\pi_i(\hat{a}_1)\pi_j(\hat{a}_2)}{\pi_i(a_1)\pi_j(a_2)}\right)\right]$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \mu + \lambda\left[1 - \sum_{j=1}^{n}\pi_j(a_2)\left(\frac{\pi_i(\hat{a}_1)\pi_j(\hat{a}_2)}{\pi_i(a_1)\pi_j(a_2)}\right)\right]$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \mu + \lambda\left[1 - \frac{\pi_i(\hat{a}_1)}{\pi_i(a_1)}\sum_{j=1}^{n}\pi_j(a_2)\left(\frac{\pi_j(\hat{a}_2)}{\pi_j(a_2)}\right)\right]$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \mu + \lambda\left[1 - \frac{\pi_i(\hat{a}_1)}{\pi_i(a_1)}\underbrace{\sum_{j=1}^{n}\pi_j(\hat{a}_2)}_{=1}\right]$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \mu + \lambda\left(1 - \frac{\pi_i(\hat{a}_1)}{\pi_i(a_1)}\right)$$

$$\iff \quad \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right) = \frac{1}{u'(w_1(x_i))},$$

where the last equality comes from condition (3.25). Therefore,

$$\frac{1}{u'(w_1(x_i))} = \mathbb{E}\left(\frac{1}{u'(w_2(x_{ij}))}\middle| x_i\right).$$

■

Appendix B

B1. ***Proof of Proposition 3.4.***

First, see that $V^y \geq V^x$ iff

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) \left[S - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \right) \right] \geq \pi(s|\overline{a}) \left[ns - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})} \right) \right]$$

Notice that the BPC is equivalent to the following condition:

$$\sigma(S|\, \overline{a},\ \hat{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}),$$

which, using the fact that $\sigma(S|\, \overline{a},\ \overline{\boldsymbol{a}}_{-i}) \geq \sigma(S|\, \overline{a},\ \hat{\boldsymbol{a}}_{-i})$, in turn implies that:

$$\sigma(S|\, \overline{a},\ \overline{\boldsymbol{a}}_{-i}) \geq \pi(s|\overline{a}).$$

Therefore, it is sufficient to show that the following inequality holds:

$$S - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \right) \geq ns - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})} \right)$$

$$\iff \quad S - ns \geq n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})} \right) - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})} \right)$$

$$\iff \quad S - ns \geq n \left(\underbrace{\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})}}_{=w_S} - \underbrace{\frac{c(\overline{a}) - c(\hat{a})}{\pi(s|\overline{a}) - \pi(s|\hat{a})}}_{=w_s} \right)$$

$$\iff \quad S - ns \geq n(w_S - w_s)$$

$$\iff \quad \frac{S}{n} - s \geq w_S - w_s$$

■

B2. ***Proof of Proposition 3.6.***

The proof closely mirrors that of Propositions 3.1 and 3.2.

From the argument provided for the proof of Proposition 3.1, it is straightforward to see that $w_F = 0$. As for the case of w_S, note that the repeated game incentive compatibility constraint for agent i should take into account the threat of the non-i agents playing $\hat{\boldsymbol{a}}_{-i}$ infinitely. Such constraint is thus:

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i})w_S - c(\overline{a}) + \frac{\delta}{1-\delta}\left[\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i})w_S - c(\overline{a})\right] \geq \sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})w_S - c(\hat{a}) + \frac{\delta}{1-\delta}\left[\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\hat{a})\right] \tag{IC-R}$$

The LHS of (IC-R) is agent i's first-period utility if $y = S$ provided he chose $\overline{a}$, plus his discounted utility over infinite periods if $x = S$ conditioned on $a_i = \overline{a}$. Notice that $\boldsymbol{a}_{-i} = \overline{\boldsymbol{a}}_{-i}$ over every period, since the agent always exerts $\overline{a}$. The RHS is the agent's first-period utility if $y = S$ provided he was disobedient on his chosen action (though the rest of the agents did choose $\overline{\boldsymbol{a}}_{-i}$), plus his discounted utility over infinite periods if $x = S$ conditioned on $a_i^t = \hat{a}$ and $\boldsymbol{a}_{-i}^t = \hat{\boldsymbol{a}}_{-i}$ $\forall t > 1$ (since i was punished for being disobedient in $t = 1$). Some simple rearranging of (IC-R) yields:

$$\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i})w_S - c(\overline{a}) \geq (1-\delta)\sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i})w_S + \delta\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})w_S - c(\hat{a}), \tag{IC-R'}$$

Naturally, the principal will choose an optimal w_S such that the incentive compatibility constraint is binding. Therefore, one can rearrange and write (IC-R') with equality such that

$$w_S = \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \overline{\boldsymbol{a}}_{-i}) - (1-\delta)\sigma(S|\hat{a}, \overline{\boldsymbol{a}}_{-i}) - \delta\sigma(S|\hat{a}, \hat{\boldsymbol{a}}_{-i})}.$$

■

B3. ***Proof of Proposition 3.8.***

As in Corollary 3.2, one can simply input n times the repeated scheme on the principal's vNM utility to obtain V_R^y. Then, $V_R^y > V^y$ iff

$$V_R^y = \sigma(S|\overline{a}, \boldsymbol{\overline{a}_{-i}}) \left[S - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - (1-\delta)\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) - \delta\sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}})} \right) \right]$$

$$> \sigma(S|\overline{a}, \boldsymbol{\overline{a}_{-i}}) \left[S - n \left(\frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i})} \right) \right] = V^y$$

$$\iff \quad \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - (1-\delta)\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) - \delta\sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}})} < \frac{c(\overline{a}) - c(\hat{a})}{\sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i})}$$

$$\iff \quad \sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - (1-\delta)\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) - \delta\sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}}) > \sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i})$$

$$\iff \quad \sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) + \delta\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) - \delta\sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}}) > \sigma(S|\overline{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i})$$

$$\iff \quad \delta \left[\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) - \sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}}) \right] > 0,$$

which holds since $\delta > 0$ and $\sigma(S|\hat{a}, \boldsymbol{\overline{a}}_{-i}) > \sigma(S|\hat{a}, \boldsymbol{\hat{a}_{-i}})$. ■

References

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53:91–96.

