University of Ottawa

uOttawa

Doctoral Thesis

Essays on Incentives, Justice, and Mechanism Design

Author:

Ghislain Junior

Sidie

Thesis submitted to the University of Ottawa

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Economics

Faculty of Social Sciences

University of Ottawa

Declaration of Authorship

All chapters of this thesis are self-contained research articles. I, Ghislain Junior Sidie, acknowledge the contribution of Pr. Roland Pongou, my supervisor, for the research associated with the first and third chapters of this thesis. In both cases, his contribution is equal to my own, although the work in this thesis is in my own words.

Contents

List of Figures

Abstract

Ghislain Junior Sidie

Essays on Incentives, Justice, and Mechanism Design

Chapter 1 develops an axiomatic foundation for the classical problem of paying workers in settings where their actions are not observed by the employer. The latter observes the distribution of workers' abilities, has precise or imprecise information on the level of output, and demands fairness. We first characterize workers' pay thanks to a set of axioms developed under conditions of uncertainty and information asymmetry, assuming the employer is Bayesian. This characterization leads to a closed-form reward rule called the *informational Bayesian value* (IBV). Then, we analyze worker disappointment under the IBV and find that average disappointment is equal to zero for each worker. This latter result is robust to considering all inputs-based types of disappointment, including ex-post, interim, and ex-ante disappointment. The analysis implies that a worker is never disappointed in the long run when the pay rule utilized in an organization under information asymmetry is fair.

Chapter 2 analyzes agents' incentives in a production economy where there are both incomplete information and fairness considerations in the allocation of the economy's surplus. Agents hold private information (type or productivity), with each type profile defining a state of the economy. They choose between remaining inactive and participating in production by selecting from a finite set of available actions. This environment defines a new class of Bayesian economies. I show the existence of a pure-strategy Bayesian Nash equilibrium. Additionally, I uncover conditions that guarantee the efficiency and uniqueness of the equilibrium. Importantly, in the absence of fairness principles, an equilibrium may not exist. The findings are robust to several extensions of the basic model. I develop an application to mechanism design. When incentives are aligned with fairness, I uncover a mechanism that is always Pareto-efficient, and that is incentive-compatible and individually rational under certain conditions. This finding has implications for the design of mechanisms in production economies that are compatible with fairness and efficiency while providing incentives for truthful reporting of private information.

Chapter 3 considers the problem of evaluating the utility of participating in organizations that involve multiple interacting agents and are characterized by imperfect information in the sense that agents' types are uncertain ex-ante. We show that this problem admits a unique solution along the lines of expected utility theory, and that this solution is fully characterized by a natural extension of the notions of neutrality to ordinary risk and strategic risk introduced by Roth (1977). In addition, we analyze the implications of Roth's risk neutrality for incentives and equilibrium behaviors. We show the existence of rationalizable organizations, wherein uncertainty over agents' actions arises as an equilibrium play. We develop an application to the firm, showing that, under Roth's risk neutrality, workers' equilibrium behaviors do not depend on whether or not the firm possesses information on individual costs. Our second application is to two-sided assignment markets involving sellers and buyers of indivisible goods. We show that, under Roth's risk neutrality, there exists a pricing mechanism that achieves the expected utility of all agents when the optimal matching is unique, and that only achieves the expected utility of sellers when there is more than one optimal matching.

Acknowledgements

I am deeply grateful to God for blessing me with motivation, patience, and good health, which have been indispensable in completing this thesis. I would also like to take this opportunity to express my heartfelt gratitude to the many individuals whose invaluable support and contributions have made my Ph.D. journey a successful and fulfilling experience.

First and foremost, I wish to express my deepest gratitude to my advisor, Professor Roland Pongou. Roland, thank you for your continuous support, constant availability, extraordinary patience, and remarkable humility throughout this journey. Your guidance and encouragement have been invaluable every step of the way.

My sincere thanks also go to the thesis committee members - Professors Aggey Semenov, Gamal Atallah, and Eric Stephens - for their time, encouragement, insightful comments, and questions always aimed at improving the quality of all the chapters of my thesis. I would like to extend my appreciation to the external examiner, Professor Lars Ehlers, for his time and valuable comments on my thesis. This acknowledgment also goes to all faculty members in the Department of Economics at the University of Ottawa. Special mention and thanks to Professors Catherine Deri Armstrong, Vicky Barham, Yazid Dissou, Paul Makdissi, Vikram Manjunath, Myra Mohnen, and Louis-Philippe Morin. In addition, I would like to express my gratitude to the following University administrative staff: Alexie, Guilherme, Martine, and Nour.

I am deeply grateful to the seminar participants at the University of Ottawa and the World Bank Group for their thoughtful and constructive feedback. I owe special thanks to Pr. Bertrand Tchantcho, Pr.Jean-Baptiste Tondji, Dr. Nicolas Côté, Dr. Woubet Kassa, and Dr. Marie-Christelle Mabeu for their generous support and stimulating discussions, which greatly enriched the development of this thesis.

I gratefully acknowledge the PhD Completion Grant in memory of Professor Serge Nadeau, which provided crucial support during the writing of the final chapter of this thesis. This generous assistance afforded me the time and focus needed to bring the project to completion, and I deeply appreciate the recognition and encouragement it represents.

I would not have been able to complete this dissertation without the unwavering love and support of my wife, Daniele Raissa Sidie. Dani stood by my side with steadfast faith, providing encouragement and thoughtful reflection during moments of uncertainty and challenge. Her unwavering dedication and belief in this journey were instrumental in bringing this work to fruition. Our daughters, Melina Sidie Ngounou, Enora Johanne Sidie Mekueko, and Kayla Sidie Tchengwa, have been an incredible blessing, bringing immense joy and inspiration during the final stages of this journey. Their presence has been a source of boundless encouragement and motivation.

I honor the memory of my father, Jeremie Sidie, whose unwavering dedication and sacrifices made it possible for me to reach this level of education. I deeply wish he were still with us to witness the culmination of his efforts and the fulfillment of his dreams through the completion of this thesis. I have learned the valuable lessons of my mom, Charlotte Sidie, particularly to stay strong when things are hard. I gratefully acknowledge her continuing love, counsel, and support. I also thank all the other members of my family, particularly Christelle Tchabock Sidie, Thierry Ngounou Sidie, Dady Nganso Sidie, Gladys Tchakoute Sidie, and Arlette Njonkou Sidie.

Last but certainly not least, I would like to extend my heartfelt thanks to all my friends, especially Daryl Feleu, Daryl Julien, Victorien Kenfack, Adomako Kofi, Nyavor Kunche, Aoudou Njingouo, Raphael Tchoukui, Raoul Tekam, Jules Yaba, and Brice Yakam. Your support has been a source of strength throughout this long journey. I am also deeply grateful to my fellow economics doctoral colleagues, particularly Emile Ndoumbe and Yannick Ngongang for the mutual support and camaraderie we shared along the way.

To my late Dad: Jeremie Sidie, whose sacrifices laid the foundation for my doctoral journey

My Mum: Charlotte Sidie

My family: Kayla, Enora, Melina, and Daniele

General Introduction

Economic environments characterized by incomplete and asymmetric information are generally prone to inefficiencies stemming from adverse selection and moral hazard. Adverse selection occurs when agents possess private information before entering into an agreement, leading to possible misrepresentation or strategic behavior that distorts outcomes. In contrast, moral hazard arises when agents' actions after entering an agreement are unobservable, enabling behaviors that deviate from optimal effort or performance.

These challenges have been a central focus in the contract theory literature, which traditionally adopts a positive approach to address them. This approach involves designing contracts that align incentives to mitigate informational frictions and inefficiencies. Two critical constraints underpin this methodology: the participation constraint, which ensures agents willingly engage in the contract, and the incentive compatibility constraint, which incentivizes agents to truthfully reveal private information and act in accordance with the contract's objectives.

By systematically incorporating these constraints, the positive approach offers a robust framework for addressing inefficiencies in environments marked by incomplete and asymmetric information. Building on this foundation, the primary objective of this thesis is to develop a normative approach for coping with the challenges of adverse selection and moral hazard, which are inherent to such environments. This normative perspective seeks not only to mitigate these informational frictions but also to provide prescriptive guidance for designing solutions that align with broader economic and social objectives. Central to this approach is the requirement that contract design be fundamentally grounded in principles of distributive justice, ensuring fairness in outcomes while maintaining efficiency.

The principles of distributive justice require that the distribution of rewards be aligned with individual contributions. This implies that more productive workers should be paid more and equally productive workers should be paid identically. These principles have traditionally been explored within the framework of cooperative games. However, this model is constrained by strong assumptions that overlook essential aspects of real-world scenarios. A particularly limiting assumption is that of complete information, which requires all participants to share common knowledge about every element of the game. This thesis contributes to the literature by generalizing classical principles of distributive justice to more realistic models characterized by incomplete and asymmetric information. In addition, it examines how these generalized principles address the key challenges of adverse selection and moral hazard. By doing so, the thesis not only broadens the applicability of distributive justice theories but also provides insights into their effectiveness in mitigating inefficiencies and promoting equitable outcomes under informational constraints.

The first chapter addresses the problem of paying workers in a setting where their actions are not completely observed by their employer. We address this problem following an axiomatic approach by assuming that the employer has a preference for fairness and may have precise or imprecise information on the level of output produced. We formalize classical principles of fairness under conditions of uncertainty and information asymmetry, and use them to uniquely characterize a reward rule called the informational Bayesian value (IBV). The latter defines an output-based contract offered by an uninformed employer to employees under the assumption that the employer is fair and uses Bayesian updating to update his priors upon observing the output. Building on this analysis, we examine worker disappointment—the discrepancy between a worker's realized and expected pay—under the IBV. We find that the level of disappointment depends on the level of realized output. When output is sufficiently low, all workers may be disappointed. However, the output threshold below which a worker is disappointed varies across workers. Our main finding is that, if there is no constraint on output and workers have monetary utility, then average disappointment is zero for each worker. This result is found to be robust when considering all inputs-based types of disappointment, including ex-ante, interim, and ex-post disappointments. Ex-ante disappointment is disappointment a priori, that is, before any level of effort has been supplied. Interim disappointment is disappointment arising after a worker has supplied an effort level (but has no information about other workers' actions). Ex-post disappointment is disappointment arising after all workers have supplied their efforts. The economic implication of this analysis is that a potential worker is never disappointed a priori if the organization has a fair compensation pay scheme.

The second chapter explores workers' incentives under fair pay schemes in a production economy where agent productivity is private information. This chapter investigates whether incorporating fairness principles into surplus allocation can induce self-enforcing behaviors, incentivize truthful revelation of private information, and enhance overall efficiency.

While traditional studies on adverse selection and moral hazard focus primarily on optimizing outcomes under incomplete information, they often overlook fairness considerations.

This chapter bridges this gap by integrating fairness into a Bayesian framework to propose a novel pay scheme. This scheme not only aligns individual incentives with truthful behavior but also addresses equity concerns, thereby broadening the scope of traditional contract theory. The results show that fair pay schemes guarantee the stability of the economy by ensuring equilibrium existence. In addition, they incentivize workers to reveal their true productivity and supply their best effort under a *technological monotonicity* condition. This finding offers valuable insights for designing mechanisms in production economies that reconcile fairness and efficiency while promoting truthful reporting of private information.

Distributive justice refers to the perceived fairness of outcomes, which is deeply influenced by individuals' ethical perspectives and can vary widely across different people. The third chapter analyses the risk posture of decision-makers who have preferences for principles of distributive justice, particularly in contexts of uncertainty. It distinguishes between two types of risk: the ordinary risk involving the uncertainty that arises from the chance mechanism involved in lotteries, and the strategic risk involving the uncertainty that arises from the interaction of strategic agents in uncertain situations. By analyzing how these forms of risk affect decision-making under fairness considerations, the chapter uncovers a key result: decision-makers who adhere to basic principles of distributive justice have preferences represented by a unique cardinal utility function if and only if they are neutral to both ordinary and strategic risks. This neutrality ensures consistency in their ethical judgments and decision-making processes, regardless of the source of uncertainty. These insights provide a deeper understanding of how fairness considerations intersect with risk attitudes, offering a robust framework for integrating principles of justice into decision-making processes even in uncertain situations. This contributes to broader discussions on ethical decision-making and its practical implications in uncertain environments.

Chapter 1

Bayesian Fairness: Reward Rule and Disappointment under Information Asymmetry

1.1 Introduction

How should workers be compensated when their efforts are not observed? The problem of paying workers when their actions are not completely observed by their employer is classical. This problem has been traditionally studied in contract theory following a positive approach. In this paper, we follow an axiomatic approach by assuming that the employer has a preference for fairness and may have precise or imprecise information on the level of output produced. We formalize basic principles of fairness under conditions of uncertainty and information asymmetry, and use them to uniquely characterize a reward rule called the *informational Bayesian value* (IBV). The latter defines an *output-based contract* offered by an uninformed employer to employees under the assumption that the employer is fair and uses Bayesian updating to update his priors upon observing the output. Building on this analysis, we examine worker disappointment—the discrepancy between a worker's realized and expected pay—under the IBV. For robustness, we consider all forms of disappointment—including ex-post, interim, and ex-ante disappointment—which are defined based on the timing of input supply by workers.

The notion of fairness we follow refers to basic principles of distributive justice in organizations that produce private (or market) goods. It is defined as the perceived fairness of outcomes (Adams, 1965). It differs from procedural justice—the perceived fairness of the process by which outcomes are determined (Lind and Tyler, 1988). It is argued that these principles are more likely to emerge as a consensus in settings where individuals are unaware of their respective positions, i.e., behind a veil of ignorance as termed by Rawls (1971). It has been found that unfair outcomes in organizations are associated with downside effects such as lower job performance (Gilliland, 1994) and poor work attitudes (Folger and Cropanzano, 2001). It is also argued that fair pay schemes are desirable not only because they are ethical, but also because they are generally free from any disputation by workers, and they may reduce free-riding behaviors in teamwork (Holmstrom, 1982). Therefore, identifying fair pay rules should be a major concern for policymakers and employers. This is even more important in environments where employers cannot fully observe workers' actions, given the high prevalence of such environments in real life. In such situations, it is not clear how fairness issues should be addressed.

To address these questions, we introduce a model of an organization with incomplete and asymmetric information. It is defined as a tuple $\mathcal{F} = (N, (X_i)_{i\in N}, f, \rho)$, where N is a finite set of workers, $X_i \subseteq \mathbb{R}$ is a compact and convex set of effort (or ability) levels that worker i can supply, f is a continuous function representing the technology (which could be the production function or the surplus function) of the organization, and it maps each effort profile $x = (x_1, x_2, ..., x_n) \in \prod_{i=1}^{n} X_i$ to a measurable output $f(x) \in \mathbb{R}$, and $\rho = (\rho_i)_{i\in N}$ is a profile of probability density functions describing the ability distributions of workers (or the *employer's beliefs* over workers' abilities). The employer does not know the effort level supplied by a worker, but he knows the probability distribution from which his ability is drawn.[1]

After workers have supplied an effort profile $x \in X$, we assume that the employer, lacking information on x, may have precise or imprecise information about the output produced $y = f(x)$. We denote by I the employer's perception (information set) about the output y. It is formally defined as a non-empty subset of the set of all possible outputs that can be produced. If the employer observes the output accurately, then I is a singleton containing the level of output y; otherwise, I is a non-empty subset of $\mathbb{R}$ different from a singleton.[2]

With this model, we precisely address the following questions:

1. What pay can a worker expect given I?

2. What implications do *expected* and *realized* pays have for worker disappointment?

We answer the first question by assuming that the employer has a preference for fairness. We define basic ideals of fairness that extend the classical axioms to conditions of uncertainty and asymmetric information. These principles are stated as follows: (1) efficiency a posteriori: no amount of the expected output should be wasted. These axioms are quite intuitive; (2)

[1] The employer assumes that the realized ability of a worker determines his action or effort supply.

[2] As I becomes larger, information imprecision regarding output increases. Interestingly, imprecision may also be interpreted as output being a random variable whose distribution is unknown to the employer and whose support is I.

bilateral fairness a posteriori: each pair of workers contributes equally to each other's pay. They are inspired by the classical works of Shapley (1953) and Myerson (1977) undertaken in a full information context. However, because our environment significantly differs from the full information environment considered in these pioneering works, our formalization of these principles differs considerably. We characterize a unique fair pay scheme, called the *informational Bayesian value* (IBV) (Theorem 1). When the output is known accurately, that is I is a singleton, the informational Bayesian value is simply a sharing rule of the surplus associated with this observed output among the workers. However, our finding is more general as the information set I can be any subset of the set of achievable output levels.

Accordingly, the informational Bayesian value can be considered as an *output-based contract* between an uninformed employer and several employees in a competitive environment wherein competition drives the employer's profit to zero (see Example 2). In this sense, our work complements the literature on output-based contracts by following an axiomatic approach, which departs significantly from the positive approach followed in other studies (Akerlof (1970), Holmstrom (1982), Grossman and Hart (1983)), and the robustness-based approach (Carroll (2015), Miao and Rivera (2016), Dai and Toikka (2022)).

From the treatment of the first question, we derive implications for how a worker's *expected* pay (that is, pay before output realization) and *realized* pay (that is, pay obtained after output is realized and known precisely) affect disappointment, which answers our second question. A worker is disappointed if his realized pay is lower than his expected pay; otherwise, he is happy. We find that the level of disappointment depends on the level of realized output. When output is sufficiently low, all workers may be disappointed. However, the output threshold below which a worker is disappointed varies across workers. Our main finding here is that, if there is no constraint on output and workers have monetary utility, average disappointment is zero for each worker (Theorem 2). This result is found to be robust when considering all inputs-based types of disappointment, including *ex-ante*, *interim*, and *ex-post* disappointments. Ex-ante disappointment is disappointment a priori, that is, before any level of effort has been supplied. Interim disappointment is disappointment arising after a worker has supplied an effort level (but has no information about other workers' actions). Ex-post disappointment is disappointment arising after all workers have supplied their efforts. The economic implication of this analysis is that a potential worker is never disappointed a priori if the organization has a fair compensation pay scheme.

Our model has a wide variety of applications owing to the fact that the definition of technology in our setup follows the neoclassical approach. In this paper, we focus on two applications to classical economic problems. In particular, we develop new applications to profit-maximizing firms facing uncertainty in the supply of inputs, and to Cournot oligopoly. Uncertainty in input supply is a prevalent issue in certain industries. This is the case, for instance, when key inputs (e.g., crops) are supplied by countries likely to face political instability or high variability in environmental conditions, or by individuals with changing health conditions. We show how the informational Bayesian value can be used to share profit in these settings.

Our paper can also be viewed as contributing to the literature on the theories of value. These theories have classically been developed under the assumption of full information and have been applied to various classes of problems, including operations management and supply chains (Gopalakrishnan et al. (2021)), cost allocations (Dubey (1982)), fair division (Moulin, 1992), contract design (Winter, 2002), political power measurement (Shapley and Shubik (1954), Freixas (2005), Freixas, Marciniak, and Pons (2012), Pongou and Tchantcho (2021)), network centrality measurement (Grofman and Owen (1982), Pongou and Tondji (2018)), queueing problems (Maniquet, 2003), input valuation in discrete settings (Pongou and Tondji, 2018), unfairness and income inequality (Aguiar, Pongou, and Tondji (2018), Aguiar et al. (2019)), among others. However, extensions of these theories to problems involving asymmetric information and informational constraints have received little attention. This is a gap our paper begins to fill. Also, our setup, which augments the model of a neoclassical firm by incorporating uncertainty in input supply or beliefs, addresses several limitations of classical models of production and is likely to inspire a wide variety of applications. In this respect, our applications to profit-maximizing firms facing uncertain input supply and the Cournot oligopoly model are new. Finally, our study can be seen as laying the groundwork for a normative approach to contract theory. We do not address the problem of workers' incentives under fairness, which is an interesting question for future research. Addressing this question requires that the normative approach be developed in the first place, which is what we have done by introducing the IBV along with a supportive axiomatic foundation.

The rest of this paper proceeds as follows. Section 3.1 describes the model of an organization with incomplete information and introduces other preliminary concepts. In section 1.3, we extend classical fairness principles under conditions of incomplete information and uniquely characterize the *informational Bayesian value* (IBV). Worker disappointment is investigated in section 1.4. Applications are provided in section 1.5. Section 1.6 concludes.

1.2 The model of an organization with incomplete and asymmetric information

An organization with incomplete information is a list $\mathcal{F} = (N, X, f, \rho)$ where

- $N = \{1, 2, ..., n\}$ is a finite set of workers;
- $X = \prod_{i=1}^{n} X_i$ is the set of effort profiles, where $X_i = [0, \overline{x_i}]$ is a compact and convex set of the effort levels of worker i;
- f is a continuous function representing the technology (which could be the production function or the surplus function) of the organization, mapping each effort profile $x = (x_1, ..., x_n) \in X$ to a real number output $f(x)$;
- $\rho = (\rho_i)_{i \in N}$ is a profile of probability distributions of workers' abilities, where for each worker i, ρ_i is a probability density function whose support is the set X_i.[3] Formally, we have $\rho : X \longrightarrow [0, +\infty)$, with $\int_X \rho(x)dx = 1$ and for each $i \in N$ and $x_i \in X_i$, $\rho_i(x_i) = \int_{X_{-i}} \rho(x_i, x_{-i})dx_{-i}$ where $X_{-i} = \prod_{j=1, j \neq i}^{n} X_j$ and $x_{-i} \in X_{-i}$. If workers' abilities are independently distributed, then the probability density function is given by $\rho(x) = \prod_{i \in N} \rho_i(x_i)$.

Information asymmetry. Information asymmetry arises from the fact that the employer does not observe the effort supply (x) of workers; each worker's action is only known to that worker. After workers have chosen an effort profile $x \in X$, we assume that the employer observes the output level $y = f(x)$ with precision or imprecision. In general, the perception of output by the employer is described by a set I of real numbers, which is the support of a probability distribution that is unknown a priori.[4] Therefore, the size of the set I gives an indication of how precise the employer's information about the output level y is. The set I is in general situated between two extreme cases. The first extreme case is when the employer has no information at all on the level of output that is produced, in which case $I = \mathbb{R}$. The other extreme case is when the employer knows exactly the level of output produced, in which case $I = \{y\}$. Other intermediate cases, where I is neither the real line nor a singleton, are informative of the extent to which the employer is aware of the level of output produced. For example, $I = [4, 7]$ means that the employer knows that the level of output produced is

[3]The employer observes ρ, but does not observe x. The realization of ability determines a worker's effort supply, which is why we assume for simplicity that the support of ρ_i is X_i for each worker i. In this sense, ρ_i is also interpreted as the employer's *belief* over worker i's ability. We also note that ρ can also be a joint cumulative distribution function over the set X.

[4]In this sense, output can be interpreted as being a stochastic outcome with an unknown distribution whose support is I.

between 4 and 7, but he doesn't have a precise idea of what the output level is. Throughout the paper, we shall refer to I as the employer's information set.

We can summarize the timing of the production process as follows:

(i) each worker $i \in N$ chooses an effort level $x_i \in X_i$, giving rise to an effort profile $x \in X$;

(ii) the employer, knowing ρ but not x, perceives the output $y = f(x)$ by forming $I \subseteq \mathbb{R}$;

(iii) workers' payoffs are received.

For any information set $I \subseteq \mathbb{R}$, we denote by $I_f = \{x \in X : f(x) \in I\}$ the set of effort profiles whose images belong to I. For technical reasons, we require that I_f be a convex set or a union of convex sets.

Other interpretations. We note that the model of an organization with incomplete information has several other interpretations depending on the context. For example, N can be the set of firms in an industry producing a homogeneous product, and X_i the set of outputs firm i can produce. In this case, the technology f can be the industry's profit or surplus function.

Output-based reward rule or contract. One main goal of this paper is to determine an output-based reward rule or contract that the employer can use to pay workers. This is done in the next section by assuming that the employer is Bayesian and demands fairness.

1.3 Bayesian fairness: an axiomatic foundation

This section addresses the question of what pay can a worker expect from a fair employer when the latter only observes the distribution of workers' abilities and has precise or imprecise information about the level of output produced. The notion of fairness here refers to basic principles of distributive justice that a pay scheme should satisfy. In a perfect information setting, Myerson (1977) proposes intuitive axioms of fairness (efficiency, bilateral fairness (balanced contribution)). We generalize these basic axioms under conditions of uncertainty and information asymmetry between the workers and the employer. The generalized axioms are called *efficiency a posteriori* and *bilateral fairness a posteriori*. We show that these two axioms uniquely characterize a pay scheme that we call the *informational Bayesian value (IBV)*; the latter is an **output-based reward rule or contract** that an employer can use to pay workers under conditions of information asymmetry.

1.3.1 Axioms of fairness under asymmetric information

We start by defining the notion of an output-based pay scheme or contract.

Definition 1 : *A pay scheme is a function ϕ which maps any organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$ and any information set I to a vector $(\phi_i(\mathcal{F}|I))_{i\in N}$, where $\phi_i(\mathcal{F}|I)$ is a real number representing the pay of worker i conditional on the employer's information set I.*

If no ambiguity is possible, we will denote $\phi_i(\mathcal{F}|I)$ and $\phi_i(f|I)$ interchangeably.

We also define the notion of an organization with incomplete information in the absence of a worker.

Definition 2 : *Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information and $i \in N$ a worker. The organization $\mathcal{F}$ in the absence of worker i is the organization denoted $\mathcal{F}^{-i} = (N \setminus \{i\}, X_{-i}, f^{-i}, \rho_{-i})$, where $X_{-i} = \prod_{j\in N\setminus\{i\}} X_j$, $\rho_{-i} = (\rho_j)_{j\in N\setminus\{i\}}$ and f^{-i} is defined for any $x_{-i} \in X_{-i}$ by: $f^{-i}(x_{-i}) = f(0, x_{-i})$.*

An organization with incomplete information in which a worker is absent is equivalent to having that worker supplying zero effort level at every effort profile. We also define an active worker below.

Definition 3 : *Let $x \in X$ be an effort profile. A worker $i \in N$ is active at x if he supplies a positive effort level, that is, if $x_i > 0$. If $x_i = 0$, then we say that worker i is inactive at x. We denote by $|x| = |\{i \in N : x_i > 0\}|$ the number of active workers at x.*

The following definition introduces a binary relation over the set of effort profiles.

Definition 4 : *Let $x, a \in X$ be two effort profiles and $i \in N$ a worker. a is induced by x, denoted by $a \trianglelefteq x$, if $\forall j \in N$, $a_j \neq 0 \Longrightarrow a_j = x_j$. In other words, a is induced by x if every active worker at a supplies the same effort level as at x.*[5] *a is strictly induced by x, denoted $a \triangleleft x$, if $a \trianglelefteq x$ and $a \neq x$. a is strictly induced by x via worker i, denoted $a \triangleleft_0^i x$, if $a \triangleleft x$ and $a_i = 0$.*

Next, we define the notion of a worker's marginal contribution.

Definition 5 : *Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information, $i \in N$ a worker, and $x, a \in X$ two effort profiles such that $a \triangleleft_0^i x$. The marginal contribution of worker i relative to the effort profiles a and x is given by:*
$mc(i, f, a, x) = f(a + x_i e_i) - f(a)$, where e_i is the i-th unit vector $(0, ..., 0, 1, 0, ..., 0)$.

[5] For instance we have $(0, 1, 0, 0, 5, ..., 7) \trianglelefteq (2, 1, 5, 0, 5, ..., 7)$.

We are now ready to formalize two basic principles of distributive justice under uncertainty and asymmetric information.

Efficiency a posteriori: A pay scheme ϕ is efficient a posteriori if for any organization with incomplete information , $\mathcal{F} = (N, X, f, \rho)$ and any employer's information set I,

$$\sum_{i \in N} \phi_i(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) f(x)\, dx}{\int_{I_f} \rho(x)\, dx}.$$

Bilateral fairness a posteriori: A pay scheme ϕ satisfies bilateral fairness a posteriori if for any organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$, any information set I, and any workers $i, j \in N$,

$$\phi_i(\mathcal{F}|I) - \phi_i(\mathcal{F}^{-j}|I^{-j}) = \phi_j(\mathcal{F}|I) - \phi_j(\mathcal{F}^{-i}|I^{-i}),$$

where for all $k = i, j$, I^{-k} is such that: $f^{-k}(x_{-k}) \in I^{-k}$ if and only if there exists $x_k \in X_k$ such that $f(x_k, x_{-k}) \in I$.

The efficiency a posteriori principle says that the expected output of the organization should be entirely shared among the different workers, where the expectation is based on the employer's updated beliefs. Bilateral fairness a posteriori, which extends an axiom introduced by Myerson (1977), requires the contribution of a worker j to a worker i's pay to be equal to the contribution of worker i to worker j's pay. This indeed reflects the notion of fairness.

1.3.2 Main result: an output-based reward rule

Below, we show that the two axioms of fairness defined above characterize a unique pay scheme called the informational Bayesian value.

Theorem 1 : *One and only one pay scheme satisfies the axioms of efficiency a posteriori and bilateral fairness a posteriori. This pay scheme is called the informational Bayesian value, denoted IBV, and defined for every organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$, any information set I, and every worker $i \in N$ by:*

$$IBV_i(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) Sh_i(f, x)\, dx}{\int_{I_f} \rho(x)\, dx}, \tag{1.1}$$

where the function $Sh_i(f,x)$ is defined for all $x \in X$ by:

$$Sh_i(f,x) = \sum_{a \triangleleft_0^i x} \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} mc(i,f,a,x) \tag{1.2}$$

The informational Bayesian value, $IBV_i(\mathcal{F})$, of a worker i can be interpreted as the conditional expectation of the weighted marginal contribution of i in the production process of the organization, where the expectation is based on the employer's updated beliefs. The beliefs have been updated to incorporate the informational constraint faced by the employer. For a given effort profile x, the value $Sh_i(f,x)$ is the average marginal contribution of worker i over the set of effort profiles that are strictly induced by x via worker i. So, $Sh_i(f,x)$ can be interpreted as the average contribution of worker i to output $f(x)$. It is the alternative pay scheme of the informational Bayesian value in the setting where the employer perfectly observes the effort levels of workers. Indeed, if the employer observes an effort profile x then he will provide each worker i with the pay $Sh_i(f,x)$. This pay scheme generalizes the classical Shapley value (Shapley, 1953) in transferable-utility environments. In fact, if the effort level set of each worker is the pair $\{0,1\}$, then the classical Shapley value of a transferable-utility game $v : 2^N \to \mathbb{R}$, is simply $Sh_i(f,x)$, where $x = (1,1...,1)$ and $f : y \mapsto f(y) = v(\{i \in N : y_i = 1\})$. Therefore, $Sh_i(f,x)$ will be called the Shapley pay of worker i at the effort profile x under technology f. Unlike the classical Shapley value, the informational Bayesian value is defined over convex sets and does not assume that the effort of each worker is known (or given) a priori.

If the employer has no information about the level of output produced, that is $I = \mathbb{R}$, the informational Bayesian value $IBV_i(\mathcal{F}|\mathbb{R})$ will be simply called the informational ex-ante value (IEV), denoted by $IEV(\mathcal{F})$, and defined for all $i \in N$ by:

$$IEV_i(\mathcal{F}) = \int_0^{\overline{x_n}} \dots \int_0^{\overline{x_1}} \rho(x)\, Sh_i(f,x)\, dx_1 dx_2 \dots dx_n \tag{1.3}$$

The IEV can be interpreted as the fair pay each worker can expect prior to joining the organization. It can also be viewed as the initial contract offered to workers before a given performance has been realized. This contract is generally based on the employer's beliefs drawn from workers' observable characteristics such as age, gender, level of education, past experiences, and so on.

Notation 1 : *$IBV_i(\mathcal{F}|I)$ and $IEV_i(\mathcal{F})$ will often be denoted by $IBV_i(f|I)$ and $IEV_i(f)$, respectively.*

For clarity of exposition, the proof of Theorem 1 and all the subsequent results are provided in the Appendix.

Example 1 : *Consider an organization with incomplete information consisting of two workers. The probability density function, ρ, is such that the effort levels x_1 and x_2 are independently and uniformly distributed on $[0,1]$ and $[0,2]$, respectively. So, we have $\rho(x) = \frac{1}{2}$ for all $x \in X$. Let f be a linear technology defined by*

$$f(x_1, x_2) = 2x_1 + 3x_2.$$

Let us determine the informational Bayesian value for different employer information sets.

- *Assume the employer has no information on output, that is $I = \mathbb{R}$. For every $i = 1, 2$, we have: $IEV_i(f) = \frac{1}{2}\int_0^2\int_0^1 Sh_i(f,x)\ dx_1dx_2$. For all $x \in [0,1]\times[0,2]$, $Sh_1(f,x) = \frac{1}{2}(f(x_1,0) - f(0,0)) + \frac{1}{2}(f(x_1,x_2) - f(0,x_2)) = 2x_1$ and $Sh_2(f,x) = 3x_2$. This implies that $IEV_1(f) = 1$ and $IEV_2(f) = 3$. Therefore, workers 1 and 2 can expect a pay of 1 and 3, respectively, if the employer has no information on the output produced.*
- *Assume the employer has imprecise information about the output and thinks that the level of output produced is between 4 and 8, that is $I = [4,8]$. Figure 1.1 depicts the set $I_f = \{(x_1,x_2) \in [0,1]\times[0,2] : 4 \leq 2x_1 + 3x_2 \leq 8\}$ of effort profiles that satisfy this constraint.*

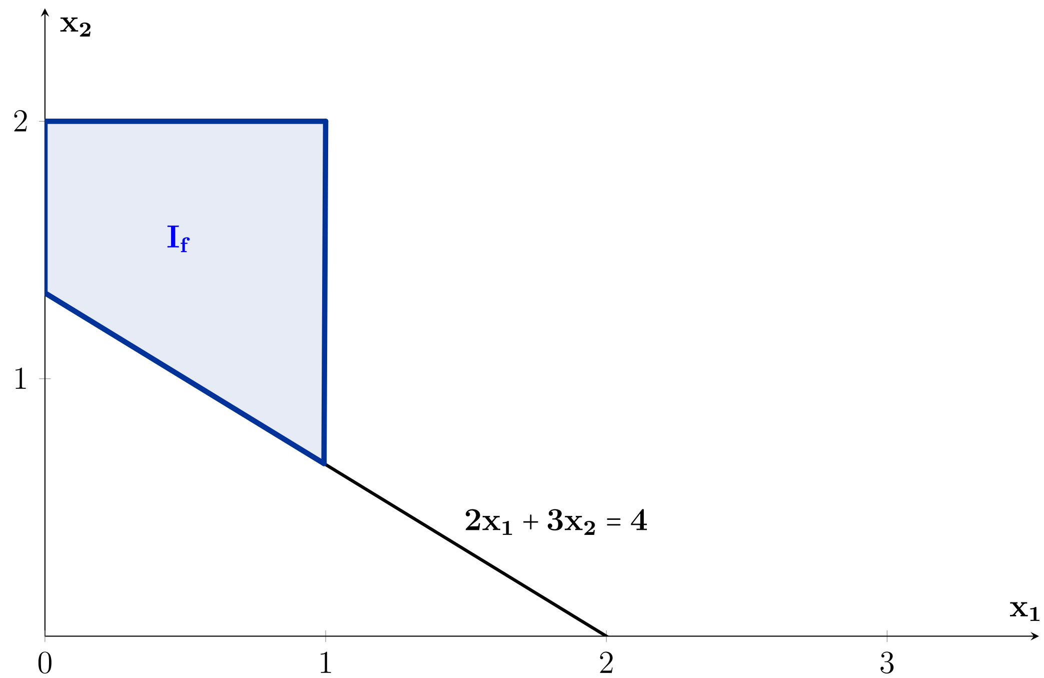

Figure 1.1: The set I_f of effort profiles producing a level of output between 4 and 8.

We have that $P(I) = \int_{I_f} \rho(x)\ dx = \frac{1}{2}\int_0^1\int_{\frac{4-2x_1}{3}}^{2} dx_2dx_1 = \frac{1}{2}$. It follows that:

$IBV_1(f|I) = \int_0^1 \int_{\frac{4-2x_1}{3}}^{2} 2x_1 dx_2 dx_1 = \frac{10}{9}$ *and* $IBV_2(f|I) = \int_0^1 \int_{\frac{4-2x_1}{3}}^{2} 3x_2\, dx_2 dx_1 = \frac{40}{9}$

Therefore, $IBV_1(f|I) > IEV_1(f)$ *and* $IBV_2(f|I) > IEV_2(f)$*. This example suggests that workers will experience an increase in their pay if the employer has imprecise information and thinks that the level of output produced by workers is between 4 and 8.*

- *Assume the employer has precise information about output and knows that the level of output produced is equal to 7, that is* $I = \{7\}$*. The set of effort profiles* I_f *that satisfy the informational constraint is given by:*

$$I_f = \{(x_1, x_2) \in [0,1] \times [0,2] : 2x_1 + 3x_2 = 7\} = \{(1 - \frac{1}{2}t, \frac{5}{3} + \frac{1}{3}t) \in [0,1] \times [0,2] : t \in [0,1]\}$$

We find that:

$$IBV_1(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) Sh_1(f,x)\, dx}{\int_{I_f} \rho(x)\, dx} = \frac{3}{2}.$$

$$IBV_2(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) Sh_2(f,x)\, dx}{\int_{I_f} \rho(x)\, dx} = \frac{11}{2}.$$

Therefore, both workers are better-off relative to the situation where the employer has no information on the level of output produced. However, this result cannot be generalized to any level of output; in particular, it is overturned for relatively low levels of output. This is the object of the next section.

1.4 Worker disappointment

In this section, we analyze the discrepancy between a worker's expected pay (that is, pay before output realization) and realized pay (that is, pay after output realization). Here, we assume that the employer has precise information on the level of output produced ($I = \{y\}$). When output is not realized yet, this corresponds to the case where the employer has no information at all on the output ($I = \mathbb{R}$). Disappointment is defined in the psychology literature as the feeling of dissatisfaction that follows the failure of expectations (Higgins, 1987). Consistent with this definition, we measure the disappointment of a worker i in an organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$ when an output level y is observed by the function $D_i(\mathcal{F}, y) = IBV_i(\mathcal{F}|I(y)) - IEV_i(\mathcal{F})$ (that is, realized pay minus expected pay), where $I(y) = \{y\}$. Worker i is: disappointed if $D_i(\mathcal{F}, y) < 0$ (that is, realized pay is smaller than expected pay); not disappointed if $D_i(\mathcal{F}, y) = 0$; and happy if $D_i(\mathcal{F}, y) > 0$. We find that disappointment depends on realized output, and it significantly varies across workers for different output levels. We illustrate this using the example below.

Example 2 : *Consider the organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$ analyzed in Example 1, where $N = \{1, 2\}$, $X = [0, 1] \times [0, 2]$, $f(x) = 2x_1 + 3x_2$ and $\rho(x) = \frac{1}{2}$. Now assume that the employer perfectly observes an output level $y \in [0, 8]$. We analyze how worker disappointment depends on realized outputs and varies between the two workers. First, let us determine how an uninformed employer who only observes output will pay the workers. This is equivalent to designing a contract under incomplete information, wherein the employer only observes the output produced but cannot observe workers' actions. Consistent with our theory, we assume the employer uses the informational Bayesian value associated to the information $I(y) = \{y\}$. In this case, the set of effort profiles that satisfy the constraint is given by $I_f(y) = \{(x_1 x_2) \in [0, 1] \times [0, 2] : 2x_1 + 3x_2 = y\}$. We can show that, for each worker, the informational Bayesian value associated to this information set is given by:*

$$IBV_1(\mathcal{F}|I(y)) = \frac{\int_{I_f(y)} \rho(x) Sh_1(f, x)\, dx}{\int_{I_f(y)} \rho(x)\, dx} = \begin{cases} \frac{y}{2} & if \quad y \in [0, 2] \\ 1 & if \quad y \in [2, 6] \\ \frac{y}{2} - 2 & if \quad y \in [6, 8] \end{cases}$$

$$IBV_2(\mathcal{F}|I(y)) = \frac{\int_{I_f(y)} \rho(x) Sh_2(f, x)\, dx}{\int_{I_f(y)} \rho(x)\, dx} = \begin{cases} \frac{y}{2} & if \quad y \in [0, 2] \\ y - 1 & if \quad y \in [2, 6] \\ \frac{y}{2} + 2 & if \quad y \in [6, 8] \end{cases}$$

Figure 1.2 depicts the informational Bayesian value (realized pay) of each worker as a function of the observed output y, as well as the workers' informational ex-ante values (expected pays) in the absence of any information about the output produced. We remark that the relationship between the informational Bayesian value and the informational ex-ante value depends on the observed output, which can be used by the employer as a signal to retrieve the actions chosen by the workers. Interestingly, a Bayesian employer who observes an output level of $y \in [0, 8]$ rewards workers equally if output is sufficiently low ($y \in [0, 2]$). For intermediate values of output ($y \in [2, 6]$), he gives the less productive worker (worker 1) his ex-ante informational value, and the more productive worker (worker 2) obtains the residual output. For sufficiently large values of output ($y \in [6, 8]$), both workers get a payment higher than their informational ex-ante value, with the payment of the more productive worker departing *the most from his informational ex-ante value.*[6] *It follows that the informational Bayesian value can be likened to an incentive compatibility contract that provides strong incentives to high-ability workers to choose a sufficiently high effort level in order to distinguish themselves from low-ability workers. As a result, low-ability workers must respond by choosing a higher effort level if they want to experience an increase in their pay relative to the informational ex-ante value. This example highlights a separating fair contract in the sense that the two workers receive different pay; this is more attractive to the high-ability worker when compared to the situation where the employer doesn't observe the output produced. This is consistent with the model of signaling and the principal-agent problem with hidden information (Akerlof (1970), Spence (1974), Hart and Holmström (1987), Varian (1992)).*

It follows that the low-ability worker will be disappointed if realized output is less than 2. He will not be disappointed if realized output is between 2 and 6, and will be happy if it is greater than 6. As for the high-ability worker, he will be disappointed if realized output is smaller than 4 and will be happy if it is greater than 4. So, both workers are disappointed if output is sufficiently low and happy if it is sufficiently high.

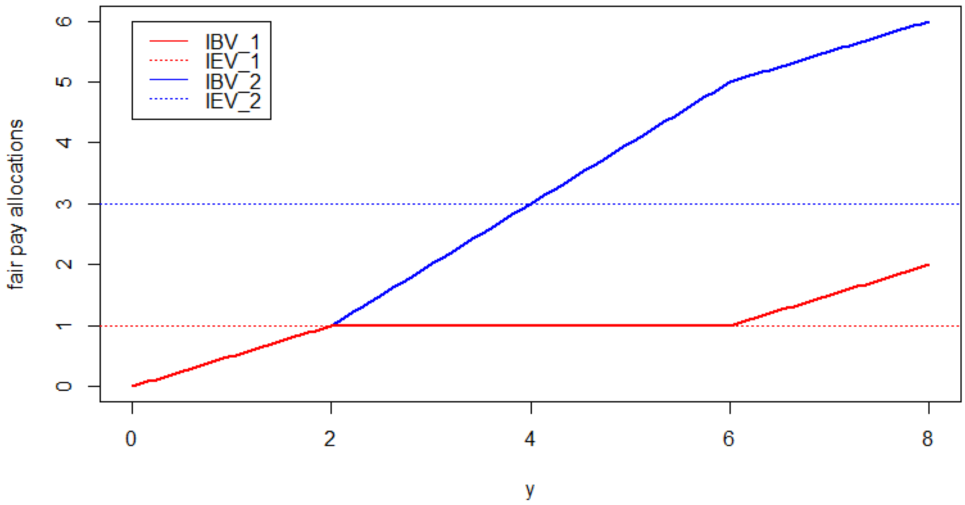

Figure 1.2: informational Bayesian value of workers associated to the informational constraint $I(y) = \{y\}$ as a function of the observed output y.

[6]In our framework of an organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$, worker i is said to be more productive than worker j if $X_j \subseteq X_i$ and for all $x, a \in X$ such that $a \lhd_0^i x$, we have $mc(i, f, a, x) \geq mc(j, f, \tau_{ij}(a), \tau_{ij}(x))$. That is, if the maximum effort level of j is less than or equal to the maximum effort level of i and the marginal contribution of i relative to any pair of effort profiles is greater than or equal to the marginal contribution of j relative to a corresponding pair of effort profiles in which the effort levels of i and j are swapped.

Despite the fact that worker disappointment depends on realized output level and varies between workers, we show below that, for each worker, the average disappointment is zero. Before stating this result, let us define the notion of a partition of informational constraints.

Definition 6 : *Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information. A set $\mathcal{I}_f$ of informational constraints forms a partition of informational constraints if, for every effort profile $x \in X$, there exists a unique informational constraint $I \in \mathcal{I}_f$ such that $f(x) \in I$. Simply put, $\mathcal{I}_f$ is a partition of informational constraints if each effort profile belongs to exactly one element of $\mathcal{I}_f$.*

Consider the following illustrative example.

Example 3 : *Let $\mathcal{R}_f = \{y \in \mathbb{R} : \exists x \in X, f(x) = y\}$ be the set of realizable outputs of f. The set $\mathcal{K}_f = \{I(y) = \{y\} : y \in \mathcal{R}_f\}$ forms a partition of informational constraints. For each $y \in \mathcal{R}_f$, if $T^{1y} = (-\infty, y)$ and $T^{2y} = [y, +\infty)$, then the set $\mathcal{T}_f^y = \{T^{1y}, T^{2y}\}$ also forms a partition of informational constraints.*

Now, we are ready to state the next result.

Theorem 2 : *For any organization with incomplete information, $\mathcal{F} = (N, X, f, \rho)$, any partition of informational constraints $\mathcal{I}_f$, and any worker $i \in N$, we have:*

$$\mathbb{E}_{\mathcal{I}_f}\left(IBV_i(\mathcal{F}|I)\right) = IEV_i(\mathcal{F}). \text{ In particular, } \mathbb{E}\left(D_i(\mathcal{F}, y)\right) = 0$$

This result states that the informational ex-ante value of any organization with incomplete information is the average of the informational Bayesian value over any partition of informational constraints. In particular, taking the partition of informational constraints to be $\mathcal{K}_f = \{I(y) = \{y\} : y \in \mathcal{R}_f\}$ yields: $\mathbb{E}\left(D_i(\mathcal{F}, y)\right) = \mathbb{E}_{\mathcal{K}_f}\left(IBV_i(\mathcal{F}|I(y)) - IEV_i(\mathcal{F})\right) = \mathbb{E}_{\mathcal{K}_f}\left(IBV_i(\mathcal{F}|I(y))\right) - ESV_i(\mathcal{F}) = 0$ for all workers i, which means that the average level of disappointment for each worker is null.