Bertsekas, D. P. (1974). Necessary and Sufficient Conditions for Existence of an Optimal Portfolio. *Journal of Economic Theory*, 8(2):235–247.

Che, Y. K. and Yoo, S. W. (2001). Optimal Incentives for Teams. *American Economic Review*, 91(3):525–541.

Fleckinger, P. (2012). Correlation and Relative Performance Evaluation. *Journal of Economic Theory*, 147(1):93–117.

Fleckinger, P., Martimort, D., and Roux, N. (forthcoming, 2024). Should They Compete or Should They Cooperate? The View of Agency Theory. *Journal of Economic Literature*.

Friedman, J. W. (1971). A Non-cooperative Equilibrium for Supergames. *Review of Economic Studies*, 38(1):1–12.

Fudenberg, D., Levine, D., and Maskin, E. (1994). The Folk Theorem with Imperfect Public Information. *Econometrica*, 62(5):997.

Green, J. and Laffont, J.-J. (1977). Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods. *Econometrica*, 45(2):427.

Green, J. and Stokey, N. L. (1983). A Comparison of Tournaments and Contracts. *Journal of Political Economy*, 91(3):349–364.

Grossman, S. J. and Hart, O. D. (1983). An Analysis of the Principal-Agent Problem. *Econometrica*, 51(1):7–45.

Groves, T. (1973). Incentives in Teams. *Econometrica*, 41(4):617.

Halac, M., Lipnowski, E., and Rappoport, D. (2021). Rank Uncertainty in Organizations. *American Economic Review*, 111(3):757–86.

Holmström, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10(1):74–91.

Holmström, B. (1982). Moral Hazard in Teams. *Bell Journal of Economics*, 13(2):324.

Holmström, B. (2017). Pay for Performance and Beyond. *American Economic Review*, 107(7):1753–77.

Holmström, B. and Milgrom, P. (1990). Regulating Trade Among Agents. *Journal of Institutional and Theoretical Economics*, 146(1):85–105.

Holmström, B. and Milgrom, P. (1994). The Firm as an Incentive System. *American Economic Review*, 84(4):972–991.

Hurwicz, L. (1960). Optimality and Informational Efficiency in Resource Allocation Processes. In Arrow, K., Karlin, S., and Suppes, P., editors, *Mathematical Methods in the Social Sciences*. Stanford University Press.

Itoh, H. (1991). Incentives to Help in Multi-Agent Situations. *Econometrica*, 59(3):611.

Itoh, H. (1993). Coalitions, Incentives, and Risk Sharing. *Journal of Economic Theory*, 60(2):410–427.

Laffont, J.-J. and Martimort, D. (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Lambert, R. A. (1983). Long-Term Contracts and Moral Hazard. *Bell Journal of Economics*, 14(2):441.

Lazear, E. P. and Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5):841–864.

Ma, C., Moore, J., and Turnbull, S. (1988). Stopping Agents from 'Cheating'. *Journal of Economic Theory*, 46(2):355–372.

Mankiw, N. G. (1998). *Principles of Microeconomics*. Harcourt Publishers.

Marschak, J. (1955). Elements for a Theory of Teams. *Management Science*, 1(2):127–137.

Marschak, J. and Radner, R. (1972). *Economic Theory of Teams*. Yale University Press.

Maskin, E., Dasgupta, P., and Hammond, P. (1979). The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility. *Review of Economic Studies*, 46(2).

Matsushima, H. (1989). Efficiency in Repeated Games with Imperfect Monitoring. *Journal of Economic Theory*, 48(2):428–442.

Milgrom, P. and Roberts, J. (1990). Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities. *Econometrica*, 58(6):1255.

Milgrom, P. and Shannon, C. (1994). Monotone Comparative Statics. *Econometrica*, 62(1):157.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, 38(2):175–208.

Mirrlees, J. A. (1999). The Theory of Moral Hazard and Unobservable Behaviour: Part I. *Review of Economic Studies*, 66(1):3–21.

Mookherjee, D. (1984). Optimal Incentive Schemes with Many Agents. *Review of Economic Studies*, 51(3):433–446.

Myerson, R. B. (1979). Incentive Compatibility and the Bargaining Problem. *Econometrica*, 47(1):61.

Myerson, R. B. (1982). Optimal Coordination Mechanisms in Generalized Principal–Agent Problems. *Journal of Mathematical Economics*, 10(1):67–81.

Nalebuff, B. and Stiglitz, J. (1983). Prices and Incentives: Towards a General Theory of Compensation and Competition. *Bell Journal of Economics*, 14:21–43.

Rogerson, W. P. (1985). Repeated Moral Hazard. *Econometrica*, 53(1):69–76.

Topkis, D. M. (1978). Minimizing a Submodular Function on a Lattice. *Operations Research*, 26(2):305–321.

Vickrey, W. (1945). Measuring Marginal Utility by Reactions to Risk. *Econometrica*, 13(4):319.

Vives, X. (1990). Nash Equilibrium with Strategic Complementarities. *Journal of Mathematical Economics*, 19(3):305–321.