Example 4 : *For illustration, let us continue with Example 1 by considering the partition of informational constraints to be $\mathcal{T}_f^4 = \{T^{14}, T^{24}\}$, where $T^{14} = [0, 4[$ and $T^{24} = I = [4, 8]$. We have shown that $IBV_1(\mathcal{F}|T^{24}) = \frac{10}{9}$ and $IBV_2(\mathcal{F}|T^{24}) = \frac{40}{9}$. Now let us compute $IBV_i(\mathcal{F}|T^{14})$, $i = 1, 2$. Figure 1.3 depicts the set $T_f^{14} = \{(x_1, x_2) \in [0, 1] \times [0, 2] : 2x_1 + 3x_2 < 4\}$ of effort profiles that satisfy the constraint T^{14}.*

Again, we have $P(T^{14}) = \int_{I_f} \rho(x)\, dx = \frac{1}{2}$. As a result, we get that:

$$IBV_1(\mathcal{F}|T^{14}) = \int_0^1 \int_0^{\frac{4-2x_1}{3}} 2x_1 dx_2 dx_1 = \frac{8}{9}, \text{ and}$$

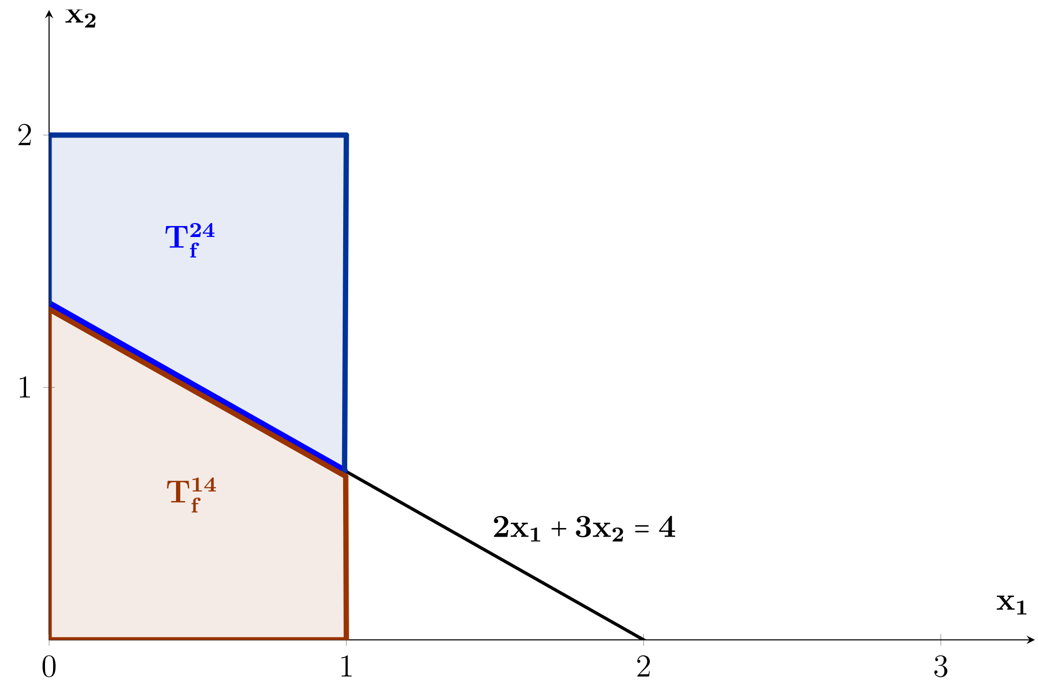

Figure 1.3: The sets T_f^{14} and T_f^{24} of effort profiles producing an output strictly less than 4 and greater than or equal to 4, respectively.

$$IBV_2(\mathcal{F}|T^{14}) = \int_0^1 \int_0^{\frac{4-2x_1}{3}} 3x_2 dx_2 dx_1 = \frac{14}{9}$$

In the absence of constraints, the contribution of each worker to the production process of the organization is $IEV_1(\mathcal{F}) = 1$ *and* $IEV_2(\mathcal{F}) = 3$. *Now, if the organization faces an informational constraint that requires output to be between 0 and 4, then the contribution of each worker becomes* $IBV_1(T^{14}) = \frac{8}{9}$ *and* $IBV_1(T^{14}) = \frac{14}{9}$. *Therefore, workers perform more poorly under the informational constraint* $T^{14} = [0, 4[$ *than when there are no constraints. Moreover, the following equalities hold:*

$$\mathbb{E}_{\mathcal{T}_f^4}(IBV_1(\mathcal{F}|I)) = P(T^{14})IBV_1(\mathcal{F}|T^{14}) + P(T^{24})IBV_1(\mathcal{F}|T^{24}) = IEV_1(\mathcal{F})$$

$$\mathbb{E}_{\mathcal{T}_f^4}(IBV_2(\mathcal{F}|I)) = P(T^{14})IBV_2(\mathcal{F}|T^{14}) + P(T^{24})IBV_2(\mathcal{F}|T^{24}) = IEV_2(\mathcal{F}).$$

The notion of disappointment we have been focusing on so far was based only on observed output. It will be interesting, for robustness, to investigate whether Theorem 2 will still go through when considering alternative definitions of disappointment based on workers' effort levels. For this purpose, we consider three concepts of disappointment: **ex-post disappointment**, **interim disappointment**, and **ex-ante disappointment**. Ex-post disappointment results after an effort profile has been supplied. It is defined, for each player i and each effort profile x, as the difference in the worker's fair pay when the employer perfectly observes the effort profile $x \in X$ versus when he has no such information but observes the corresponding output level $f(x)$ precisely. More formally, it is defined by:

$$D_i^{exp}(\mathcal{F},x) = Sh_i(f,x) - IBV_i(\mathcal{F}|I(f(x)) \text{ for all } i \in N \text{ and } x \in X$$

Interim disappointment results after an effort level has been supplied. It is defined, for each player i and each effort level $x_i \in X_i$ as the expectation of the worker's ex-post disappointment, where the expectation is based on the effort levels supplied by all workers but worker i. More formally, it is defined by:

$$D_i^{int}(\mathcal{F},x_i) = \int_{X_{-i}} \rho(x_{-i} \mid x_i)\left(Sh_i(f,(x_i,x_{-i})) - IBV_i(\mathcal{F}|I(f(x_i,x_{-i}))\right) dx_{-i} \text{ for all } i \in N \text{ and } x_i \in X_i$$

Ex-ante disappointment refers to a disappointment a priori, that is, before any supply of effort. It is defined for each player i as the expectation of the worker's ex-post disappointment, where the expectation is based on the effort levels supplied by all workers. More formally, it is defined by

$$D_i^{exan}(\mathcal{F}) = \int_X \rho(x)\left(Sh_i(f,x) - IBV_i(\mathcal{F}|I(f(x))\right) dx \text{ for all } i \in N$$

Our next result states that ex-ante disappointment equals zero for every worker. Alternatively, a potential worker for an organization is never disappointed a priori. This intuitively makes sense since a disappointed worker a priori will never choose to join the organization he intended to work for. Implications of this result are that the average ex-post disappointment is equal to zero, and the average interim disappointment is equal to zero for every worker. These results can be summarized as follows.

Theorem 3 : *For any organization with incomplete information, $\mathcal{F} = (N,X,f,\rho)$ and any worker $i \in N$, we have:*

$$D_i^{exan}(\mathcal{F}) = \mathbb{E}_X\left(D_i^{exp}(\mathcal{F},x)\right) = \mathbb{E}_{X_i}\left(D_i^{int}(\mathcal{F},x_i)\right) = 0$$

This finding implies that if production is repeated over time, no worker will be disappointed in the long run under IBV despite the presence of uncertainty and information asymmetry. We illustrate it in the example below.

Example 5 : *Consider the organization with incomplete information $\mathcal{F} = (N,X,f,\rho)$ from the previous example, where $N = \{1,2\}$, $X = [0,1]\times[0,2]$, $f(x) = 2x_1 + 3x_2$ and $\rho(x) = \frac{1}{2}$. Let* *us focus on worker 1. We found that for all* $x \in X$*:* $Sh_1(f,x) = 2x_1$ *and*

$$IBV_1(\mathcal{F}|I(f(x))) = \begin{cases} \frac{2x_1+3x_2}{2} & if \quad f(x) \in [0,2] \\ 1 & if \quad f(x) \in [2,6] \\ \frac{2x_1+3x_2}{2} - 2 & if \quad f(x) \in [6,8] \end{cases} = \begin{cases} \frac{2x_1+3x_2}{2} & if \quad x \in \mathcal{C}_1 \\ 1 & if \quad x \in \mathcal{C}_2 \\ \frac{2x_1+3x_2}{2} - 2 & if \quad x \in \mathcal{C}_3 \end{cases},$$

where $\mathcal{C}_1 = \{x \in X : f(x) \in [0,2]\} = [0,1] \times [0, \frac{2-2x_1}{3}]$, $\mathcal{C}_2 = \{x \in X : f(x) \in [2,6]\} = [0,1] \times [\frac{2-2x_1}{3}, \frac{6-2x_1}{3}]$, *and* $\mathcal{C}_3 = \{x \in X : f(x) \in [6,,8]\} = [0,1] \times [\frac{6-2x_1}{3}, 2]$ *are depicted in Figure 1.4.*

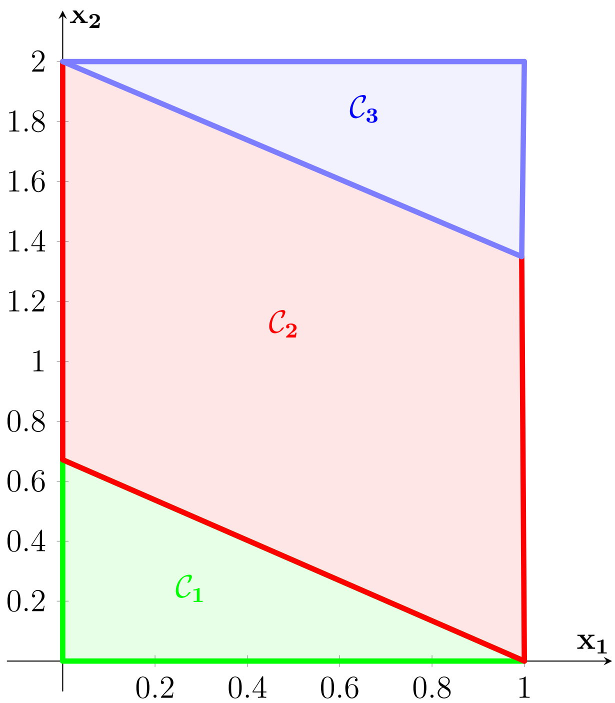

Figure 1.4: The set $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ of effort profiles.

The worker's ex-post disappointment is given for all $x \in X$ *by:*

$$D_1^{exp}(\mathcal{F},x) = Sh_1(f,x) - IBV_1(\mathcal{F}|I(f(x))) = \begin{cases} \frac{2x_1-3x_2}{2} & if \quad x \in \mathcal{C}_1 \\ 2x_1 - 1 & if \quad x \in \mathcal{C}_2 \\ \frac{2x_1-3x_2}{2} + 2 & if \quad x \in \mathcal{C}_3 \end{cases}$$

It follows that the worker's ex-ante disappointment is equal to:

$$D_1^{exan}(\mathcal{F}) = \int_X \rho(x)\, D_1^{exp}(\mathcal{F},x)\, dx = \frac{1}{2}\left(\int_{\mathcal{C}_1} \frac{2x_1-3x_2}{2}\, dx + \int_{\mathcal{C}_2} (2x_1-1)\, dx + \int_{\mathcal{C}_3} \left(\frac{2x_1-3x_2}{2} + 2\right) dx\right)$$

Simple calculations show that:

$$\int_{\mathcal{C}_1} \frac{2x_1-3x_2}{2}\, dx = \int_0^1 \int_0^{\frac{2-2x_1}{3}} \frac{2x_1-3x_2}{2}\, dx_2 dx_1 = \int_0^1 \left(-x_1^2 + \frac{4}{3}x_1 - \frac{1}{3}\right) dx_1 = 0,$$

$$\int_{\mathcal{C}_2}(2x_1-1)\,dx=\int_0^1\int_{\frac{2-2x_1}{3}}^{\frac{6-2x_1}{3}}(2x_1-1)\,dx_2dx_1=\int_0^1\frac{4}{3}(2x_1-1)\,dx_1=0,\ \ and$$

$$\int_{\mathcal{C}_3}\left(\frac{2x_1-3x_2}{2}+2\right)dx=\int_0^1\int_{\frac{6-2x_1}{3}}^{2}\left(\frac{2x_1-3x_2}{2}+2\right)dx_2dx_1=\int_0^1\left(x_1^2-\frac{2}{3}x_1\right)dx_1=0.$$

It follows that $D_1^{exan}(\mathcal{F})=0$ *as expected.*

We end this study with some applications of the models developed in the previous sections.

1.5 Some applications

There is a wide variety of applications of our theory. In this section, we provide only two applications: an application to a profit-maximizing firm facing uncertainty in input supply, and another application to the Cournot model.

1.5.1 A profit-maximizing firm facing uncertainty in input supply

Consider a firm consisting of n agents who supply their effort levels under uncertainty to produce a single output. Assume the firm exhibits price-taking behavior and that agents have different abilities, which implies that the unit price of agents' effort levels is valued differently. Let p be the price of one unit of output produced and w_i be the unit price of the effort level of agent $i\in N=\{1,2,...,n\}$. Each agent i can supply an effort level from 0 up to a maximum value $\overline{x_i}$, so that $X_i=[0,\overline{x_i}]$ is agent i's effort level set and $X=\prod_{i\in N}X_i$ is the set of effort level profiles. let ρ be a probability density function over X characterizing the uncertainty affecting the supply of agents' effort levels. Output is produced using a concave technology $f:X\longrightarrow\mathbb{R}$. We would like to find the fair share of the firm's profit each agent can expect to receive.

In the absence of uncertainty, the profit maximization problem of the firm can be stated as $\pi(p,w)=\max_{x\in X}\pi(x)=pf(x)-\sum_{i\in N}w_ix_i$. Let $x^*=(x_1^*,...,x_n^*)$ be the solution of this problem. Under uncertainty, agent $i\in N$ will never supply an effort level beyond the optimal level x_i^* because any additional effort level supplied will decrease the firm's marginal revenue. Whereas any effort level below the optimal level is likely to be supplied by the agent due to the uncertainty affecting the supply of effort. The set of effort profiles that is consistent with this optimal behavior is given by $I_f=\{x\in X:x_i\leq x_i^*\text{ for all }i\in N\}$. The fair share of each agent $i\in N$ is then defined as the informational Bayesian value of the firm associated to the informational constraint I induced by I_f. That is,

$$IBV_i(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) Sh_i(f,x)\, dx}{\int_{I_f} \rho(x)\, dx}, \quad i \in N.$$

As an illustration, consider a firm consisting of two agents 1 and 2, with effort levels $X_1 = [0,2]$ and $X_2 = [0,3]$, respectively. Assume their effort levels are independently and uniformly distributed over the set $X = [0,2] \times [0,3]$. Let the technology of the firm be defined for any $x = (x_1, x_2) \in X$ by $f(x) = x_1^{\frac{1}{3}} x_2^{\frac{1}{2}}$. Assume the factors' prices are given by $p = 6$, $w_1 = 2$ and $w_2 = 3$. The maximization problem under complete information is: $\max_{x \in X} 6x_1^{\frac{1}{3}} x_2^{\frac{1}{2}} - 2x_1 - 3x_2$. It follows that the optimal levels of effort supplied by agents are: $x_1^* = x_2^* = 1$. The set of effort profiles that is consistent with this optimal behavior is given by $I_f = [0,1] \times [0,1]$. Therefore, the fair share of agent 1 and agent 2 are respectively $IBV_1(\mathcal{F}|I) = \int_0^1 \int_0^1 3x_1^{\frac{1}{3}} x_2^{\frac{1}{2}} - 2x_1 \, dx_1 dx_2 = \frac{1}{2}$ and $IBV_2(\mathcal{F}|I) = \int_0^1 \int_0^1 3x_1^{\frac{1}{3}} x_2^{\frac{1}{2}} - 3x_2 \, dx_1 dx_2 = 0$.

We found that the fair share of agent 1 is higher than that of agent 2. A rationale for this result is that agent 2's level of effort costs the firm more than the effort level of agent 1. Whereas the marginal revenue of the firm benefits both agents equally.

1.5.2 Cournot oligopoly in an uncertain world

We now turn our attention to an application to the classical oligopoly model. To this end, consider an industry for a homogeneous product with n firms. Each firm $i = 1, \ldots, n$ has a capacity K_i and produces an output level q_i at constant marginal cost c_i. Suppose that the industry has linear inverse demand given by $P(Q) = a - bQ$, where $a, b > 0$ and $Q = \sum_{i=1}^{n} q_i$ is the industry output.[7] The production of outputs is subject to uncertainty characterized by a probability density function ρ over the set $X = \prod_{i=1}^{n} [0, K_i]$. In the event that the industry firms come together to form a single cartel, what would be a fair allocation of the profits earned through collusion among the cartel members? Consistent with our theory, we assume that the collusive profit will be shared according to the informational Bayesian value.

This industry can be modeled as a firm with incomplete information $\mathcal{F} = (N, X, \pi, \rho, I)$, where N is a set of firms, $X = \prod_{i=1}^{n} [0, K_i]$ is the set of output profiles, and π is the cartel's profit function defined as $\pi(q_1, \ldots, q_n) = \sum_{i=1}^{n} (a - c_i - bQ) q_i$. $I \subseteq \mathbb{R}$ is the informational constraint that sets the profit realized by the cartel or the profit level the cartel aims to achieve. It can be a range of values ($I = [\underline{\pi}, \overline{\pi}]$) if the cartel profit is imprecisely known or a fixed level of profit

[7] Linear inverse demand is the standard assumption in most analyses of the Cournot model (see, e.g., Singh and Vives (1984), Bimpikis, Ehsani, and Ilkılıç (2019), Jing-Yuan and Smeers (1999)).

($I = \{\pi^*\}$), if the realized profit or the targeted profit is known precisely by the cartel. The latter case encompasses situations where the cartel decides to produce the optimal level of output and earns the highest profit $\pi^* = \max_{q_1,\dots,q_n} \pi(q_1, q_2, \dots, q_n)$). However, the optimal level of output can be achieved as an interior solution only if marginal costs across all firms are equalized. Our approach provides a broader scope for distributing any level of profit among cartel members, regardless of whether their marginal costs are equal or not. Under a general informational constraint I on the cartel's profit, the fair share of firm i that complies with the constraint I is given by:

$$IBV_i(\mathcal{F}|I) = \frac{\int_{I_\pi} \rho(q) Sh_i(\pi, q)\, dq}{\int_{I_\pi} \rho(q)\, dq},$$

where $I_\pi = \{(q_1, \dots, q_n) \in X : \pi(q_1, \dots, q_n) \in I\}$.

Illustration. Consider a cartel consisting of two firms. The firms' capacities and marginal costs are given by $K_1 = 60$, $K_2 = 100$, $c_1 = 20$, and $c_2 = 50$. Let $a = 1000$ and $b = 5$. The output productions of firms are independently and uniformly distributed. Assume the cartel has made a profit of 25000, and let us determine the fair contribution of each firm to this profit. The informational constraint is $I = \{25000\}$. The set of output profiles that satisfy the constraint is given by $I_\pi = \{(q_1, q_2) \in [0, 60] \times [0, 100] : \pi(q_1, q_2) = 25000\}$, where the cartel profit $\pi(q_1, q_2)$ is defined by

$$\pi(q_1, q_2) = (980 - 5(q_1 + q_2))\, q_1 + (950 - 5(q_1 + q_2))\, q_2 = 980q_1 + 950q_2 - 5(q_1 + q_2)^2.$$

A parametrization of the set I_π of production profiles $(q_1, q_2) \in X$ that satisfy the constraint I leads to:

$$I_\pi = \left\{ \left(t, 95 - t - \sqrt{6t + 4025} \right) : t \in [0, 98 - \sqrt{4604}] \right\}.$$

A simple computation shows that $Sh_1(\pi, q) = 980q_1 - 5q_1^2 - 5q_1q_2$ and $Sh_2(\pi, q) = 950q_1 - 5q_2^2 - 5q_1q_2$ for all $q = (q_1, q_2) \in [0, 60] \times [0, 100]$. It follows that the fair share of each firm out of the cartel profit is:

$$IBV_1(\mathcal{F}|I) = \frac{\int_{I_\pi} \rho(q) Sh_1(\pi, q)\, dq}{\int_{I_\pi} \rho(q)\, dq} = 12463.7$$

$$IBV_2(\mathcal{F}|I) = \frac{\int_{I_\pi} \rho(q) Sh_2(\pi, q)\, dq}{\int_{I_\pi} \rho(q)\, dq} = 12536.3$$

It appears that firm 2 enjoys a large share of the cartel's profit due to its higher production capacity relative to firm 1.

1.6 Conclusion

In this paper, we develop an axiomatic foundation for the problem of paying workers in a production environment where their actions are not observed by the employer. The latter has beliefs over workers' abilities (or observes the distributions from which abilities are drawn), has precise or imprecise information on output, and demands fairness. We extend classical axioms of fairness under conditions of uncertainty and asymmetric information and use them to characterize a unique pay scheme called the informational Bayesian value (IBV). The IBV defines an *output-based contract* between an uninformed employer and informed employees. Then, we analyze the implications of IBV for worker disappointment. We show that worker disappointment depends on the level of realized output, and varies across workers. However, despite this heterogeneity, the average disappointment is zero for each worker. This means that if production is repeated over time, no worker will be disappointed in the long run under IBV despite the fact that information is incomplete and asymmetric. Finally, we develop applications to some classical economic problems, including profit-maximizing firms that face uncertainty in input supply, and the Cournot oligopoly model.

1.7 Appendix: Proofs of Results

Proof (Theorem 1) **:** ***Sufficiency:*** *In this part of the proof, we show that the informational Bayesian value, IBV, satisfies the principles of efficiency a posteriori and bilateral fairness a posteriori. Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information and I the employer's information set. Before proceeding with the proof, let us present an alternative formulation of the informational Bayesian value. Consider the conditional probability density function $\rho(\cdot \mid I)$ obtained from the employer's beliefs ρ using Bayes' rule and defined over X by:*

$$\rho(x \mid I) = \begin{cases} \frac{\rho(x)}{\int_{I_f} \rho(x)\, dx} & if \quad x \in I_f \\ 0 & if \quad x \notin I_f \end{cases}$$

It follows that for all $k \in N$,

$$IBV_k(\mathcal{F}|I) = \frac{\int_{I_f} \rho(x) Sh_k(f,x)\, dx}{\int_{I_f} \rho(x)\, dx} = \int_X \rho(x \mid I) Sh_k(f,x)\, dx.$$

Efficiency a posteriori: *The following equalities hold.*

$$\begin{aligned} \sum_{i \in N} IBV_i(\mathcal{F}|I) &= \sum_{i \in N} \int_X \rho(x \mid I) Sh_i(f,x)\, dx \\ &= \int_X \rho(x \mid I) \sum_{i \in N} Sh_i(f,x)\, dx \\ &= \int_X \rho(x \mid I) f(x)\, dx \\ &= \frac{\int_{I_f} \rho(x) f(x)\, dx}{\int_{I_f} \rho(x)\, dx} \end{aligned}$$

The penultimate equality holds because $\sum_{i \in N} Sh_i(f,x) = f(x)$. Therefore, the informational Bayesian value satisfies the efficiency a posteriori axiom.

Bilateral fairness a posteriori: *Let i and j be two workers and I an information set. It holds that:*

$$\begin{aligned}
IBV_i(\mathcal{F}^{-j}|I^{-j}) &= \int_{X_{-j}} \rho(x_{-j}|I^{-j})\, Sh_i(f^{-j}, x_{-j})\, dx_{-j} \\
&= \frac{\int_{I^{-j}_{f^{-j}}} \rho_{-j}(x_{-j})\, Sh_i(f^{-j}, x_{-j})\, dx_{-j}}{\int_{I^{-j}_{f^{-j}}} \rho_{-j}(x_{-j}) dx_{-j}} \\
&= \frac{\int_{I^{-j}_{f^{-j}}} \left[\int_{X_j} \rho(x_j, x_{-j}) dx_j\right] Sh_i\left(f, (0, x_{-j})\right) dx_{-j}}{\int_{I^{-j}_{f^{-j}}} \left[\int_{X_j} \rho(x_j, x_{-j}) dx_j\right] dx_{-j}} \\
&= \frac{\int_{I^{-j}_{f^{-j}}} \int_{X_j} \rho(x)\, Sh_i\left(f, (0, x_{-j})\right) dx}{\int_{I^{-j}_{f^{-j}}} \int_{X_j} \rho(x) dx} \\
&= \int_X \rho(x \mid I)\, Sh_i\left(f, (0, x_{-j})\right) dx
\end{aligned}$$

The last equality holds because for all $x_{-j} \in I^{-j}_{f^{-j}}$, there exists $x_j \in X_j$ such that $(x_j, x_{-j}) \in I_f$ and for $x_{-j} \in I^{-j}_{f^{-j}}$ such that $x = (x_j, x_{-j}) \notin I_f$ for all $x_j \in X_j$, we have that $\rho(x|I) = 0$. It follows that:

$$IBV_i(\mathcal{F}|I) - IBV_i(\mathcal{F}^{-j}|I^{-j}) = \int_X \rho(x \mid I)\, [Sh_i(f, x) - Sh_i(f, (0, x_{-j}))dx] \tag{1.4}$$

Likewise, we have that:

$$IBV_j(\mathcal{F}|I) - IBV_j(\mathcal{F}^{-i}|I^{-i}) = \int_X \rho(x \mid I)\, [Sh_j(f, x) - Sh_j(f, (0, x_{-i}))dx] \tag{1.5}$$

From equations (1.4) and (1.5), to show that $IBV_i(\mathcal{F}|I) - IBV_i(\mathcal{F}^{-j}|I^{-j}) = IBV_j(\mathcal{F}|I) - IBV_j(\mathcal{F}^{-i}|I^{-i})$, it suffices to show that $Sh_i(f, x) - Sh_i(f, (0, x_{-j})) = Sh_j(f, x) - Sh_j(f, (0, x_{-i}))$ for all $x \in X$. That is:

$$Sh_i(f, x) - Sh_j(f, x) = Sh_i(f, (0, x_{-j})) - Sh_j(f, (0, x_{-i})) \text{ for all } x \in X. \tag{1.6}$$

Let $x \in X$. If $x_i = 0$ or $x_j = 0$, then it follows that: $Sh_i(f, x) - Sh_j(f, x) = 0 = Sh_i(f, (0, x_{-j})) - Sh_j(f, (0, x_{-i}))$. Now assume workers i and j are active at x, that is $x_i \neq 0$ and $x_j \neq 0$. It *follows that:*

$$
\begin{aligned}
Sh_i(f,x) &= \sum_{a \lhd_0^i x} \varphi(a,x)\,[f(a+x_ie_i) - f(a)] \\
&= \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,x)\,[f(a+x_ie_i) - f(a)] + \sum_{\substack{a \lhd_0^i x \\ a_j\neq 0}} \varphi(a,x)\,[f(a+x_ie_i) - f(a)] \\
&= \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,x)\,[f(a+x_ie_i) - f(a)] + \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a+x_je_j,x)\,[f(a+x_je_j+x_ie_i) - f(a+x_je_j)]
\end{aligned}
$$

The last equality follows from the fact that for all $a, x \in X$ *such that* $a \lhd_0^i x$, $a_j \neq 0$ *implies that* $a_j = x_j$.

Similarly, we have that:

$$
Sh_j(f,x) = \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,x)\,[f(a+x_je_j) - f(a)] + \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a+x_ie_i,x)\,[f(a+x_ie_i+x_je_j) - f(a+x_ie_i)]
$$

But for all $a \in X$ *such that* $a \lhd_0^i x$ *and* $a_j = 0$, *we have that*

$$
\varphi(a+x_ie_i,x) = \frac{(|a|+1)!(|x|-|a|-2)!}{(|x|)!} = \varphi(a+x_je_j,x).
$$

This implies that:

$$
\begin{aligned}
Sh_i(f,x) - Sh_j(f,x) &= \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,x)\,[f(a+x_ie_i) - f(a+x_je_j)] \\
&+ \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a+x_je_j,x)\,[f(a+x_ie_i) - f(a+x_je_j)] \\
&= \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \left(\varphi(a,x) + \varphi(a+x_je_j,x)\right)\,[f(a+x_ie_i) - f(a+x_je_j)]
\end{aligned}
$$

But for all $a \in X$ *such that* $a \lhd_0^i x$ *and* $a_j = 0$,

$$
\varphi(a,x) + \varphi(a+x_je_j,x) = \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} + \frac{(|a|+1)!(|x|-|a|-2)!}{(|x|)!} = \frac{(|a|)!(|x|-|a|-2)!}{(|x|-1)!}
$$

It follows that:

$$
Sh_i(f,x) - Sh_j(f,x) = \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \frac{(|a|)!(|x|-|a|-2)!}{(|x|-1)!}\,[f(a+x_ie_i) - f(a+x_je_j)] \tag{1.7}
$$

On the other hand, we have that:

$$\begin{aligned} Sh_i(f,(0,x_{-j})) &= \sum_{a \lhd_0^i (0,x_{-j})} \varphi(a,(0,x_{-j}))\,[f(a+x_ie_i)-f(a)] \\ &= \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,(0,x_{-j}))\,[f(a+x_ie_i)-f(a)] \end{aligned}$$

Similarly, we have that:

$$Sh_j(f,(0,x_{-i})) = \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \varphi(a,(0,x_{-i}))\,[f(a+x_je_j)-f(a)]$$

But for all $a \in X$ such that $a \lhd_0^i x$ and $a_j = 0$, we have that:

$$\varphi(a,(0,x_{-j})) = \frac{(|a|)!(|x|-|a|-2)!}{(|x|-1)!} = \varphi(a,(0,x_{-i})).$$

It follows that:

$$Sh_i(f,(0,x_{-j})) - Sh_j(f,(0,x_{-i})) = \sum_{\substack{a \lhd_0^i x \\ a_j=0}} \frac{(|a|)!(|x|-|a|-2)!}{(|x|-1)!}\,[f(a+x_ie_i)-f(a+x_je_j)] \quad (1.8)$$

From equations (2.10) and (2.11), it follows that Equation (2.9) holds. Hence, the informational Bayesian value satisfies bilateral fairness a posteriori.

Necessity: *In this part of the proof, we prove the uniqueness of the IBV. Consider another procedure ϕ which satisfies the axioms of efficiency a posteriori and bilateral fairness a posteriori. We would like to show that, for any organization with incomplete information $\mathcal{F} = (N, X, f, \rho)$ and any information set I, we have $\phi_i(\mathcal{F}|I) = IBV_i(\mathcal{F}|I)$, for each worker $i \in N$.*

We proceed by induction on the number n of workers in the organization. If we have only one worker in an organization with incomplete information, that is if $n = 1$, then the efficiency a posteriori axiom implies that this worker would get all of the expected output produced under the pay schemes IBV and ϕ. So, the proof is verified for $n = 1$. Now assume that the two pay schemes are identical for all organizations with incomplete information with less than n workers, and all information sets. Let $\mathcal{F} = (N, X, f, \sigma)$ be an organization with n workers and I be an information set. By induction, we have that $IBV_i(\mathcal{F}^{-j}|I^{-j}) = \phi_i(\mathcal{F}^{-j}|I^{-j})$ for any

$i, j \in N$. As a result, for all $i, j \in N$, the axiom of stochastic bilateral fairness implies that:

$$IBV_i(\mathcal{F}|I) - IBV_j(\mathcal{F}|I) = IBV_i(\mathcal{F}^{-j}|I^{-j}) - IBV_j(\mathcal{F}^{-i}|I^{-i}) = \phi_i(\mathcal{F}^{-j}|I^{-j}) - \phi_j(\mathcal{F}^{-i}|I^{-i}) = \phi_i(\mathcal{F}|I) - \phi_j(\mathcal{F}|I)$$

That is:

$$IBV_i(\mathcal{F}|I) - IBV_j(\mathcal{F}|I) = \phi_i(\mathcal{F}|I) - \phi_j(\mathcal{F}|I), \text{ for all } i, j \in N.$$

Fixing i and summing over $j \in N$ yields:

$$IBV_i(\mathcal{F}|I) - \sum_{j \in N} IBV_j(\mathcal{F}|I) = \phi_i(\mathcal{F}|I) - \sum_{j \in N} \phi_j(\mathcal{F}|I), \text{ for all } i \in N.$$

The axiom of efficiency a posteriori applies and we get:

$$IBV_i(\mathcal{F}|I) - \int_X \sigma(x|I) f(x)\, dx = \phi_i(\mathcal{F}|I) - \int_X \sigma(x|I) f(x)\, dx, \text{ for all } i \in N.$$

It follows that:

$$IBV_i(\mathcal{F}|I) = \phi_i(\mathcal{F}|I), \text{ for all } i \in N.$$

Hence, the informational Bayesian value, IBV, is the unique pay scheme that satisfies simultaneously the axioms of efficiency a posteriori and bilateral fairness a posteriori.

Proof (Theorem 2) : *Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information. We prove the result for finite partitions of informational constraints, but this can be easily extended to the infinite case. Let $\mathcal{I}_f$ be a finite partition of informational constraints, and $i \in N$ a worker. We have that:*

$$\begin{aligned}
\mathbb{E}_{\mathcal{I}_f}(IBV_i(\mathcal{F}|I)) &= \sum_{I \in \mathcal{I}_f} P(I) IBV_i(\mathcal{F}|I) \\
&= \sum_{I \in \mathcal{I}_f} \left(\int_{I_f} \rho(x)\, dx \right) \left\{ \frac{\int_{I_f} \rho(x) Sh_i(f, x)\, dx}{\int_{I_f} \rho(x)\, dx} \right\} \\
&= \sum_{I \in \mathcal{I}_f} \int_{I_f} \rho(x) Sh_i(f, x)\, dx \\
&= \int_X \rho(x) Sh_i(f, x)\, dx \\
&= IEV_i(\mathcal{F}).
\end{aligned}$$

■

Proof (Theorem 3) : *Let $\mathcal{F} = (N, X, f, \rho)$ be an organization with incomplete information and $i \in N$ a worker. It follows from the definition that:*

$$D_i^{exan}(\mathcal{F}) = \mathbb{E}_X\left(D_i^{exp}(\mathcal{F}, x)\right) = \mathbb{E}_{X_i}\left(D_i^{int}(\mathcal{F}, x)\right).$$

Let us show that $D_i^{exan}(\mathcal{F}) = 0$. The following equalities hold:

$$\begin{aligned}
D_i^{exan}(\mathcal{F}) &= \int_X \rho(x)\left(Sh_i(f,x) - IBV_i(\mathcal{F}|I(f(x)))\right)dx \\
&= \int_X \rho(x)\, Sh_i(f,x)\, dx - \int_X \rho(x)\, IBV_i(\mathcal{F}|I(f(x))\, dx \\
&= IEV_i(\mathcal{F}) - \int_{\mathcal{R}_f}\left(\int_{f^{-1}(y)} \rho(x) dx\right) IBV_i(\mathcal{F}|I(y))\, dy \\
&= IEV_i(\mathcal{F}) - \mathbb{E}_{\mathcal{K}_f}\left(IBV_i(\mathcal{F}|I(y))\right) \\
&= IEV_i(\mathcal{F}) - IEV_i(\mathcal{F}) \\
&= 0
\end{aligned}$$

The penultimate equality holds because of Theorem 2. ∎

Chapter 2

Acting under Incomplete Information: Mechanism Design, Justice, and Efficiency

2.1 Introduction

Adverse selection and moral hazard problems arise in economic environments with incomplete and asymmetric information. Theoretical investigations of these issues have traditionally followed a positive approach (Rothschild and Stiglitz, 1978; Akerlof, 1970; Spence, 1974; Hart and Holmström, 1987; Varian, 1992). This approach typically involves solving optimization problems, such as maximizing the principal's payoff while ensuring that the agents' individual rationality and incentive compatibility constraints are satisfied. In this chapter, I incorporate fairness considerations into the classical framework of a production economy with incomplete information. In this new environment, I study individual incentives and behaviors, showing that *fairness* defines a type of contract that always leads to stable or self-enforcing actions on the part of economic agents. I also identify structural conditions under which such actions are optimal in the sense of enhancing efficiency. Finally, I develop an application for mechanism design, showing that in some economies, fairness induces truthful revelation of private information (such as productivity or ability) and incentivizes agents to contribute their best efforts, leading to optimal production.

It is generally argued that fairness is crucial not only because of ethical considerations but also for ensuring efficiency and long-term stability. When resources are allocated unfairly, inefficiencies can arise as a result of demotivation, free-riding, and systemic risks (Holmstrom, 1982; Gilliland, 1994; Folger and Cropanzano, 2001). In environments where asymmetric information is present, the knowledge of how fairness affects economic behavior and efficiency is very limited. The classic problem of adverse selection, where individuals' private information distorts incentives and resource allocation illustrates well some of the challenges that incomplete information poses. However, so far, fairness considerations have not been incorporated in this framework.

To illustrate, consider a production environment where a company relies on a team of workers with different skills to complete a project. Each worker has private information about their ability or productivity. Some workers are highly skilled, while others have lower productivity, but only each worker knows how productive they truly are. The manager must decide how to distribute the project's output (such as bonuses or wages) based on contributions that are not fully observable to the others. This scenario presents a classic problem of adverse selection: less skilled workers might contribute less but still hope to receive a share of the output, while highly skilled workers may fear they won't be fairly compensated for their higher productivity.

The challenge is to design a pay scheme that *fairly* distributes the output among workers, ensuring that each worker's contribution is reflected while addressing the information asymmetry that could otherwise lead to adverse selection and/or moral hazard. I address this problem by allowing workers to choose their level of effort based on private information about their productivity, along with beliefs about the productivity of other workers.

Unlike traditional models that assume all participants have full knowledge of each other's productivity, I consider an economy of incomplete information (Bayesian economy) where each agent holds a finite set of private information (types) that encloses information on their productivity and has the option to remain inactive or participating in the production by choosing an action from a finite set.[1] The model allows for heterogeneity among agents, in the sense that different agents can have different types and different action sets. Each type profile defines a state of the economy in which agents' actions are converted into measurable output by the state's technology. I assume that each agent has beliefs about the types of other agents drawn from a common prior distribution and derives utility from their pay, which reflects their contribution to the output produced. The model unfolds in three stages, each defined by the extent of information available to the agents

1. **Ex-ante stage:** Nature draws a type profile $\theta = (\theta_1, \ldots, \theta_n)$ that specifies the productivity of each agent in the economy, according to a prior probability distribution $\mathbb{P}$. At this stage, agents are unaware of both their own types and those of others.

[1] I have chosen to use singular "they/them" as a gender-neutral pronoun.

2. **Interim stage:** Each agent learns their own type $\theta_i \in \Theta_i$, chooses an action x_i from a finite set X_i, and forms beliefs about the types of other agents using $\mathbb{P}$.

3. **Ex-post stage:** Agents know the type profile θ as well as the action profile $x = (x_1, ..., x_n) \in X$. The output $f_\theta(x)$ is produced, and each agent receives their pay.

I assume that each type profile θ defines a specific state of the economy, characterized by a technology f_θ that maps each action profile x to a measurable output $f_\theta(x)$. An agent's pay in each state should reflect their contribution to the resulting output.

In this model, I address several questions outlined as follows:

1. How should agents be *fairly* compensated at the ex-post stage?

2. Do fair pay schemes defined at the ex-post stage induce *self-enforcing* and *efficient* behaviors among agents at the interim/ex-ante stage?

3. Do agents act optimally and truthfully under fair pay schemes?

The first question has garnered significant attention, especially in the literature on the theories of value (Moulin, 1995; Algaba, Fragnelli, and Sánchez-Soriano, 2019; Winter, 2002; Roth, 1988). While much of the existing research has been conducted within the framework of complete information, this study approaches the question through the lens of a model with incomplete information, inspired by the Bayesian formulation of Harsanyi (1967) and Harsanyi (1968). Building upon these studies, I follow a normative approach by setting a number of desirable properties–referred to as the principles of fairness–that a reasonable pay scheme should satisfy. However, *fairness* itself is a nuanced concept, and I aim to disentangle two distinct notions of fairness—*fairness-as-reciprocity* and *fairness-as-fair-equality-of-opportunity*—that are under consideration. Fairness-as-reciprocity focuses on the idea that social institutions are fair when those who contribute more to the common good receive more in return. Workers who exert higher effort and contribute to greater output should be rewarded proportionately, while free-riders—those who benefit from the efforts of others without contributing themselves—should not receive undue rewards. Reciprocity demands that the distribution of rewards be aligned with individual contributions, preventing exploitation or unfair advantages (Intropi, 2024; Lister, 2020; Lister, 2013; Binmore, 2004; Widerquist, 1999). Whereas fairness-as-fair-equality-of-opportunity, drawing from John Rawls' theory (Rawls, 1971), emphasizes the importance of ensuring a level playing field. It is unfair, on this view, when agents' outcomes are determined by factors beyond their control, such as genetic endowment, family background, or disabilities. A fair society should ensure that all agents have equal opportunities to develop their abilities and compete on fair terms. Any resulting inequalities in pay or output should reflect differences in effort or choices, not uncontrollable circumstances.

The basic model mainly focuses on economies whose conception of fairness aligns with fairness-as-reciprocity. However, I later extend this model to encompass economies with a more inclusive understanding of fairness.

In the basic model, agents choose their actions based on private information about their types. These actions are then converted by the technology of the realized state into an output, which is distributed among the agents as pay or rewards. In this setting, I introduce fairness principles that govern the ex-post distribution of output. These principles are outlined as follows:

1. *Ex-post inactivity*: each agent's pay should be zero if all agents remain inactive.

2. *Ex-post efficiency*: the output produced should be entirely distributed among agents of the economy.

3. *Ex-post bilateral fairness*: any pair of agents should contribute equally to each other's pay.

Importantly, I show that these principles of fairness characterize a unique pay scheme called the *ex-post fair pay* (Theorem 4). The ex-post fair pay generalizes the classical Shapley value (Shapley, 1953), known as the most conceptually elegant and appealing solution concept in cooperative game theory. However, our model extends that of cooperative games in many respects. First, it is defined in an incomplete information setting. Second, in a cooperative game, all agents have the same set of actions that contains only two options–*participate* and *not participate* in the game. In addition, it is assumed that the grand coalition forms, meaning that all agents decide to participate in the game. Our model goes beyond these assumptions by allowing different agents to have different action sets that may contain more than two actions. Third, in our model, each action profile defines a cooperative game among the agents, so our setting can be viewed as a multi-game setting, where different games are played at different action profiles. Consequently, our formalization of the principles of fairness characterizing the Shapley value differs from those considered in this literature (Shapley, 1953; Young, 1985; Myerson, 1977; Maniquet, 2003; Aguiar et al., 2019).

To address the question concerning the implications of fair pay schemes for agents' incentives and efficiency at both the interim and ex-ante stages, I associate a Bayesian game to an economy with incomplete information and define an equilibrium in such an economy as a pure-strategy Bayesian Nash equilibrium of its associated Bayesian game.[2] Interestingly, I find that an economy that fully adheres to fairness principles always admits an equilibrium (Theorem 5). However, an equilibrium may not exist if some fairness principles are violated. As a result, Theorem 5 identifies a class of finite Bayesian games that always have a pure-strategy Bayesian Nash equilibrium.[3]

Although an equilibrium always exists in any fair Bayesian economy, this equilibrium may not be Pareto-efficient. I uncover structural conditions that guarantee the existence of a Pareto-efficient equilibrium. Specifically, I find that a fair Bayesian economy admits a Pareto-efficient equilibrium if agents' beliefs are drawn from a uniform distribution and the technology in any state of the economy is weakly monotonic. In addition, the equilibrium is unique if the technologies are strictly monotonic (Theorem 6). These results suggest that fair pay schemes lead to both production stability and efficiency under relatively mild conditions.

In the first part of this study (sections 2.2, 2.3, and 2.4), the model of an economy with incomplete information is kept as simple and clean as possible. The second part (section 2.5) then shows how the logic of the result persists under various extensions that either remedy unrealistic features of the basic model or otherwise enrich it. First, in the basic model, I assume that the entire economy bears the action costs of its agents, effectively setting each agent's individual cost to zero. While this assumption can be theoretically defended, it does not reflect the reality of most organizations, where agents are typically responsible for all or part of the costs associated with their actions. To account for this, the first extension relaxes this assumption by assigning each agent the cost of their actions, contingent on the realized state of nature. Second, while the basic model perfectly satisfies fairness-as-reciprocity, it does not address fairness-as-fair-equality-of-opportunity. This omission could invite objections: if productivity differences arise due to circumstances beyond workers' control, such as disabilities or socioeconomic disadvantages, it is not enough to simply reward higher contributions. Workers who are less productive through no fault of their own might find the pay scheme unfair if they are disadvantaged from the outset. To address this, I extend the basic model to include progressive taxation and redistribution that reflects a commitment to fairness-as-fair-equality-of-opportunity. In this extended model, the economy not only rewards contributions based on effort and productivity but also insures individuals against life's contingencies, such as disabilities or unequal access to resources. This ensures that workers who are unable to be

[2]Since the analysis of equilibrium existence at the ex-ante stage is equivalent to that at the interim stage (Harsanyi, 1967), I will focus on the equilibrium analysis at the interim stage.

[3]It is well known that a finite Bayesian game may not admit a pure strategy Bayesian Nash equilibrium.

as productive as others still receive support, aligning the model with Rawlsian ideals of a just society (Rawls, 1971). Lastly, I consider a more comprehensive extension that combines key aspects of the two previous extensions. For each extension, I establish the existence of equilibrium and identify conditions that ensure the equilibrium is Pareto-efficient and unique, highlighting the robustness of the results obtained in the basic model.

Finally, I address the question of whether fairness principles incentivize agents to participate optimally in production and act in accordance with their true types by leveraging our findings in the context of mechanism design. I proceed by considering mechanisms that specify the actions agents should take for each possible realization of their types. I am particularly interested in mechanisms that satisfy the following desiderata:

1. *Individual rationality,* which says that every agent ends up at least as well-off as being inactive.
2. *Pareto-efficiency,* which says that no deviation from the actions prescribed by the mechanism can make any agent better off without harming another.
3. *Strategy-proofness,* which says that no agent can ever profit from lying about their type.

The chapter's contribution to this literature is to define a class of mechanisms $\mathcal{M}^*$ that satisfy these desiderata. More precisely, each mechanism $\mathcal{M}^*$ is defined within a Bayesian economy and assigns to each type profile the action profile that maximizes the technology of the realized state of the economy. I show that $\mathcal{M}^*$ is Pareto-efficient if the Bayesian economy is fair, and it is individually rational and strategy-proof if the Bayesian economy is fair and every state of the economy is endowed with a weakly monotonic technology. This finding has implications for the design of mechanisms in productive economies that are compatible with fairness and efficiency while allowing truthful reporting of private information.

2.1.1 Related literature

The main contribution of this study lies in its integration of fairness considerations into the resolution of classical problems involving adverse selection (Akerlof, 1970; Spence, 1974) and moral hazard (Holmström, 1979) within a production economy characterized by incomplete information. Unlike traditional approaches that focus primarily on optimizing principal-agent relationships (Hart and Holmström, 1987), this study extends the framework by demonstrating that fairness-based contracts can induce self-enforcing and efficient behaviors among rational agents operating under asymmetric information. Consequently, this work advances the literature on contract theory by proposing a normative approach to contracting, complementing recent studies on robustness-based contracts, where the principal knows only a subset of the agent's possible actions and evaluates contracts based on worst-case performance (Carroll, 2015; Dai and Toikka, 2022; Miao and Rivera, 2016; Marku, Ocampo, and Tondji, 2024). In our model, the principal is one of the agents responsible for distributing the output produced. Her role can be made more explicit by limiting her actions to two options: *contracting* and *not contracting*. This study departs from previous research by assuming that the principal values fairness and assesses contracts based on her beliefs about the productivity of other agents.

This study also contributes to several other strands of literature. First, it adds to the literature examining how to allocate output among individuals who generated it. A central challenge in this area is to design pay schemes that are both fair and efficient, reflecting the varying contributions of individuals. Among the solutions developed, the Shapley value has emerged as a prominent method for addressing this allocation problem. Traditionally, the Shapley value has been applied in settings characterized by complete information, where all individual contributions are fully known and measurable. Within this framework, it has proven to be a powerful tool across a wide array of applications. In operations management and supply chains, the Shapley value has been used to optimize centralized decision-making (Gopalakrishnan et al., 2021), Kemahlıoğlu-Ziya and Bartholdi III (2011). It has been extensively employed in cost allocation, ensuring equitable distribution of expenses among stakeholders (Littlechild and Owen, 1973; Dubey, 1982). Additionally, it has played a critical role in fair division, contract design, and political power measurement, offering a rigorous method for evaluating influence and decision-making authority (Moulin, 1992; Winter, 2002; Shapley and Shubik, 1954; Freixas, Marciniak, and Pons, 2012; Pongou and Tchantcho, 2021; Demeze-Jouatsa, Pongou, and Tondji, 2024). Beyond these areas, the Shapley value has been instrumental in addressing bankruptcy problems (Aumann and Maschler, 1985), assessing network centrality (Grofman and Owen, 1982), solving queueing problems (Maniquet, 2003), and measuring unfairness and income inequality (Aguiar, Pongou, and Tondji, 2018). These applications demonstrate the Shapley value's versatility and effectiveness across various fields, but they all share the assumption of complete information. Our study extends this literature by addressing the allocation problem within a model of incomplete information, a scenario more reflective of real-world conditions where contributions are often only partially observable. I adapt the Shapley value framework to operate under these constraints, offering a novel approach to resource allocation that maintains fairness and efficiency even when information is incomplete. This study, therefore, complements the scarce literature that has applied the

Shapley value in settings characterized by uncertainty and asymmetric information (Krasa and Yannelis, 1994; Pongou and Tondji, 2018; Takeng, 2022). However, it distinguishes itself through its expanded scope and more in-depth analysis.

This study also contributes to the literature exploring the conditions under which a pure strategy Bayesian Nash equilibrium exists in Bayesian games. Seminal contributions in this area include the works of Radner and Rosenthal (1982) and Milgrom and Weber (1985), who imposed conditions of independent atomless types and private values. Athey (2001) further advances this literature by establishing equilibrium existence under the single-crossing condition. More recently, He and Sun (2019) introduced a coarser inter-player information requirement to demonstrate the existence of a pure strategy equilibrium. In contrast to these studies, I establish equilibrium existence for a class of Bayesian games through a normative approach, requiring that agents' utilities reflect their fair contribution to the game.

Finally, this study contributes to the mechanism design literature, a field dedicated to crafting frameworks and rules that incentivize self-interested agents to achieve socially desirable outcomes. Central to this endeavor is the design of mechanisms that fulfill several critical objectives, including efficiency, strategy-proofness, individual rationality, and fairness. However, these desiderata are often proven to be logically incompatible in various settings, such as social choice theory (Gibbard, 1973; Satterthwaite, 1975; Green and Laffont, 1977), matching theory (Roth, 1982; Alcalde and Barberà, 1994; Nesterov, 2017; Alva and Manjunath, 2020), and pure exchange economies (Serizawa, 2002; Momi, 2017). This study advances the literature by identifying a class of mechanisms that satisfy all these desiderata, thereby complementing existing research that has uncovered possibility results under specific preference domain restrictions (Bogomolnaia and Moulin, 2004; Bogomolnaia, Moulin, and Stong, 2005; Chen et al., 2013; Pycia and Ünver, 2017; Manjunath and Westkamp, 2021).

This study complements the recent work of Demeze-Jouatsa, Pongou, and Tondji (2024). The latter paper analyzes individual incentives and efficiency in political economies governed by principles of distributive justice. This chapter differs in several important respects. Demeze-Jouatsa, Pongou, and Tondji (2024) assume a complete information setting and so do not address the incentives problem in economies where agents' productivity is private information. I differ in that I assume incomplete and asymmetric information, meaning that an agent's productivity is privately known and not observable by others, resulting in a fundamentally different incentive structure and equilibrium concept. Also, the scope of this study is different. In particular, I study conditions under which agents truthfully report their productivity and act optimally under a fair pay rule. I uncover a mechanism that is always Pareto-efficient, and that is incentive-compatible and individually rational under a mild monotonicity assumption. All these findings are new. I also develop new applications, showing, for instance, how fairness can address the moral hazard problem.

This paper contributes to and extends a growing literature at the intersection of incentives and fairness by developing a normative framework that embeds formal principles of distributive justice into allocation rules under incomplete information. In doing so, it complements and contrasts with several recent theoretical and empirical studies. Sinha and Anastasopoulos (2017) propose mechanisms that implement social objectives incorporating fairness by transforming agents' utilities through a concave aggregator. Their approach introduces fairness indirectly—through reduced variance in utilities—while maintaining implementability. In contrast, this study imposes fairness directly in the allocation of the surplus generated in the economy and characterizes allocation rules that preserve these fairness principles while inducing truthful behavior and effort provision in settings where agent productivity is private information. Velez (2011) also addresses the tension between fairness and incentives but does so in the context of indivisible goods and quasi-linear preferences. He shows that Generalized Money Rawlsian Fair (GMRF) solutions, which satisfy envy-freeness and efficiency, can be implemented in both Nash and strong Nash equilibria despite agents' incentives to misreport preferences. This work diverges by focusing on production rather than exchange and on contribution-based justice rather than envy-freeness. In particular, my fairness criteria depend on the marginal contribution of each agent to collective output, offering an alternative foundation for justice under informational constraints.

On the empirical side, several studies provide support for the behavioral and organizational significance of fairness. Ismail, Abdul-Halim, and Joarder (2015) find that distributive justice mediates the effect of career incentives on employee performance, suggesting that perceived fairness in outcomes enhances motivation and productivity. Similarly, Klein and Colauto (2020) investigate how perceptions of organizational justice within incentive contracts shape the alignment between personal and organizational goals. Their work, grounded in agency theory, emphasizes the behavioral dimensions of incentive design and shows that fairness–particularly distributive, procedural, and interactional justice–strengthens goal congruence in decentralized firms. By contrast, my analysis develops a theoretical foundation for embedding fairness directly into the structure of incentive schemes under asymmetric information. I characterize allocation rules that satisfy fairness principles while inducing truthful revelation and optimal effort from agents whose productivity is privately known. In doing so, my study provides the normative underpinnings for the types of justice perceptions that Klein and Colauto (2020) identify as empirically significant. Together, the two studies offer a complementary perspective: theirs documents the behavioral consequences of justice in practice, while mine shows how fairness can be operationalized through incentive design in environments of incomplete information. Finally, Sung, Choi, and Kang (2017) examine how incentive pay influences firm performance through employee commitment and competence, contingent on contextual factors such as procedural justice climate and environmental turbulence. Notably, they find that the effect of incentive pay on employee commitment turns negative in firms with a low procedural justice climate, emphasizing that when employees perceive the process of determining rewards as unfair, incentive pay can backfire. Their analysis focuses on procedural justice–the perceived fairness of the process by which outcomes are determined or distributed (Lind and Tyler, 1988). In contrast, this study centers on distributive justice, which concerns the perceived fairness of outcomes (Adams, 1965), particularly how the produced output is shared among contributors in environments with incomplete information. By formalizing fairness principles rooted in reciprocity and fair equality of opportunity, this study shows that fair outcome-based pay schemes can induce efficient, self-enforcing, and incentive-compatible behaviors, even under asymmetric information. Thus, while Sung, Choi, and Kang (2017) highlight the critical role of fair procedures in the success of incentive systems, this study demonstrates how fairness in outcome distribution structurally supports efficient participation and truthful behavior, providing theoretical underpinnings that generalize and deepen the understanding of their empirical findings. In sum, whereas existing empirical studies highlight the importance of perceived justice for performance, and existing theoretical studies explore fairness within complete information or utility-based frameworks, my contribution lies in demonstrating how fairness can be operationalized and sustained within a formal mechanism design framework under informational constraints, thereby linking normative principles to behavioral outcomes.

2.1.2 Organization of the chapter

The remainder of the chapter is organized as follows. I define the model and required concepts in section 2.2. A characterization of fair Bayesian economies is provided in section 2.3. Section 2.4 analyzes equilibrium existence and efficiency in fair Bayesian economies. Extensions of the basic model are considered in section 2.5. Application to mechanism design is provided in section 2.6. Section 2.7 concludes.

2.2 The model of a Bayesian economy: the basic model

This section formally introduces the model of a Bayesian economy and provides real-world examples to illustrate its application.

2.2.1 Model definition

An economy with incomplete information is defined as a tuple $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$, where:

- $N = \{1, 2, ..., n\}$ is a finite set of agents. Agents can be individuals, workers, firms, countries, and so on.

- Θ_i is a finite set of types (private information) of agent $i \in N$. $\Theta = \prod_{i \in N} \Theta_i$ is the set of agents' type profiles. Each element $\theta = (\theta_1, \theta_2, ..., \theta_n) \in \Theta$ determines a state of the economy. For notational convenience, I represent $\theta \in \Theta$ as $\theta = (\theta_i, \theta_{-i})$ given an agent $i \in N$. Here, θ_{-i} is the type profile of all agents except agent i, and Θ_{-i} denotes the set of all such type profiles.

- X_i is a finite set of actions of agent i. $X = \prod_{i \in N} X_i$ is the set of action profiles.

- $o_i \in X_i$ represents the inaction point of agent i, which reflects the scenario where the agent chooses to remain inactive in a given economic state. Let $o = (o_i)_{i \in N}$ denote the profile of agents' inaction points, representing a state where all agents choose to remain inactive.

- p_i is the belief function of agent i. It is a mapping from Θ_i to $\Delta(\Theta_{-i})$, the set of probability distributions over Θ_{-i}. That is, for any possible type $\theta_i \in \Theta$, p_i specifies a probability distribution $p_i(.|\theta_i)$ over the set Θ_{-i}, which represents agent i's beliefs about the types of other agents when their own type is θ_i.

- $f : \Theta \longrightarrow \mathbb{R}^X$ is a function that defines the technology in each state of the economy. It maps each type profile θ to a technology $f(\theta) = f_\theta : X \longrightarrow \mathbb{R}$ that satisfies $f_\theta(o) = 0$. f_θ is the technology (production function) of the economy in the state θ, and can also be interpreted as the surplus function. It maps each action profile $x \in X$ to a real number output $f_\theta(x) \in \mathbb{R}$.[4] I denote by $F(X) = \{g : X \longrightarrow \mathbb{R},$ such that $g(o) = 0\}$ the collection of technologies that generate zero output at the inaction profile.

[4]For the sake of exposition, I normalize the technology f_θ to zero at the inaction profile in any state θ. Alternative normalizations could be applied without impacting our results; in those cases, f_θ would represent the net surplus function.

- $\Phi : F(X) \times X \longrightarrow \mathbb{R}^n$ is a pay scheme that assigns each pair $(g, x) \in F(X) \times X$ to a vector $\Phi(g, x) = (\Phi_i(g, x))_{i \in N}$, where $\Phi_i(g, x)$ represents the pay of agent i associated to the output $g(x)$ generated using the technology g and the action profile x. It can also be interpreted as agent i's pay associated to the action profile x under the technology g.

- For any $i \in N$, $u_i : \Theta \times X \longrightarrow \mathbb{R}$ represents the utility function of agent i. It is defined for any state $\theta \in \Theta$ and any action profile $x \in X$ by $u_i(\theta, x) = \Phi_i(f_\theta, x)$.[5] In this basic model, I assume that the utility of an agent after a state of the economy θ has been realized and the action profile x has been taken by agents is determined by the pay they receive under the technology f_θ given the action profile x. The functional form of the agents' utility is kept simple, at some cost of realism, which will be addressed later.

The beliefs $(p_i)_{i \in N}$ are consistent if there is some common prior distribution over the set of type profiles Θ such that each agent's beliefs given his type are just the conditional probability distributions that can be computed from the prior distribution. That is, there exists some probability distribution $\mathbb{P} \in \Delta(\Theta)$ such that:

$$p_i(\theta_{-i}|\theta_i) = \frac{\mathbb{P}(\theta_i, \theta_{-i})}{\sum\limits_{t_{-i} \in \Theta_{-i}} \mathbb{P}(\theta_i, t_{-i})} \; \forall \theta_i \in \Theta_i;\; \theta_{-i} \in \Theta_{-i};\; \forall i \in N, \tag{2.1}$$

where $\sum\limits_{t_{-i} \in \Theta_{-i}} \mathbb{P}(\theta_i, t_{-i}) > 0$ for all $i \in N$. That is, the probability that an agent i is of type θ_i must be positive. This assumption is essential for the validity of equation 2.1 and for the relevance of equilibrium analysis in economies with incomplete information.

This study focuses on economies with incomplete information in which agents form consistent beliefs based on the prior probability distribution $\mathbb{P}$, updated according to Bayes' rule. Therefore, I will refer to these economies as *Bayesian economies*. In a Bayesian economy, differences in beliefs among agents can be explained by differences in information. Economies with incomplete information and inconsistent beliefs, where variations in beliefs cannot be attributed to differences in information but rather to subjective opinions, are not covered in this study and will deserve particular attention in future research.

2.2.2 Bayesian economies in the real world: some examples

The model of a Bayesian economy formalizes a wide range of real-life situations. Below are some illustrative examples.

[5] $u_i(\theta, x)$ can be any positive monotonic transformation of $\Phi_i(f_\theta, x)$, where the transformation may be different for each agent.

Example 6 (Firm) : *A firm can be modeled as a Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$, *where* N *is the set of workers,* Θ_i *is the set of abilities that worker* i *possesses,* X_i *is the set of effort levels he can supply. Each worker can choose to be inactive or shirk by exerting no effort, this is captured by the inaction point* o_i. *Each profile of workers' ability* $\theta \in \Theta$ *gives rise to a technology* $f(\theta) = f_\theta$ *that maps each profile of workers' effort* $x \in X$ *into a measurable output* $f_\theta(x)$. *Worker* i*'s pay associated to the output* $f_\theta(x)$ *is given by* $\Phi_i(f_\theta, x)$, *which coincides with the worker's utility in the basic model.*

Example 7 (Agrarian economy) : *An agrarian economy can be modeled as a Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$, *where* N *consists of a landowner (agent 1) and a finite number of laborers (agents 2 up to n) (Shapley and Shubik, 1967). The landowner has a parcel of land available for various agricultural purposes. The landowner' set of types* Θ_1 *can reflect her stance on environmental issues, ranging from ecological ignorance or pursuit of economic gains to ecological mindfulness. Her set of actions,* X_1, *includes different levels of fertilizer used on her land, with the inaction point* o_1 *being the complete avoidance of fertilizer use. A laborer's type is determined by his productivity, and his set of actions consists of varying levels of labor he can supply, with the point of inaction corresponding to zero labor supply. Each profile of agents' type* $\theta \in \Theta$ *gives rise to a production function* $f(\theta) = f_\theta$ *that maps each profile of agents' action* $x \in X$ *to the monetary value of the amount* $f_\theta(x)$ *of suitable crops that can be produced. The pay of agent* i *associated to the wealth* $f_\theta(x)$ *is given by* $\Phi_i(f_\theta, x)$.

Example 8 (Two-sided networked economy) : *A two-sided networked economy can be modeled as a Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$, *where* N *is the set of individuals partitioned into two sets* N_1 *and* N_2, *representing respectively the majority and the minority groups* ($|N_1| > |N_2|$). *For each* $i \in N$, Θ_i *is the set of abilities that individual* i *has to form new relationships within or across groups. Based on the degree of sociability, an individual can be "in-group oriented" (form links only with individuals of the same group), "other-group oriented" (form links only with individuals of the other group), and "mixed oriented" (form links with individuals of any group). An action of an individual is a vector* $x_i \in X_i = \{0, 1\}^n$, *which specifies* i*'s intention to form a direct link with other individuals. Thus,* $x_{ij} = 1$ *means that individual* i *is willing to have a relationship with individual* j, *otherwise* $x_{ij} = 0$. *I assume that* $x_{ii} = 1$ *for all* $i \in N$ *and that the inaction point is the null vector. Each profile of individual actions* $x \in X$ *gives rise to a unique network* $g(x)$. *Each profile of individuals' type* $\theta \in \Theta$ *gives rise to a function* $f(\theta) = f_\theta$ *that maps each profile of individuals' action* $x \in X$ *to a measure of the integration of the minority group in the network* $g(x)$. *Individual* i*'s contribution to the integration level* $f_\theta(x)$ *is given by* $\Phi_i(f_\theta, x)$.

Example 9 (Financial market and systemic risk) : *A financial market can be modeled as a Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$, where N represents a finite set of interconnected banks that are linked through financial relationships such as investment decisions. While this interconnectivity can have its advantages in promoting financial stability and economic growth, it also poses risks, such as the possibility of contagion, where the failure of one bank can have a ripple effect on others. The types available to an individual bank are determined by its lending capacity or its ability to finance risky projects, ranging from small-sized banks to large-sized banks. Large-sized banks are able to finance high-stake projects which require long-term and more risk-tolerant financing that other small or medium sized banks are either unwilling or ill-equipped to provide. Examples of such banks include national and multinational development banks. A bank's set of actions consists of its investment decisions in a variety of assets such as stocks, bonds, and real estate. More precisely, the actions set, X_i, of a bank $i \in N$ consists of vectors $a_i = (a_{i1}, a_{i2}, a_{i3})$, where a_{i1}, a_{i2}, and a_{i3} stand respectively for the levels of investment in stocks, bonds, and real estate. The inaction point is naturally defined by the null vector. Each profile of banks' type $\theta \in \Theta$ gives rise to a function $f(\theta) = f_\theta$ that maps each profile of banks' investment decisions $x \in X$ to a level of system-wide risk. Bank i's contribution to the system-wide risk $f_\theta(x)$ is given by $\Phi_i(f_\theta, x)$.*

Example 10 (International trade) : *An international trade can be modeled as a Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$, where N represents a finite set of countries involved in a global value chain (GVC). A country's type reflects its ability to participate in GVCs, which is generally determined by the country's endowment (natural resources, human capital, geographic and logistical factors) and political and economic factors. Thus a country's ability can be classified into low (L), medium (M), and high ability (H), so that $\Theta_i = \{L, M, H\}$ for any country i. Each country $i \in N$ can decide to participate ($x_i = 1$) or not to participate ($x_i = 0$) in the GVC, so country i's set of actions is $X_i = \{0, 1\}$. Each profile of countries' type $\theta \in \Theta$ gives rise to a function $f(\theta) = f_\theta$ that maps each profile of countries' participation decisions in the GVC, $x \in X$, to the value of trade flows generated by participating countries. Country i's contribution to the value of the trade flows generated, $f_\theta(x)$, is given by $\Phi_i(f_\theta, x)$.*

All these illustrations tend to demonstrate that our model of Bayesian economy is very general and lends itself to several applications.

The question of how should the pay scheme Φ of a Bayesian economy distribute the generated output among agents at any realized state of the economy is still pending. I address this question by assuming that the economy is "*fair*." The next section introduces the notion of fairness in a Bayesian economy and characterizes fair Bayesian economies.

2.3 Fairness in a Bayesian economy: an axiomatic foundation

This section introduces the concept of fairness in a Bayesian economy, focusing on the ideal of fairness-as-reciprocity.

A Bayesian economy, $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$, is fair if the pay scheme Φ satisfies basic principles of distributive justice. These principles are intuitive, self-justifying, and translate into the perceived fairness of the outcome. They have been applied to various contexts, but their formalization differs across studies.[6] I formalize these principles in our framework below. Before diving into it, some definitions and notations are in order.

Definition 7 : *Let $x \in X$ be a profile of actions. An agent $i \in N$ is active at x if he is not at the inaction point, that is, if $x_i \neq o_i$. If $x_i = o_i$, then I say that agent i is inactive at x. I denote by $|x| = |\{i \in N : x_i \neq o_i\}|$ the number of active agents at x.*

The following definition introduces a binary relation on the set of action profiles.

Definition 8 : *Let $x, a \in X$ be two profiles of actions and $i \in N$ an agent. a is induced by x, denoted by $a \trianglelefteq x$, if $\forall j \in N$, $a_j \neq o_j \implies a_j = x_j$. In other words, a is induced by x if every active agent at a chooses the same action as at x.[7] a is strictly induced by x, denoted $a \triangleleft x$, if $a \trianglelefteq x$ and $a \neq x$. a is strictly induced by x via agent i, denoted $a \triangleleft_o^i x$, if $a \triangleleft x$ and $a_i = o_i$.*

Next, I define the notion of an agent's marginal contribution.

Definition 9 : *Let $i \in N$ be a worker, and $x, a \in X$ two action profiles such that $a \triangleleft_o^i x$. Given a production function $g \in F(X)$, the marginal contribution of agent i relative to the action profiles a and x is given by:*

$$mc(i, g, a, x) = g(x_i, a_{-i}) - g(a),$$

where $(x_i, a_{-i}) \in X$ is the action profile in which agent i chooses x_i, and every other agent j chooses a_j.

Definition 10 : *Given a production function $g \in F(X)$, an agent $i \in N$ is said to be unproductive at the action profile x if its marginal contribution relative to any action profile a, $a \triangleleft_o^i x$, is*

[6] Some references include Shapley (1953), Myerson (1980), Young (1985) Tijs and Driessen (1986), Gopalakrishnan et al. (2021), Maniquet (2003), Demeze-Jouatsa, Pongou, and Tondji (2024), Aguiar, Pongou, and Tondji (2018), Pongou and Tondji (2018).

[7] For instance we have $(o_1, x_2, o_3, o_4, x_5, ..., x_n) \trianglelefteq (x_1, x_2, o_3, x_4, x_5, ..., x_n)$.

equal to zero, that is $mc(i,g,a,x) = 0$ for all $a \in X$ such that $a \lhd_o^i x$. Moreover, if $g \equiv f_\theta$, then I say that agent i is unproductive at the action profile $x \in X$ in the state of the economy $\theta \in \Theta$.

A permutation of N is a bijection of N onto itself. I denote by $\mathcal{S}_n$ the set of all permutations of N. For any permutation $\pi \in \mathcal{S}_n$ and any function $g \in F(X)$, I define the function $\pi g \in F(\pi X)$ as follows:

$$\begin{aligned} \pi g: \quad \pi X = \prod_{i \in N} X_{\pi(i)} &\longrightarrow \mathbb{R} \\ \pi x = \left(x_{\pi(i)}\right)_{i\in N} &\longmapsto \pi g(\pi x) = g(x) \end{aligned}$$

I now formalize the principles of distributive justice in a Bayesian economy.

Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$ be a Bayesian economy.

(EI) Ex-post Inactivity: Φ satisfies ex-post inactivity if for all $\theta \in \Theta$ and x such that $x_i = o_i$ for all $i \in N$,

$$\Phi_i(f_\theta, x) = 0 \text{ for all } i \in N.$$

(EE) Ex-post Efficiency: Φ is ex-post efficient if for all $\theta \in \Theta$, and $x \in X$,

$$\sum_{i\in N} \Phi_i(f_\theta, x) = f_\theta(x).$$

(EBF) Ex-post Bilateral Fairness: Φ satisfies ex-post bilateral fairness if for all $\theta \in \Theta$, $x \in X$,

$$\Phi_i(f_\theta, x) - \Phi_i(f_\theta, (o_j, x_{-j})) = \Phi_j(f_\theta, x) - \Phi_j(f_\theta, (o_i, x_{-i})) \text{ for all } i, j \in N.$$

(EU) Ex-post Unproductivity: Φ satisfies ex-post unproductivity if for any agent $i \in N$, $\theta \in \Theta$, and $x \in X$ such that agent i is unproductive in the state θ at the action profile x,

$$\Phi_i(f_\theta, x) = 0$$

(EA) Ex-post Anonymity: Φ is ex-post anonymous if for all $\theta \in \Theta$, $x \in X$, and $\pi \in \mathcal{S}_n$,

$$\Phi_{\pi(i)}(\pi f_\theta, \pi x) = \Phi_i(f, \theta) \text{ for all } i \in N$$

(EM) Ex-post Marginality: Φ satisfies ex-post marginality if for all $\theta, \theta' \in \Theta$, $x \in X$, and $i \in N$ such that $mc(i, f_\theta, a, x) \geq mc(i, f_{\theta'}, a, x)$ for all $a \lhd_o^i x$,

$$\Phi_i(f_\theta, x) \geq \Phi_i(f_{\theta'}, x)$$

Conditional on the state of economy $\theta \in \Theta$, EI is quite innocuous and requires zero pay to every agent at the inaction profile. EE requires the output produced to be entirely shared among agents. EBF states that claims between any pair of agents should be balanced at any action profile. Specifically, for any action profile $x \in X$, $\Phi_i(f_\theta, x)$ is agent i's pay if agent j takes action x_j and $\Phi_i(f_\theta, (o_j, x_{-j}))$ would be his pay if agent j decides to be inactive at x. So, the difference pay, $\Phi_i(f_\theta, x) - \Phi_i(f_\theta, (o_j, x_{-j}))$, represents the claim of agent j to agent i at x. Similarly, $\Phi_j(f_\theta, x) - \Phi_j(f_\theta, (o_i, x_{-i}))$ represents the claim of agent i to agent j at x. EBF requires the two claims to be equal, so no one agent should have a superior claim over another. EBF generalizes to our environment the balanced contribution axiom introduced by Myerson (1980). EU states that unproductive agents at an action profile should receive nothing. EA requires agents' pay to be independent of their labels or names. Importantly, EA implies that equally productive agents should be paid identically. Finally, EM states that if an agent is more productive in one state of the economy than in another state, then this agent should not be paid less in the former state.

Definition 11 : *A Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$ is a Myerson economy if the pay scheme Φ satisfies EI, EE and EBF.*

Definition 12 : *A Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$ is fair if the pay scheme Φ satisfies EE, EBF, EU, EA, and EM.*

It is obvious that a pay scheme that complies with EU will also satisfy EI. As a result, a Bayesian economy that is fair is also a Myerson economy. What about the converse? Are all Myerson economies fair? I show below that the two economies are equivalent.

Theorem 4 : *A Myerson economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i) \rangle$ admits a unique pay scheme $\Phi \equiv$ **EFP** defined for all $\theta \in \Theta$, $x \in X$, and any agent $i \in N$ by:*

$$EFP_i(f_\theta, x) = \sum_{a \lhd_o^i x} \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} mc(i, f_\theta, a, x)$$

*In addition, **EFP** satisfies EU, EA, EM.*

The value $EFP_i(f_\theta, x)$ is the weighted average marginal contribution of agent i in the state of the economy θ over the set of effort profiles that are strictly induced by x via agent i. So, $EFP_i(f_\theta, x)$ can be interpreted as the average contribution of agent i to the output level $f_\theta(x)$ in the state of the economy θ. This pay scheme generalizes the classical Shapley value (Shapley, 1953) in transferable-utility environments. In fact, if the action set of each agent is the pair $\{0, 1\}$, then the classical Shapley value of a transferable-utility game $v : 2^N \to \mathbb{R}$, is simply $EFP_i(f, x)$, where $x = (1, 1..., 1)$ and $f : y \mapsto f(y) = v(\{i \in N : y_i = 1\})$. The value $EFP_i(f_\theta, x)$ will be called the ex-post fair pay of agent i associated to the action profile x in the state of the economy θ. Indeed the ex-post fair pay extends the classical Shapley value as the action set of an agent can contain more than two options, but also different agents can have different action sets.

For clarity of exposition, the proof of Theorem 4 and all the subsequent results are provided in the Appendix.

As per Theorem 4, proving that a Bayesian economy is fair is equivalent to demonstrating its classification as a Myerson economy, which entails showing that its pay scheme adheres to EI, EE, and EBF criteria. In contrast, to demonstrate that a Bayesian economy is unfair, it suffices to show that its pay scheme doesn't comply with at least one of the fairness axioms: EE, EBF, EU, EA, and EM. Let us illustrate this with an example.

Example 11 : *Consider an economy with two workers, worker 1 and worker 2. Workers' types are determined by their abilities. Worker 1 has an average ability* (A)*, and worker 2 is of low ability* (L) *with a probability of 1/2 and high ability* (H) *with a probability of 1/2, so that the workers' type sets are given by* $\Theta_1 = \{A\}$ *and* $\Theta_2 = \{L, H\}$*. The type of worker 1 is common knowledge since* Θ_1 *is a singleton. Workers' actions are contingent upon the number of periods they are available to work. Each period can be seen as a four-hour time frame. The average worker is available to work only for two periods; while worker 2 can work for one or two periods. Assuming each worker can decide to be inactive (not working or working for zero period), it follows that workers' action sets are given by* $X_1 = \{0, 2\}$ *and* $X_2 = \{0, 1, 2\}$*. The production function of the economy in the states of the economy* (A, L) *and* (A, H) *are respectively given by:*

$$f_{(A,L)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 6 & if \quad x = (0,1) \\ 5 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 14 & if \quad x = (2,1) \\ 12 & if \quad x = (2,2) \end{cases} \quad \text{and } f_{(A,H)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 10 & if \quad x = (0,1) \\ 15 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 25 & if \quad x = (2,1) \\ 30 & if \quad x = (2,2) \end{cases}$$

We observe that the low-ability worker exerts a negative externality on the average worker since their combined output levels when working independently are greater than the level of output they generate when working together. Whereas the high-ability worker exerts a positive externality on the average worker since their combined output levels when working *independently are lower than the level of output they generate when working together. We also note that the low-ability worker is more productive when he works for one period than when he works for two periods. Let the utility functions, u_1 and u_2, for the two possible type profiles be given as in Table 2.1. Consider the pay scheme Φ defined for all workers $i \in N = \{1,2\}$, all states of the economy $\theta \in \Theta = \{(A,L),(A,H)\}$ and all action profiles $x \in X = X_1 \times X_2$ by: $\Phi(f_\theta, x) = u_i(\theta, x)$. The pay scheme Φ satisfies EI since all workers get a payoff of zero at the inaction point $o = (0,0)$ in every state of the economy. We also observe that the sum of workers' payoffs at every action profile and every type profile is equal to the associated output. Therefore, Φ satisfies EE. In addition, the claim of the average worker to the high-ability worker ($\theta = (A,H)$) when the action profile $x = (2,2)$ is taken is* 2.5 *since* $\Phi_1(f_{(A,H)},(2,2)) - \Phi_1(f_{(A,H)},(2,0)) = 12.5 - 10 = 2.5$, *which is equal to the claim of the latter to the former since* $\Phi_2(f_{(A,H)},(2,2)) - \Phi_2(f_{(A,H)},(0,2)) = 17.5 - 15 = 2.5$. *Therefore, claims between the two workers in the state of the economy $\theta = (A,H)$ at the action profile $x = (2,2)$ are balanced. It can be demonstrated similarly that this holds true for all type profiles and all effort profiles. Therefore, Φ satisfies EBF. It follows that the Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$ is fair.*

Table 2.1A: (u_1, u_2) at (A,L)

		x_2 = 0	x_2 = 1	x_2 = 2
x_1	0	(0, 0)	(0, 6)	(0, 5)
	2	(10, 0)	(9, 5)	(8.5, 3.5)

Table 2.1B: (u_1, u_2) at (A,H)

		x_2 = 0	x_2 = 1	x_2 = 2
x_1	0	(0, 0)	(0, 10)	(0, 15)
	2	(10, 0)	(12.5, 12.5)	(12.5, 17.5)

Table 2.1: pay scheme Φ of a fair Bayesian economy.

Example 12 : *Consider another Bayesian economy $\mathcal{B}'$ with the same characteristics as in $\mathcal{B}$ except for the utility u and the pay scheme Φ that are replaced by u' and Φ' described in Table 2.2. The pay scheme Φ' satisfies EI and EE, but violates EBF. In fact, given the action profile* $(2,1)$ *in the state of the economy* (A,L), *the low-ability worker has no claim to the average worker since* $\Phi_1(f_{(A,L)},(2,1)) - \Phi_1(f_{(A,L)},(2,0)) = 0$. *While the latter has a claim of* 3 *to the former since* $\Phi_2(f_{(A,L)},(2,1)) - \Phi_2(f_{(A,L)},(0,1)) = 3$. *It follows that the Bayesian economy $\mathcal{B}'$ is not a fair economy.*

Table 2.2A: (u'_1, u'_2) at (A, L)

		x_2 = 0	1	2
x_1	0	(0, 0)	(2, 4)	(2.5, 2.5)
	2	(7, 3)	(7,7)	(2, 10)

Table 2.2B: (u'_1, u'_2) at (A, H)

		x_2 = 0	1	2
x_1	0	(0, 0)	(3, 7)	(5, 10)
	2	(7, 3)	(2, 23)	(15, 15)

Table 2.2: pay scheme Φ' of a Bayesian economy that is unfair.

2.4 Equilibrium existence and efficiency in a fair Bayesian economy

In the previous section, I defined fairness principles in a Bayesian economy and characterized fair Bayesian economies. In this section, I show that these principles of fairness are sufficient to guarantee the existence of an equilibrium in Bayesian economies and establish conditions under which the equilibrium is unique and socially welfare-maximizing.

2.4.1 Equilibrium existence

I first show that a Bayesian economy can be modeled as a Bayesian game and define an equilibrium in a Bayesian economy as a pure strategy Bayesian Nash equilibrium of its associated Bayesian game.

Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ be a Bayesian economy. The Bayesian game associated to the Bayesian economy $\mathcal{B}$ is defined as $\Gamma^{\mathcal{B}} = \langle N, (\Theta_i), (X_i), (p_i), (u_i)\rangle$, where N is a finite set of players; Θ_i is a finite set of types for player i, and X_i is player i's set of actions; p_i is player i's beliefs about other players' types; $u_i : \Theta \times X \longrightarrow \mathbb{R}$ is the payoff function of player $i \in N$. That is, for any profile of actions and types $(\theta, x) \in \Theta \times X$, the function u_i specifies a number $u_i(\theta, x)$ that represents the payoff that player i would get if the players' actual types were all as in θ and players chose their actions as specified in x. In a fair Bayesian economy, $u_i(\theta, x)$ is defined as the ex-post fair pay associated to the production function f_θ and the action profile x. That is,

$$u_i(\theta, x) = EFP_i(f_\theta, x) = \sum_{a \lhd_0^i x} \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} [f_\theta(x_i, a_{-i}) - f_\theta(a)]$$

A pure strategy for a player i in the Bayesian game $\Gamma^{\mathcal{B}}$ is a function $s_i : \Theta_i \longrightarrow X_i$ that specifies player i's action choice for each realization of his type; $s_i(\theta_i)$ for a given $\theta_i \in \Theta$ would specify the pure action that player i would play if his type were θ_i. I denote by S_i the set of player i's pure strategies.

Let $s = (s_1, s_2, ..., s_n) \in S = \prod_{i \in N} S_i$ be a strategy profile and $i \in N$ a given player. Assume the current type of player i is θ_i. Then, the expected utility of player i is given by:

$$U_i((s_i, s_{-i})|\theta_i) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\, u_i(\theta_i, t_{-i}, s_i(\theta_i), s_{-i}(t_{-i}))$$

If the Bayesian economy is fair, then player i's expected utility becomes:

$$U_i((s_i, s_{-i})|\theta_i) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\, EFP_i\left(f_{(\theta_i,\, t_{-i})}, (s_i(\theta_i), s_{-i}(t_{-i}))\right)$$

Definition 13 : *Let $\Gamma^{\mathcal{B}}$ be the Bayesian game associated to a Bayesian economy $\mathcal{B}$. A strategy profile $(s_1^*, s_2^*, ..., s_n^*)$ is a pure strategy Bayesian Nash equilibrium of the game $\Gamma^{\mathcal{B}}$ if $\forall i \in N$; $\forall s_i \in S$; $\forall \theta_i \in \Theta_i$,*

$$U_i((s_i^*, s_{-i}^*)|\theta_i) \geq U_i((s_i, s_{-i}^*)|\theta_i)$$

Hence, a pure strategy Bayesian equilibrium of a Bayesian game is a strategy profile that maximizes each player's expected utility given their type and beliefs about other players' types and strategies.

In addition, the equilibrium $s^* = (s_1^*, s_2^*, ..., s_n^*)$ is said to be Pareto-efficient if it cannot be Pareto-dominated, that is, there is no strategy profile s, such that $U_i(s|\theta_i) \geq U_i(s^*|\theta_i)$ for all $i \in N$ and $\theta_i \in \Theta_i$, and $U_i(s|\theta_i) > U_i(s^*|\theta_i)$ for some i.

Definition 14 : *An equilibrium of a Bayesian economy is defined as a pure strategy Bayesian Nash equilibrium of its associated Bayesian game.*

A Bayesian economy may not admit an equilibrium. In fact, consider the Bayesian economy $\mathcal{B}'$ as defined in Example 12, with payoffs in every state of the economy given by Table 2.2. This economy doesn't admit an equilibrium: if player 1 decides to be inactive (by playing 0), then player 2's best response would be to choose 1 if his ability is low and 2 if his ability is high. But the best response of player 1 to this strategy would be to be active by choosing 2. Furthermore, if player 1 decides to be active (by playing 2), then player 2's best response would be to choose 2 if his ability is low and 1 if his ability is high. But the best response of player 1 to this strategy would be to be inactive. As a result, there is no pure strategy profile such that each player would play their best response, given their type. Therefore, no equilibrium exists for this Bayesian economy.

In analyzing the subgames of this Bayesian economy (Tables 2.2A and 2.2B), one might be tempted to argue that the economy has no equilibrium because none of its subgames have a (pure strategy) Nash equilibrium. This argument is misleading, as illustrated in Example 13 below, which provides a Bayesian economy that doesn't have an equilibrium, yet each of its subgames admits a Nash equilibrium.

Example 13 : *Consider a two-agent Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, \Phi, (u_i)\rangle$, *where* $N = \{1, 2\}$; $\Theta_1 = \{\theta_1\}$ *and* $\Theta_2 = \{\theta_{21}, \theta_{22}\}$; $X_1 = \{a_1, a_2\}$ *and* $X_2 = \{b_1, b_2\}$; *the inaction profile* $o = (b_1, a_1)$; *agent 1 has only one type* (θ_1), *therefore his type is perfectly known to player 2* ($p_2(\theta_1|\theta_{2j}) = 1, j = 1, 2$). *While agent 2 has two types* (θ_{21} *and* θ_{22}) *and agent 1's beliefs about agent 2's types are such that* $p_1(\theta_{21}|\theta_1) = p_1(\theta_{22}|\theta_1) = \frac{1}{2}$; *the production functions of the economy in the states of the economy* (θ_1, θ_{21}) *and* (θ_1, θ_{22}) *are respectively given by:*

$$f_{(\theta_1,\theta_{21})}(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ 3 & if \quad x = (a_1, b_2) \\ 1 & if \quad x = (a_2, b_1) \\ 2 & if \quad x = (a_2, b_2) \end{cases} \quad and \ f_{(\theta_1,\theta_{22})}(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ -1 & if \quad x = (a_1, b_2) \\ 1 & if \quad x = (a_2, b_1) \\ -2 & if \quad x = (a_2, b_2) \end{cases}$$

Let the utility functions, u_1 *and* u_2, *be given as in Table 2.3. Consider the pay scheme* Φ *defined for all workers* $i \in N = \{1, 2\}$, *all states of the economy* $\theta \in \Theta$ *and all action profiles* $x \in X = X_1 \times X_2$ *by:* $\Phi(f_\theta, x) = u_i(\theta, x)$. *Each subgame of this Bayesian economy admits a Nash equilibrium. In fact, the strategy profile* (a_2, b_1) *is the unique Nash equilibrium at the state* (θ_1, θ_{21}), *and the strategy profile* (a_1, b_2) *is the unique Nash equilibrium at the state* (θ_1, θ_{22}). *However, this Bayesian economy has no equilibrium. In fact, if agent 1 chooses the action* a_1, *then agent 2's best response would be to choose* b_2 *regardless of his type. The best response of agent 1 to this strategy depends on his expected utility. If he chooses* a_1, *then he would get an expected utility of* $-\frac{1}{2}$, *and if he chooses* a_2, *he would get an expected utility of 0. So agent 1's best response to agent 2 choosing* b_2 *at each type would be to choose* a_2. *Therefore, action* a_1 *cannot be played at any equilibrium. It can also be shown that action* a_2 *cannot be played at any equilibrium. It follows that this Bayesian economy doesn't have an equilibrium.*

Table 2.3A: (u_1, u_2) at (θ_1, θ_{21})

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0)	(1, 2)
	a_2	(0, 1)	(2, 0)

Table 2.3B: (u_1, u_2) at (θ_1, θ_{22})

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0)	(-2, 1)
	a_2	(-1, 2)	(-2, 0)

Table 2.3: Bayesian economy with no equilibrium.

The preceding two examples illustrate Bayesian economies for which no equilibrium exists. The common feature between these two economies is that they are not fair. In particular, their pay schemes violate the EBF principle. The following result ensures the existence of an equilibrium in fair Bayesian economies.

Theorem 5 : *Any fair Bayesian economy admits an equilibrium.*

Theorem 5 uncovers a new class of Bayesian games that always admit a pure strategy Bayesian Nash equilibrium. Along the same lines, it has implications for the existence of pure strategy Nash equilibrium for a wide class of finite strategic form games defined under complete information. Specifically, let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, EFP, (u_i) \rangle$ be a fair Bayesian economy. If the set of private information of each agent has only one element, that is $|\Theta_i| = 1 \ \forall i \in N$, then the players' private information becomes common knowledge among players, and the Bayesian game associated to this Bayesian economy is simply a finite strategic form game of complete information. Therefore, Theorem 5 uncovers a new class of finite strategic form games that always admit a pure strategy Nash equilibrium. The next example illustrates this finding.

Example 14 : *Consider the Bayesian economy $\mathcal{B}$ as defined in Example 11, with payoffs in every state of the economy recalled below. An equilibrium of this Bayesian economy is a pair (s_1^*, s_2^*), with $s_1^* \in X_1 = \{0, 2\}$ and $s_2^* : \Theta_2 = \{L, H\} \longrightarrow X_2 = \{0, 1, 2\}$ such that the following equations are satisfied:*

$$u_2(A, L, s_1^*, s_2^*(L)) \geq u_2(A, L, s_1^*, x_2) \ \forall x_2 \neq s_2^*(L) \tag{2.2}$$

$$u_2(A, H, s_1^*, s_2^*(H)) \geq u_2(A, H, s_1^*, x_2) \ \forall x_2 \neq s_2^*(H) \tag{2.3}$$

$$\frac{1}{2}\left[u_1(A, L, s_1^*, s_2^*(L)) + u_1(A, L, s_1^*, s_2^*(H))\right] \geq \frac{1}{2}\left[u_1(A, L, x_1, s_2^*(L)) + u_1(A, L, x_1, s_2^*(H))\right] \forall x_1 \neq s_1^* \tag{2.4}$$

Let the strategy profile $s^* = (s_1^*, s_2^*)$ *be defined as follows:*

$$s_1^* = 2 \ and \ s_2^*(\theta_2) = \begin{cases} 1 & if \quad \theta_2 = L \\ 2 & if \quad \theta_2 = H \end{cases}$$

It holds that: $u_2(A, L, 2, 1) = 5 \geq u_2(A, L, 2, 0) = 0$ *and* $u_2(A, L, 2, 1) = 5 \geq u_2(A, L, 2, 2) = 3.5$, *so Equation 2.2 is satisfied. Also,* $u_2(A, H, 2, 2) = 17.5 \geq u_2(A, H, 2, 0) = 0$ *and* $u_2(A, H, 2, 2) = 17.5 \geq u_2(A, H, 2, 1) = 12.5$*; so Equation 2.3 is also satisfied. Finally, when worker 2 plays the strategy* s_2^**, worker 1 gets an expected utility of* 10.75 *if he plays* $s_1^* = 2$ *and an expected utility of* 0 *if he plays* 0*, so Equation 2.4 is also satisfied. Therefore, an equilibrium of this economy is for worker 1 to work for two periods and worker 2 to work for one period if his ability is low and work for two periods if his ability is high. It is the unique equilibrium of the Bayesian economy.*

Table 2.4A: (u_1, u_2) at (A, L)

		$x_2 = 0$	$x_2 = 1$	$x_2 = 2$
x_1	0	(0, 0)	(0, 6)	(0, 5)
	2	(10, 0)	(9, 5)	(8.5, 3.5)

Table 2.4B: (u_1, u_2) at (A, H)

		$x_2 = 0$	$x_2 = 1$	$x_2 = 2$
x_1	0	(0, 0)	(0, 10)	(0, 15)
	2	(10, 0)	(12.5, 12.5)	(12.5, 17.5)

Table 2.4: Fair Bayesian economy with an equilibrium

2.4.2 Equilibrium efficiency

In the preceding section, I have established the existence of equilibrium in fair Bayesian economies. This, in turn, prompts an inquiry into the unique and efficient nature of the equilibrium within these environments. I begin by presenting two examples showing that the equilibrium may not be unique or Pareto-efficient in general.

Example 15 : *Consider a two-agent fair Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, EFP, (u_i)\rangle$*, where* $N = \{1, 2\}$*;* $\Theta_1 = \{\theta_1\}$ *and* $\Theta_2 = \{\theta_{21}, \theta_{22}\}$*;* $X_1 = \{a_1, a_2\}$ *and* $X_2 = \{b_1, b_2\}$*; the inaction profile* $o = (a_1, b_1)$*; agent 1 has only one type (*θ_1*), therefore his type is perfectly known to player 2 (*$p_2(\theta_1|\theta_2 j) = 1, j = 1, 2$*). While agent 2 has two types (*θ_{21} *and* θ_{22}*) and agent 1's beliefs about agent 2's types are such that* $p_1(\theta_{21}|\theta_1) = \frac{2}{3}$ *and* $p_1(\theta_{22}|\theta_1) = \frac{1}{3}$*; The production functions of the economy in the states of the economy* (θ_1, θ_{21}) *and* (θ_1, θ_{22}) *are respectively given by:*

$$f_{(\theta_1,\theta_{21})}(x) = \begin{cases} 0 & if \quad x \in \{(a_1,b_1),(a_2,b_2)\} \\ 1 & if \quad x \in \{(a_1,b_2),(a_2,b_1)\} \end{cases} \quad and \ f_{(\theta_1,\theta_{22})}(x) = \begin{cases} 0 & if \quad x = (a_1,b_1) \\ -5 & if \quad x = (a_1,b_2) \\ -2 & if \quad x = (a_2,b_1) \\ 9 & if \quad x = (a_2,b_2) \end{cases}$$

The agents' utility functions, u_1 and u_2, in each state of the economy are provided in Table 2.5 below. This economy has three equilibria, s^*, t^*, w^*, defined by:

$$s_1^* = a_1 \ and \ s_2^*(\theta_2) = \begin{cases} b_2 & if \quad \theta_2 = \theta_{21} \\ b_1 & if \quad \theta_2 = \theta_{22} \end{cases} \quad t_1^* = a_2 \ and \ t_2^*(\theta_2) = \begin{cases} b_1 & if \quad \theta_2 = \theta_{21} \\ b_2 & if \quad \theta_2 = \theta_{22} \end{cases}, \ and$$

$$w_1^* = a_2 \ and \ w_2^*(\theta_2) = \begin{cases} b_2 & if \quad \theta_2 = \theta_{21} \\ b_2 & if \quad \theta_2 = \theta_{22} \end{cases}$$

Table 2.5A: (u_1,u_2) at (θ_1,θ_{21})

		x_2: b_1	b_2
x_1	a_1	(0,0)	(0, 1)
	a_2	(1, 0)	(0, 0)

Table 2.5B: (u_1,u_2) at (θ_1,θ_{22})

		x_2: b_1	b_2
x_1	a_1	(0,0)	(0, -5)
	a_2	(-2, 0)	(6,3)

Table 2.5: Bayesian economy with multiple equilibria

Example 16 : *Consider a two-agent fair Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, EFP, (u_i)\rangle$, where $N = \{1,2\}$; $\Theta_1 = \{\theta_1\}$ and $\Theta_2 = \{\theta_{21}, \theta_{22}\}$; $X_1 = \{a_1, a_2, a_3\}$ and $X_2 = \{b_1, b_2, b_3\}$; the inaction profile $o = (a_1, b_1)$; agent 1 has only one type (θ_1), therefore his type is perfectly known to player 2 ($p_2(\theta_1|\theta_{2j}) = 1, j = 1,2$). While agent 2 has two types (θ_{21} and θ_{22}) and agent 1's beliefs about agent 2's types are such that $p_1(\theta_{21}|\theta_1) = \frac{7}{10}$ and $p_1(\theta_{22}|\theta_1) = \frac{3}{10}$; The production function of the economy is given in Table 2.6 below:*

Table 2.6A: technology at (θ_1, θ_{21})

		x_2: b_1	b_2	b_3
x_1	a_1	0	-4	11
	a_2	8	18	7
	a_3	-7	21	20

Table 2.6B: technology at (θ_1, θ_{22})

		x_2: b_1	b_2	b_3
x_1	a_1	0	-5	-38
	a_2	-1	20	47
	a_3	-28	27	50

Table 2.6: Production function of a fair Bayesian economy with inefficient equilibrium

Agents' utility functions, u_1 and u_2, in each state of the economy are provided in Table 2.7. This economy admits a unique equilibrium, s^*, defined by:

$$s_1^* = a_2 \text{ and } s_2^*(\theta_2) = \begin{cases} b_3 & if \quad \theta_2 = \theta_{21} \\ b_2 & if \quad \theta_2 = \theta_{22} \end{cases}$$

However, this equilibrium is not Pareto-efficient since it is Pareto-dominated by the strategy profile s defined by:

$$s_1 = a_3 \text{ and } s_2(\theta_2) = \begin{cases} b_2 & if \quad \theta_2 = \theta_{21} \\ b_3 & if \quad \theta_2 = \theta_{22} \end{cases}$$

In fact, the equilibrium s^* yields an expected utility of 5 to agent 1 and a utility of 5 to agent 2 if he is of type θ_{21} and a utility of 8 if he is of type θ_{22}. While the strategy profile s yields an expected utility of 153/10 to agent 1 and a utility of 12 to agent 2 if he is of type θ_{21} and a utility of 20 if he is of type θ_{22}. Therefore, this fair Bayesian economy admits a unique equilibrium that is inefficient.

Table 2.7A: (u_1, u_2) at (θ_1, θ_{21})

		x_2: b_1	b_2	b_3
x_1	a_1	(0,0)	(0, -4)	(0,11)
	a_2	(8, 0)	(15, 3)	(2,5)
	a_3	(-7, 0)	(9, 12)	(1,19)

Table 2.7B: (u_1, u_2) at (θ_1, θ_{22})

		x_2: b_1	b_2	b_3
x_1	a_1	(0,0)	(0, -5)	(0,-38)
	a_2	(-1, 0)	(12, 8)	(42,5)
	a_3	(-28, 0)	(2, 25)	(30,20)

Table 2.7: Bayesian economy with inefficient equilibrium

I have shown that the equilibrium of a fair Bayesian economy may not be unique (Example 15) or Pareto-efficient (Example 16). Now, I investigate conditions under which a fair Bayesian economy admits an equilibrium that is unique and Pareto efficient. Before diving into it, some definitions are in order.

Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ be a Bayesian economy and $i \in N$. A weak order over X_i is a binary relation that is transitive and complete. If each X_i ($i \in N$) is endowed with a weak order $\succeq_i$, then these different orders induce on $X = \prod_{i \in N} X_i$ a transitive binary relation denoted $\succeq$ and defined for all $x, y \in X$ by: $x \succeq y$ if $x_i \succeq_i y_i$ for all $i \in N$. However, the relation $\succeq$ need not be complete.

Definition 15 : *A Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ is ex-post weakly monotonic if its production function f is ex-post weakly monotonic. That is, for all $i \in N$ and $\theta_i \in \Theta_i$, there exists a weak order $\succeq_{\theta_i}$, such that for all $\theta \in \Theta$ and $x, y \in X$, $x \succeq_\theta y \Longrightarrow f_\theta(x) \geq f_\theta(y)$.*

Definition 16 : *A Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ is ex-post strictly monotonic if its production function f is ex-post strictly monotonic. That is, for all $i \in N$ and $\theta_i \in \Theta_i$, there exists a weak order $\succeq_{\theta_i}$, such that for all $\theta \in \Theta$ and $x, y \in X$, $x \succeq_\theta y$ and $x \neq y \Longrightarrow f_\theta(x) > f_\theta(y)$.*

A Bayesian economy is ex-post weakly monotonic if, for each type of each agent, there exists a weak order over the agent's action set such that the production function in each state of the economy weakly preserves the binary relation over the set of action profiles induced by the agents' types weak orders. An ex-post strictly monotonic Bayesian economy is defined similarly, except that the production function in each state of the economy strictly preserves the binary relation over the set of action profiles induced by the agents' types weak orders. The next result guarantees the existence and/or uniqueness of Pareto-efficient equilibrium in fair Bayesian economies, provided certain conditions are met.

Theorem 6 : *Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, EFP, (u_i)\rangle$ be a fair Bayesian economy. If the economy is ex-post weakly monotonic and the agents' beliefs p_i, for all $i \in N$, are drawn from uniform distributions, then there exists an equilibrium that is Pareto-efficient. In addition, the equilibrium is unique if the economy is ex-post strictly monotonic.*

In addition to establishing conditions that guarantee the existence of a Pareto-efficient equilibrium, Theorem 6 also provides conditions that rule out the possibility of multiple equilibria within the domain of fair Bayesian economies

Let us illustrate this finding with an example.

Example 17 : *Consider the fair Bayesian economy $\mathcal{B}$ as defined in Example 14, with the technology in every state of the economy recalled in Equation 2.5 below. We have seen that this Bayesian economy admits a unique equilibrium, s^*, given by: $s_1^* = 2$ and $s_2^* : s_2^*(L) = 1$ and $s_2^*(H) = 2$. At this equilibrium, worker 1 gets an expected utility of $43/4$ and worker 2 gets a utility level of 5 if his ability is low and a utility level of 17.5 if his ability is high. Moreover, there is no other strategy profile that gives each worker a utility level at least as high as the one he derives at the equilibrium. As a result, the unique equilibrium of the economy is also Pareto-efficient. This result holds essentially because the conditions of Theorem 6 are fulfilled. First, the Bayesian economy is fair (Example 11). Second, the beliefs of each worker are drawn from a uniform distribution. Third, the Bayesian economy is strictly ex-post monotonic. To see how this last condition is satisfied, consider the weak orders defined for each type of worker as follows: $\succeq_A$: $2 \succeq_A 0$; $\succeq_L$: $1 \succeq_L 2 \succeq_L 0$, and $\succeq_H$: $2 \succeq_H 1 \succeq_H 0$. These weak orders induce in each state of the economy $\theta \in \Theta = \{(A,L),\ (A,H)\}$ a transitive binary relation $\succeq_\theta$ over the set of action profiles $X = \{0,2\} \times \{0,1,2\}$. For all $\theta \in \Theta$ and $x, y \in X$ such that $x \succeq_\theta y$ and $x \neq y$ it holds that $f_\theta(x) > f_\theta(y)$. Therefore, this fair Bayesian economy is ex-post strictly monotonic. This implies that the economy admits a unique and Pareto-efficient pay scheme.*

$$f_{(A,L)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 6 & if \quad x = (0,1) \\ 5 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 14 & if \quad x = (2,1) \\ 12 & if \quad x = (2,2) \end{cases} \quad and \quad f_{(A,H)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 10 & if \quad x = (0,1) \\ 15 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 25 & if \quad x = (2,1) \\ 30 & if \quad x = (2,2) \end{cases} \tag{2.5}$$

I relate the existence of an equilibrium that is Pareto-efficient in the fair Bayesian economy of the preceding example to the fact that its production function is ex-post weakly monotonic. However, it is worth noting that the existence of an equilibrium is due to the fact that the economy is fair rather than the monotonicity property of its production function. To illustrate this, consider the Bayesian economy in Example 12. This economy is unfair and ex-post strictly monotonic, yet it doesn't admit an equilibrium.

2.5 Extensions of the basic model

In this section, I explore three extensions of the basic model. The aim is twofold: first, to examine how the results hold up when the model is made more realistic; and second, to demonstrate how the analytical tools can be used to analyze more complex models. The first extension introduces the cost of actions for each agent in each state of the economy. The second extension introduces redistributive taxation to support the worst-off agents of the economy. The final extension combines the two previous extensions, incorporating both the cost of actions and redistributive taxation.

2.5.1 Bayesian economy with costly action

The basic economic model assumes that the costs of agents' actions are zero or are already factored into the production function in each state of the economy. The latter means that the entire economy ultimately shoulders the action costs of its agents. This is not a realistic depiction of real-life organizations in which agents are fully or partly responsible for the costs of their actions. In this section, I extend the basic model developed in Section 2.2 to explicitly incorporate state of nature-contingent action costs. Specifically, after a state of the economy $\theta \in \Theta$ has been realized, each agent $i \in N$ in the economy chooses an action $x_i \in X$ and incurs a cost $c_i(x_i, \theta)$. The action profile $x \in X$ is then converted by the state-contingent technology f_θ into an output level $f_\theta(x) \in \mathbb{R}$, which is then shared among the agents according to the pay scheme Φ. Each agent receives a net utility equal to its output share minus the action cost he incurred. To summarize, the extended Bayesian economy is a tuple $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, (c_i), \Phi, (u_i)\rangle$, where for each $i \in N$, $c_i : \Theta \times X_i \longrightarrow \mathbb{R}$ is agent i's cost function, which is equal to zero at the inaction point, that is $c_i(\theta, o_i) = 0$ for all $i \in N$ and $\theta \in \Theta$. All other elements of the model are described as in the basic model with the exception of the agents' utility functions, which are now defined for all $i \in N$ and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = \Phi_i(f_\theta, x) - c_i(\theta, x_i).^{8}$$

A few observations are in order. First, I required the cost function of each agent to depend not only on his own type but also on the types of other agents as well. This functional form is very general and fits a wide range of situations, including those where the interaction between agents of different types can affect the action cost of each type differently. Second, in the basic model, the fairness axioms characterizing a Bayesian economy are captured in the pay scheme Φ. Therefore, the characterization result of fair Bayesian economies obtained in Section 2.3 remains valid for extended Bayesian economies as well as in subsequent extensions of the basic model, which will be discussed later. I then naturally defined a fair extended Bayesian

[8] The utility function $u_i(\theta, x)$ can also be defined as any positive monotonic transformation of $\Phi_i(f_\theta, x) - c_i(\theta, x_i)$, where the transformation may be different for each agent.

economy as an extended Bayesian economy in which the agents' utilities are defined for all $i \in N$ and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = EFP_i(f_\theta, x) - c_i(\theta, x_i).$$

The results of this model are summarized below.

Theorem 7 : *Any fair extended Bayesian economy admits an equilibrium. In addition, if the agents' beliefs p_i, for all $i \in N$, are drawn from uniform distributions and the production function F defined by Equation 2.6 below is ex-post weakly monotonic, then there exists an equilibrium that is Pareto-efficient. The equilibrium is unique if the production function F is ex-post strictly monotonic.*

$$\begin{array}{ccccccc} F: & \Theta & \longrightarrow & \mathbb{R}^X & & & \\ & \theta & \longmapsto & F_\theta: & X & \longrightarrow & \mathbb{R} \\ & & & & x & \longmapsto & F_\theta(x) = f_\theta(x) - \sum_{i \in N} c_i(\theta, x_i) \end{array} \tag{2.6}$$

The argument here is an extension of the idea used in the basic model. However, the analog of Theorem 6, uncovering a condition under which the equilibrium is efficient and unique, is now a bit subtler and makes use of a generalized production function that includes the cost functions of all agents. The next example illustrates this finding.

Example 18 : *Consider the extended Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, (c_i), \Phi, (u_i)\rangle$, where $N, (\Theta_i), (X_i), (o_i), (p_i), f,$ and Φ are defined as in Example 11. The cost functions of workers are defined by:*

$$c_1(\theta, x_1) = \begin{cases} 0 & if \quad \theta \in \Theta \text{ and } x_1 = 0 \\ 8 & if \quad \theta = (A, L) \text{ and } x_1 = 2 \\ 5 & if \quad \theta = (A, H) \text{ and } x_1 = 2 \end{cases} \quad \text{and } c_2(\theta, x_2) = \begin{cases} 0 & if \quad \theta \in \Theta \text{ and } x_2 = 0 \\ 7 & if \quad \theta = (A, L) \text{ and } x_2 = 1 \\ 3 & if \quad \theta = (A, H) \text{ and } x_2 = 1 \\ 10 & if \quad \theta = (A, L) \text{ and } x_2 = 2 \\ 4 & if \quad \theta = (A, H) \text{ and } x_2 = 2 \end{cases} \tag{2.7}$$

Workers' utility functions in each state of the economy are provided in Table 2.8 below. This fair extended Bayesian economy admits a unique equilibrium $s^* = ((s_1^*, s_2^*))$ defined by:

$$s_1^* = 2 \text{ and } s_2^*(\theta_2) = \begin{cases} 0 & if \quad \theta_2 = L \\ 2 & if \quad \theta_2 = H \end{cases}$$

This equilibrium differs from the one obtained in the fair Bayesian economy in Example 11, in which agents' action costs are assumed to be zero in each state of the economy. If we relax this assumption and require the cost functions to be non-zero and defined by Equation 2.7, we find that worker 2 chooses to be inactive at the equilibrium if his ability is low. This is because it is now too costly for him to work for one period at the equilibrium, as found in Example 11. In addition, the unique equilibrium of the economy is Pareto-efficient. It can be shown that the production function F defined for all $\theta \in \{(A,L),(A,H)\}$ and all $x \in X$ by $F_\theta(x) = f_\theta - c_1(\theta, x_1) - c_2(\theta, x_2)$ is ex-post weakly monotonic, with the weak orders defined for each worker's type as follows: $\geq_A$: $2 \geq_A 0$; $\geq_L$: $0 \geq_L 1 \geq_L 2$, and $\geq_H$: $2 \geq_H 1 \geq_H 0$.

Table 2.8A: (u_1, u_2) at (A, L)

		x_2 = 0	1	2
x_1	0	(0, 0)	(0, -1)	(0, -5)
	2	(2, 0)	(1, -2)	(0.5, -6.5)

Table 2.8B: (u_1, u_2) at (A, H)

		x_2 = 0	1	2
x_1	0	(0, 0)	(0, 7)	(0, 11)
	2	(5, 0)	(7.5, 9.5)	(7.5, 13.5)

Table 2.8: Workers' payoffs in a fair extended Bayesian economy.

2.5.2 Fairness as fair-equality-of-opportunity: the inclusive Bayesian economy

In a fair Bayesian economy, the ex-post unproductivity axiom requires that the agents who contribute nothing in generating the surplus in a given state of the economy should receive nothing. While this can be justified in low-income economies, which promote minimal assistance due to economic development and social progress, high-income economies may be viewed as standing against this axiom by promoting solidarity toward low-productive agents, such as people with disabilities. This section aims to design a variation of the basic model that captures the idea of caring for the least advantageous agents of the economy while maintaining most of the fairness principles characterizing a fair Bayesian economy. Put differently, the objective is to address the problem of income inequality in a fair Bayesian economy and ensure a more equitable distribution of resources within the economy. Several techniques can be used to reduce inequalities in an economy, and the desirability and effectiveness of each technique can vary based on the economic and political context of the economy. In this study, I propose a simple and easy-to-implement policy to reduce income inequality in a fair Bayesian economy. The policy consists of taxing ex-post the wealth generated in the economy and redistributing it equally among all agents, irrespective of their productivity. The remaining wealth is then allocated according to the fairness principles outlined in the basic model.[9] This results in an economy that I call an inclusive and fair Bayesian economy. Specifically, in an inclusive and fair Bayesian economy, after a state of the economy $\theta \in \Theta$ has been realized, each agent $i \in N$ of the economy chooses an action $x_i \in X$. This gives rise to an action profile $x \in X$ that is converted by the state-contingent technology f_θ into an output level $f_\theta(x) \in \mathbb{R}$, of which a share is taxed for redistribution at a state-contingent rate $T(\theta) \in [0,1)$ and the remaining share is allocated among agents according to the ex-post fair pay. Each agent $i \in N$ then receives a payment, called the ex-post inclusive pay, denoted EIP_i^T, defined for all $\theta \in \Theta$ and $x \in X$ by:

$$EIP_i^T(f_\theta, x) = (1 - T(\theta))EFP_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n}.$$

Three observations are in order. First, to maintain fairness, I require the tax rate to be different from 1, as a maximum tax rate would compromise most of the fairness principles anchored in fair Bayesian economies, resulting in a purely egalitarian economy where the output produced is equally distributed among all agents, regardless of their productivity. I rule out this possibility by assuming $T(\theta) < 1$ for all $\theta \in \Theta$. Second, I allow the transfer received by each agent to vary depending on the level of wealth produced in the economy. As a result, the higher the wealth generated by the economy, the higher the transfer received by everyone. In particular, unproductive agents would enjoy higher social assistance if the economy's productivity return is high. Third, I allow the tax rate to be different in different states of the economy. This is reasonable and leaves room for fiscal adjustments in case the economy is hit by an unexpected shock that can affect the actions taken by agents.

Furthermore, the ex-post inclusive pay is a generalization of the ex-post fair pay to include a value of social justice, ensuring that every agent in a productive economy receives something to cover basic needs. It is equal to the ex-post fair pay if the tax rate is equal to zero. A non-zero tax rate is informative of the extent to which the economy promotes social justice and captures the trade-off between market justice and social justice. The ex-post inclusive pay is a generalization of the egalitarian Shapley value (Casajus and Huettner, 2014) defined for cooperative games. It is important to mention that our framework is more general than that of cooperative games, where agents' actions are restricted to two options–*cooperate* and *not cooperate*. In addition, our model goes beyond the traditional assumption of full information considered in cooperative games and allows each agent to have private information.

[9]For simplicity, I assume the tax is levied and redistributed efficiently, with no associated losses. Accounting for potential deadweight loss of taxation would not impact the main finding of the model.

Another important point to make is that the transition from a fair Bayesian economy to an inclusive and fair Bayesian economy is achieved through progressive taxation, where a higher tax rate is imposed on the well-off agents of the economy. This transition can also be justified on the grounds of the Pigou Dalton principle (Shorrocks and Foster, 1987; Bosmans, Lauwers, and Ooghe, 2009), which consists of a non-leaky and non-rank-switching transfer from high-income to low-income agents.[10] To illustrate this, consider a fair Bayesian economy with three individuals. In a given state of the economy, assume the contributions of individuals 1 and 2 to the economy's surplus are respectively $\frac{3}{4}$ and $\frac{1}{4}$. Individual 3 is unproductive and contributes nothing. If the economy produces a total surplus of 1000, then individual 1 would receive a payment of 750, individual 2 would receive 250, and individual 3 would receive 0, so that the distribution of individual income in the fair Bayesian economy is $(750, 250, 0)$. Now assume that the total wealth of 1000 is taxed at a rate equal to $\frac{3}{10}$ for redistribution. This taxation will generate a total wealth of 300 so that each individual would receive a minimum payment of 100. The remaining wealth of 700 is redistributed as in the fair Bayesian economy. As a result, individual 1 would receive 525, individual 2 would receive 175, and individual 3 would receive 0. The distribution of individual income in the inclusive and fair Bayesian economy is then $(625, 275, 100)$. The taxation is progressive because if we compare the individuals' revenues in the inclusive and fair Bayesian economy relative to what they get in the fair Bayesian economy, it is as if individual 1 paid a tax of 125, individual 2 received a subsidy of 25, and individual 3 received a subsidy of 100. In addition, the transfer hasn't altered the initial rank of individuals in the fair Bayesian economy as individual 1 remains the wealthiest, followed by individual 2 and then individual 3.

To summarize, the inclusive Bayesian economy is a tuple $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, T, \Phi, (u_i) \rangle$, where $T : \Theta \longrightarrow [0, 1)$ is the tax rate function. All the other elements of the model are described as in the basic model with the exception of the agents' utility functions, which are now defined for $i \in N$ and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = (1 - T(\theta))\Phi_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n}.$$[11]

I then naturally define an inclusive and fair Bayesian economy as an inclusive Bayesian economy in which the agents' utilities are given by the inclusive ex-post fair pay defined for all $i \in N$

[10] A non-leaky transfer is a transfer that is performed without any loss, while a non-rank-switching transfer means transferring without any change in the initial ranking.

[11] The utility fucntion $u_i(\theta, x)$ can also be defined as any positive monotonic transformation of $(1 - T(\theta))\Phi_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n}$, where the transformation may be different for each agent.

and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = (1 - T(\theta))EFP_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n}$$

The results of this model are summarized below.

Theorem 8 : *Any inclusive and fair Bayesian economy admits an equilibrium. In addition, if the economy is ex-post weakly monotonic and the agents' beliefs p_i, for all $i \in N$, are drawn from uniform distributions, then there exists an equilibrium that is Pareto-efficient. The equilibrium is unique if the economy is ex-post strictly monotonic.*

The next example illustrates this finding.

Example 19 : *Consider the extended Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, T, \Phi, (u_i)\rangle$, where $N, (\Theta_i), (X_i), (o_i), (p_i), f,$ and Φ are defined as in Example 11. The tax rate function T is defined by:*

$$T(\theta) = \begin{cases} \frac{1}{4} & if \quad \theta = (A, L) \\ \frac{3}{4} & if \quad \theta = (A, H) \end{cases} \tag{2.8}$$

Workers' utility functions in each state of the economy are provided in Table 2.9 below. This fair extended Bayesian economy admits a unique equilibrium $s^* = ((s_1^*, s_2^*))$ defined by:

$$s_1^* = 2 \ and \ s_2^*(\theta_2) = \begin{cases} 1 & if \quad \theta_2 = L \\ 2 & if \quad \theta_2 = H \end{cases}$$

This equilibrium coincides with the one obtained in the fair Bayesian economy of Example 11, in which the tax rate in each state of the economy is assumed to be zero. However, this need not be the case in general. In addition, the unique equilibrium of the economy is Pareto-efficient, which stems from the fact the economy is ex-post strictly monotonic, as shown in Example 17.

Table 2.9A: (u_1, u_2) at (A, L)

		x_2 = 0	x_2 = 1	x_2 = 2
x_1	0	(0, 0)	(0.75, 5.25)	(0.625, 4.375)
	2	(8.75, 1.25)	(8.5, 5.5)	(7.875, 4.125)

Table 2.9B: (u_1, u_2) at (A, H)

		x_2 = 0	x_2 = 1	x_2 = 2
x_1	0	(0, 0)	(3.75, 6.25)	(5.625, 9.375)
	2	(6.25, 3.75)	(12.5, 12.5)	(14.375, 15.625)

Table 2.9: Workers' payoffs in a fair inclusive Bayesian economy.

2.5.3 A Bayesian economy with costly actions and redistributive taxation

In this extension, I consider a comprehensive Bayesian economy that encompasses the cost of actions incurred by the agents, as well as the implementation of redistributive taxation to support unproductive agents. Formally, a comprehensive Bayesian economy is a tuple $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, T, (c_i), \Phi, (u_i) \rangle$, where each element is described as in the previous extensions with the exception of the agents' utility functions, which are now defined for all $i \in N$ and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = (1 - T(\theta))\Phi_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n} - c_i(\theta, x_i).$$

I then naturally define a fair comprehensive Bayesian economy as a comprehensive Bayesian economy in which the agents' utilities are defined for all $i \in N$ and $(\theta, x) \in \Theta \times X$ by:

$$u_i(\theta, x) = (1 - T(\theta))EFP_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n} - c_i(\theta, x_i).$$

The results of this model are summarized below.

Theorem 9 : *Any fair comprehensive Bayesian economy admits an equilibrium.*

The analog of Theorem 6, which ensures the existence of a Pareto-efficient equilibrium under the condition that the (generalized) production function of the economy is ex-post weakly monotonic, no longer holds when the costs of agents' actions are non-zero and redistributive taxation is introduced into the basic model. The next example illustrates this point.

Example 20 : *Consider a simple fair comprehensive economy $\mathcal{B}$ with two agents. Agents' types are perfectly known, $\Theta_1 = \{\theta_1\}$ and $\Theta_2 = \{\theta_2\}$, resulting in only one state of the economy $\theta = (\theta_1, \theta_2)$, which is dropped from some expressions. The agents' action sets are given by $X_1 = \{a_1, a_2\}$ and $X_2 = \{b_1, b_2\}$. The inaction profile is (a_1, b_1). The production function of the economy is given by:*

$$f(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ 2 & if \quad x = (a_2, b_1) \\ 4 & if \quad x = (a_1, b_2) \\ 6 & if \quad x = (a_2, b_2) \end{cases}$$

The cost functions of agents are such that: $c_1(a_1) = c_2(b_1) = 0$, $c_1(a_2) = 2$ *and* $c_2(b_2) = 3$. *Assume the tax rate is such that* $T(\theta) = 0.9$. *It follows that the utility function of an agent* $i = 1, 2$ *is given by* $u_i(x) = 0.1EFP_i(f, x) + 0.45f(x) - c_i(x_i)$ *for all* $x \in X$. *The utility levels enjoyed by each agent at different action profiles are outlined in Table 2.10 below. This economy admits a unique equilibrium, namely the inaction profile* (a_1, b_1). *Consistent with Theorem 7, the generalized production function* F *of this economy is defined by:*

$$F(x) = f(x) - c_1(x_1) - c_2(x_2)$$

Let's consider the weak orders, $\succeq_1$ *and* $\succeq_2$ *defined respectively over the sets* X_1 *and* X_2 *by:* $a_1 \succeq_1 a_2$ *and* $b_2 \succeq_2 b_1$. *It follows that the production function* F *is weakly monotonic. However, the unique equilibrium of the economy is not Pareto-efficient since it is Pareto-dominated by the strategy profile* (a_2, b_2).

		x_2	
		b_1	b_2
x_1	a_1	(0,0)	(1.8, -0.8)
	a_2	(-0.9, 0.9)	(0.9, 0.1)

Table 2.10: Fair comprehensive Bayesian economy with a Pareto-inefficient equilibrium

The next example is a non-trivial combination of the two preceding extensions and illustrates a fair comprehensive Bayesian economy that admits a Pareto-efficient equilibrium.

Example 21 : *Consider the fair comprehensive Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, T, (c_i), EFP, (u_i)\rangle$, *where each element of the model is defined as in Examples 18 and 21. Workers' utility functions in each state of the economy are provided in Table 2.11 below. This fair extended Bayesian economy admits a unique equilibrium* $s^* = ((s_1^*, s_2^*))$ *defined by:*

$$s_1^* = 2 \text{ and } s_2^*(\theta_2) = \begin{cases} 0 & if \quad \theta_2 = L \\ 2 & if \quad \theta_2 = H \end{cases}.$$

In addition, the unique equilibrium of the economy is Pareto-efficient.

Table 2.11A: (u_1, u_2) at (A, L)

x_1 \ x_2	0	1	2
0	(0, 0)	(0.75, -1.75)	(0.625, -5.625)
2	(0.75, 1.25)	(0.5, -1.5)	(-0.125, -5.875)

Table 2.11B: (u_1, u_2) at (A, H)

x_1 \ x_2	0	1	2
0	(0, 0)	(3.75, 3.25)	(5.625, 5.375)
2	(1.25, 3.75)	(7.5, 9.5)	(9.375, 11.625)

Table 2.11: Workers' payoffs in a fair comprehensive Bayesian economy.

2.6 Application to mechanism design

The final section of this study explores the application of our model within the realm of mechanism design. Specifically, the goal consists of developing a mechanism that incentivizes agents to truthfully reveal their private information within a fair Bayesian economy. To begin, let's first review what a mechanism entails in our framework.

Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i) \rangle$ be a Bayesian economy. A mechanism is defined as a function $\mathcal{M} : \Theta \longrightarrow X$. This means that a mechanism prescribes to agents the actions they should take for every realization of their types: $\mathcal{M}(\theta)$ specifies the actions chosen by agents when their types are given by θ. I denote by $\mathcal{M}_i(\theta)$ the action prescribed by the mechanism $\mathcal{M}$ to agent i when the type profile θ is realized.

Once the mechanism's definition is provided, the next step is to identify the desired action that the mechanism designer would like agents to choose. In pursuit of this objective, I consider a productivity-oriented social planner who seeks to induce agents to opt for the action profile that generates the highest output level. Achieving a higher output level benefits the economy as a whole by driving economic growth, improving efficiency, and raising income levels. Therefore, I am interested in the following mechanism:

$$\begin{aligned} \mathcal{M}^* : \quad \Theta &\longrightarrow X \\ \theta &\longmapsto \mathcal{M}^*(\theta) = arg\max_{x \in X} f_\theta(x) \end{aligned}$$

The mechanism $\mathcal{M}^*$ prescribes agents to take their optimal action based on their true type. These actions are deemed optimal as they maximize the output given the prevailing technology. In the context of a firm, where a worker's type is determined by their productivity and their action set represents various levels of effort, the best action for a worker is to supply the highest effort level given their productivity. Whether agents will follow the prescriptions of the mechanism depends on the incentives in place. I explore this issue within the framework of fair Bayesian economies, where incentives are shaped by principles of fairness.

A mechanism $\mathcal{M}$ induces a game among the agents of the economy. The timing of the game is described as follows:

i) Nature draws the type $\theta_i \in \Theta$ of each agent $i \in N$, giving rise to the state of the economy θ, and therefore the technology f_θ which is fixed.

ii) Each agent $i \in N$ learns his type θ_i and reports $\widehat{\theta_i}$.

iii) Agents take the actions $\mathcal{M}(\widehat{\theta})$, based on their reports.

iv) Each agent i gets a utility level $v_i(\mathcal{M}(\widehat{\theta})) = u_i(\theta, \mathcal{M}(\widehat{\theta})) = EFP_i(f_\theta, \mathcal{M}(\widehat{\theta}))$.

An important observation to highlight is that a misreport of type by an agent does not alter the underlying technology of the economy but can result in a change of the actions prescribed by the mechanism. So, misreporting types leads to a change of actions, not to a change of technology.

After detailing the essential components of a mechanism, I now shift focus to the desirable properties that a mechanism should exhibit. The literature on mechanism design has placed considerable emphasis on three key properties: strategy-proofness, individual rationality, and Pareto-efficiency. Below, I present their definitions and formalization within our framework.

A mechanism $\mathcal{M}$ is said to be strategy-proof (or incentive-compatible) in a fair Bayesian economy $\mathcal{B}$ if no agent can ever benefit by reporting a type that is different from their true one. That is, if there does not exist $\theta \in \Theta$, $i \in N$, and $\widehat{\theta_i} \in \Theta_i$ such that $v_i(\mathcal{M}(\widehat{\theta_i}, \theta_{-i})) > v_i(\mathcal{M}(\theta))$. A mechanism is strategy-proof if it is strategy-proof in any fair Bayesian economy.

A mechanism $\mathcal{M}$ is individually rational in a fair Bayesian economy $\mathcal{B}$ if each agent is at least as better-off choosing the action prescribed by the mechanism as they would be by remaining inactive, given a type profile. That is, for each $i \in N$ and $\theta \in \Theta$, $v_i(\mathcal{M}(\theta)) \geq v_i(o_i, \mathcal{M}_{-i}(\theta))$. A mechanism is individually rational if it is individually rational in any fair Bayesian economy.

A mechanism $\mathcal{M}$ is Pareto-efficient in a fair Bayesian economy $\mathcal{B}$ if, given a type profile and the action profile prescribed by the mechanism, there is no action profile that can make every agent better-off without harming another. That is, for all $\theta \in \Theta$, there is no $x \in X$ such that for each $i \in N$, $u_i(\theta, x) \geq u_i(\theta, \mathcal{M}(\theta))$, and for some $i \in N$, $u_i(\theta, x) > u_i(\theta, \mathcal{M}(\theta))$. A mechanism is Pareto-efficient if it is Pareto-efficient in any fair Bayesian economy.

With the properties of a mechanism formally defined, I now turn to examining which of these properties are satisfied by the mechanism $\mathcal{M}^*$. The next proposition shows that it is Pareto-efficient in any fair Bayesian economy.

Proposition 1 : *The mechanism $\mathcal{M}^*$ is Pareto-efficient.*

However, the mechanism $\mathcal{M}^*$ may not be strategy-proof or individually rational, as demonstrated in the next two examples.

Example 22 : *Consider a two-agent fair Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, EFP, (u_i)\rangle$, *where* $N = \{1,2\}$; $\Theta_1 = \{\theta_1\}$ *and* $\Theta_2 = \{\theta_{21}, \theta_{22}\}$; $X_1 = \{a_1, a_2\}$ *and* $X_2 = \{b_1, b_2\}$; *the inaction profile* $o = (a_1, b_1)$; *the production functions of the economy in the states* (θ_1, θ_{21}) *and* (θ_1, θ_{22}) *are respectively given by:*

$$f_{(\theta_1,\theta_{21})}(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ 4 & if \quad x = (a_1, b_2) \\ 10 & if \quad x = (a_2, b_1) \\ 8 & if \quad x = (a_2, b_2) \end{cases} \quad and \ f_{(\theta_1,\theta_{22})}(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ 1 & if \quad x = (a_1, b_2) \\ 4 & if \quad x = (a_2, b_1) \\ 7 & if \quad x = (a_2, b_2) \end{cases}$$

The agents' utility functions, u_1 *and* u_2, *in each state of the economy are provided in Table 2.12. We have that:* $\mathcal{M}^*(\theta_1, \theta_{21}) = (a_2, b_1)$ *and* $\mathcal{M}^*(\theta_1, \theta_{22}) = (a_2, b_2)$. *Agent 1 has only one type, so she cannot engage in any strategic behavior. Under the mechanism* $\mathcal{M}^*$, *she will choose action* a_2, *regardless of agent 2's type. This is not the case for agent 2, who may take advantage of the mechanism if it is profitable to do so. Assume the state of the economy* $\theta = (\theta_1, \theta_{21})$ *is realized so that agent 2's true type is* θ_{21}. *If agent 2 is truthful, then he would get a utility level equal to* $v_2(\mathcal{M}^*(\theta)) = u_2(\theta, (a_2, b_1)) = 0$. *In contrast, if he lies about his type and reports the type* θ_{22}, *then his utility level would be* $v_2(\mathcal{M}(\theta_1, \theta_{22})) = u_2((\theta_1, \theta_{21}), (a_2, b_2)) = 2$. *Thus, agent 2 benefits by reporting the type* θ_{22} *when his true type is* θ_{21}. *Hence, the mechanism* $\mathcal{M}^*$ *is not strategy-proof.*

Table 2.12A: (u_1, u_2) at (θ_1, θ_{21})

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0)	(0, 4)
	a_2	(10, 0)	(7, 1)

Table 2.12B: (u_1, u_2) at (θ_1, θ_{22})

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0)	(0, 1)
	a_2	(4, 0)	(5,2)

Table 2.12: Fair Bayesian economy in which $\mathcal{M}^*$ is not strategy-proof

Example 23 : *Consider a three-agent fair Bayesian economy* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, EFP, (u_i)\rangle$, *where* $N = \{1,2,3\}$; $\Theta_1 = \{\theta_1\}$, $\Theta_2 = \{\theta_2\}$, *and* $\Theta_3 = \{\theta_3\}$; $X_1 = \{a_1, a_2\}$, $X_2 = \{b_1, b_2\}$, *and* $X_3 = \{c_1, c_2\}$; *the inaction profile* $o = (a_1, b_1, c_1)$. *The agent types' sets* *contain only one element, so agents' types are common knowledge. The production function of the economy in the unique state* $\theta = (\theta_1, \theta_2, \theta_3)$ *is given by:*

$$f_\theta(x) = \begin{cases} 0 & if \quad x = (a_1, b_1, c_1) \\ 3 & if \quad x = (a_2, b_1, c_1) \\ 8 & if \quad x = (a_1, b_2, c_1) \\ 5 & if \quad x = (a_2, b_2, c_1) \\ 14 & if \quad x = (a_1, b_1, c_2) \\ 4 & if \quad x = (a_2, b_1, c_2) \\ 18 & if \quad x = (a_1, b_2, c_2) \\ 20 & if \quad x = (a_2, b_2, c_2) \end{cases}$$

The agents' utility functions, u_1 *and* u_2*, in each state of the economy are provided in Table 2.12. We have that:* $\mathcal{M}^*(\theta) = (a_2, b_2, c_2)$*. If agent 1 participates in the mechanism* $\mathcal{M}^*$*, then she will get a utility level equal to* $v_1(\mathcal{M}^*(\theta)) = u_1(\theta, (a_2, b_2, c_2)) = -\frac{1}{2}$*. In contrast, if she decides not to participate and stays inactive while other agents participate, then she will get a utility level equal to* $v_1(a_1, \mathcal{M}^*_{-1}(\theta)) = u_1(\theta, (a_1, b_2, c_2)) = 0$*. Thus, agent 1 has an incentive not to participate in the mechanism. Hence, the mechanism* $\mathcal{M}^*$ *is not individually rational.*

Table 2.13A: (u_1, u_2, u_3) with $x_3 = c_1$

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0,0)	(0,8,0)
	a_2	(3,0, 0)	(0,5, 0)

Table 2.13B: (u_1, u_2, u_3) with $x_3 = c_2$

		x_2: b_1	x_2: b_2
x_1	a_1	(0,0, 14)	(0, 6,12)
	a_2	(-3.5, 0, 7.5)	(-0.5, 9, 11.5)

Table 2.13: Fair Bayesian economy in which $\mathcal{M}^*$ is not individually rational

I have established that, in general, the mechanism $\mathcal{M}^*$ is not strategy-proof or individually rational. Our next result presents a specific condition under which $\mathcal{M}^*$ achieves both strategy-proofness and individually rationality. It states that $\mathcal{M}^*$ is strategy-proof and individually rational in any fair Bayesian economy endowed with an ex-post weakly monotonic production function.

Theorem 10 : *Let* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, EFP, (u_i) \rangle$ *be a fair Bayesian economy. The mechanism* $\mathcal{M}^*$ *is strategy-proof and individually rational in the economy* $\mathcal{B}$ *if the production function* f *is ex-post weakly monotonic.*

The next example illustrates this finding.

Example 24 : *Consider the fair Bayesian economy* $\mathcal{B}$ *as defined in Example 11, with the production function and the utility levels of each worker in different states of the economy recalled below. We have seen that the production function of this economy is ex-post weakly monotonic (Example 17). As a result, Theorem 10 implies that the mechanism* $\mathcal{M}^*$ *is strategy-proof and individually rational in this economy. To see this, we first observe that* $\mathcal{M}^*(A, L) = (2, 1)$ *and* $\mathcal{M}^*(A, H) = (2, 2)$*. Thus, if the two workers participate in the mechanism* $\mathcal{M}^*$*, worker 1 would choose to work for 2 periods regardless of worker 2's ability. Whereas worker 2 would choose to work for one period if his ability is low and for two periods if his ability is high. If the type profile (A, L) is realized, then we have that* $u_1((A, L), (2, 1)) = 9 > u_1((A, L), (0, 1)) = 0$ *and* $u_2((A, L), (2, 1)) = 5 > u_2((A, L), (2, 0)) = 0$*. If the type profile (A, H) is realized, then we have that* $u_1((A, H), (2, 2)) = 12.5 > u_1((A, H), (0, 2)) = 0$ *and* $u_2((A, H), (2, 2)) = 17.5 > u_2((A, H), (2, 0)) = 0$*. Therefore, each worker is better-off participating in the mechanism* $\mathcal{M}^*$ *than being inactive, given workers' abilities. It follows that the mechanism* $\mathcal{M}^*$ *is individually rational. To see how* $\mathcal{M}^*$ *is strategy-proof, we first note that worker 1 has only one ability (the average ability), so he cannot engage in any strategic behavior. If he participates in the mechanism, he will choose to work for two periods, regardless of worker 2's ability. On the other hand, worker 2 has two possible ability levels. Suppose his ability is low. If he reports truthfully, his utility would be* $v_2(\mathcal{M}^*(A, L) = u_2((A, L), (2, 1)) = 5$*. In contrast, if he falsely claims to have high ability, his utility would be* $v_2(\mathcal{M}^*(A, H) = u_2((A, L), (2, 2)) = 3.5$*. Therefore, worker 2 has no incentive to misreport his ability if it is low. Similarly, it can be shown that if his ability is high, he would not gain from reporting it as low. Hence, the mechanism* $\mathcal{M}^*$ *is strategy-proof in the fair Bayesian economy* $\mathcal{B}$*.*

$$f_{(A,L)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 6 & if \quad x = (0,1) \\ 5 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 14 & if \quad x = (2,1) \\ 12 & if \quad x = (2,2) \end{cases} \quad \text{and } f_{(A,H)}(x) = \begin{cases} 0 & if \quad x = (0,0) \\ 10 & if \quad x = (0,1) \\ 15 & if \quad x = (0,2) \\ 10 & if \quad x = (2,0) \\ 25 & if \quad x = (2,1) \\ 30 & if \quad x = (2,2) \end{cases}$$

Table 2.14A: (u_1, u_2) at (A, L)

		x_2 = 0	1	2
x_1	0	(0, 0)	(0, 6)	(0, 5)
	2	(10, 0)	(9, 5)	(8.5, 3.5)

Table 2.14B: (u_1, u_2) at (A, H)

		x_2 = 0	1	2
x_1	0	(0, 0)	(0, 10)	(0, 15)
	2	(10, 0)	(12.5, 12.5)	(12.5, 17.5)

Table 2.14: Fair Bayesian economy in which $\mathcal{M}^*$ is strategy-proof and individually rational.

2.7 Conclusion

In this chapter, I incorporate fairness considerations into a classical model of a production economy with incomplete information. This has led to a new class of Bayesian economies. My model accommodates agent heterogeneity and varying action sets and broadens the scope of existing models, illustrating how fairness principles can be effectively applied ex-post. The analysis shows that these principles not only characterize a unique pay scheme–the ex-post fair pay–but also ensure the existence of a pure-strategy Bayesian Nash equilibrium. Furthermore, I uncover structural conditions–ex-post monotonicity– necessary for achieving a Pareto-efficient equilibrium, highlighting the balance between fairness, efficiency, and stability in Bayesian economies.

Beyond the basic model, the study explores several key extensions that enhance its relevance and applicability. First, it incorporates scenarios where agents incur costs based on their type, adding realism to the analysis by accounting for the diverse burdens borne by participants. Second, it addresses the redistribution of resources to support unproductive agents through taxation, integrating principles of fairness with considerations of inclusiveness and sustainability. These extensions culminate in a more comprehensive framework that reflects complex real-world conditions, illustrating the robustness and adaptability of the proposed methodology across varied economic environments.

Finally, I develop an application to mechanism design, examining whether the fairness principles can incentivize agents to participate optimally in production and act truthfully regarding their private information. By defining a mechanism $\mathcal{M}^*$ that selects action profiles to maximize the economy's output based on agents' types, I show that this mechanism is Pareto-efficient in fair Bayesian economies. Moreover, $\mathcal{M}^*$ satisfies individual rationality and strategy-proofness, provided the economy is fair, and the underlying technologies are weakly monotonic. This finding has implications for the design of mechanisms that not only ensure fair and efficient outcomes but also promote truthful reporting of private information—key to effective decision-making in economic systems.

Overall, this study advances the theoretical understanding of fairness in Bayesian economies while providing practical insights for the design of mechanisms that achieve fair, efficient, and stable outcomes. The findings have broad implications, suggesting that with appropriately defined fairness principles, it is possible to create economic systems that are both just and conducive to productive cooperation, even in settings where information is incomplete, and agents' incentives are diverse.

2.8 Appendix A. Proofs from Section 2.3

Proof (Theorem 4) : ***Sufficiency***

In this part of the proof, I show that the ex-post fair pay EFP satisfies EI, EE, and EBF. Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, EFP, (u_i)\rangle$ be a Myerson economy.

Ex-post inactivity: *Let $\theta \in \Theta$ and $i \in N$. It is immediate that $EFP_i(f_\theta, o) = 0$. Therefore, the ex-post fair pay satisfies EI.*

Ex-post Efficiency: *Let $\theta \in \Theta$, $x \in X$ and $i \in N$. Without loss of generality, let us assume that all agents are active at x, that is, $x_i \neq o_i$ for all $i \in N$.*

$$\begin{aligned} EFP_i(f_\theta, x) &= \sum_{a \triangleleft_o^i x} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!}\left[f_\theta(x_i, a_{-i}) - f_\theta(a)\right] \\ &= \sum_{a \trianglelefteq x,\, a_i = o_i} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!}\left[f_\theta(x_i, a_{-i}) - f_\theta(a)\right] \\ &= \sum_{a \trianglelefteq x,\, a_i = o_i} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(x_i, a_{-i}) - \sum_{a \trianglelefteq x,\, a_i = o_i} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(a) \end{aligned}$$

For all $a \trianglelefteq x$, let $A(a,x) = \{i \in N : a_i = x_i\}$ and $I(a,x) = \{i \in N : a_i = o_i\}$. We have that $N = A(a,x) \cup I(a,x)$, $|A(a,x)| = |a|$ and $|I(a,x)| = |x| - |a|$. Also, for all $a \triangleleft_o^i x$, let $b = (x_i, a_{-i})$. We have $b \trianglelefteq x$ and $|b| = |a| + 1$. It follows that:

$$\begin{aligned}
\sum_{i\in N} EFP_i(f_\theta,x) &= \sum_{i\in N}\sum_{b\unlhd x,\, b_i=x_i} \frac{(|b|-1)!(|x|-|b|)!}{(|x|)!} f_\theta(b) - \sum_{i\in N}\sum_{a\unlhd x,\, a_i=o_i} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(a)\\
&= \sum_{i\in N}\sum_{b\unlhd x,\, i\in A(b,x)} \frac{(|b|-1)!(|x|-|b|)!}{(|x|)!} f_\theta(b) - \sum_{i\in N}\sum_{a\unlhd x,\, i\in I(a,x)} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(a)\\
&= \sum_{b\unlhd x}\sum_{i\in A(b,x)} \frac{(|b|-1)!(|x|-|b|)!}{(|x|)!} f_\theta(b) - \sum_{a\lhd x}\sum_{i\in I(a,x)} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(a)\\
&= \sum_{b\unlhd x} \frac{(|b|)(|b|-1)!(|x|-|b|)!}{(|x|)!} f_\theta(b) - \sum_{a\lhd x} \frac{(|x|-|a|)(|a|)!(|x|-|a|-1)!}{(|x|)!} f_\theta(a)\\
&= \sum_{b\unlhd x} \frac{(|b|)!(|x|-|b|)!}{(|x|)!} f_\theta(b) - \sum_{a\lhd x} \frac{(|a|)!(|x|-|a|)!}{(|x|)!} f_\theta(a)\\
&= \frac{(|x|)!(|x|-|x|)!}{(|x|)!} f_\theta(x)\\
&= f_\theta(x).
\end{aligned}$$

Hence, the ex-post fair pay satisfies EE.

Ex-post Bilateral Fairness: *Let $\theta \in \Theta$, $x \in X$ and $i,j \in N$. Let us show that $EFP_i(f_\theta,x) - EFP_i(f_\theta,(o_j,x_{-j})) = EFP_j(f_\theta,x) - EFP_j(f_\theta,(o_i,x_{-i}))$ for all $x \in X$. That is:*

$$EFP_i(f_\theta,x) - EFP_j(f_\theta,x) = EFP_i(f_\theta,(o_i,x_{-j})) - EFP_j(f_\theta,(o_i,x_{-i})) \text{ for all } x \in X. \tag{2.9}$$

Let $x \in X$. If $x_i = o_i$ or $x_j = o_i$, then it follows that $EFP_i(f_\theta,x) - EFP_j(f_\theta,x) = 0 = EFP_i(f_\theta,(o_j,x_{-j})) - EFP_j(f_\theta,(o_i,x_{-i}))$. Now assume workers i and j are active at x, that is $x_i \neq o_i$ and $x_j \neq o_j$. It follows that:

$$\begin{aligned}
EFP_i(f_\theta,x) &= \sum_{a\lhd_o^i x} \varphi(a,x)\,[f_\theta(a+x_ie_i) - f_\theta(a)]\\
&= \sum_{\substack{a\lhd_o^i x\\ a_j=o_j}} \varphi(a,x)\,[f_\theta(a+x_ie_i) - f_\theta(a)] + \sum_{\substack{a\lhd_o^i x\\ a_j\neq o_j}} \varphi(a,x)\,[f_\theta(a+x_ie_i) - f_\theta(a)]\\
&= \sum_{\substack{a\lhd_o^i x\\ a_j=o_j}} \varphi(a,x)\,[f_\theta(a+x_ie_i) - f_\theta(a)] + \sum_{\substack{a\lhd_o^i x\\ a_j=o_j}} \varphi(a+x_je_j,x)\,[f_\theta(a+x_je_j+x_ie_i) - f_\theta(a+x_je_j)]
\end{aligned}$$

The last equality follows from the fact that for all $a,x \in X$ such that $a \lhd_o^i x$, $a_j \neq o_j$ implies that $a_j = x_j$.

Similarly, we have that:

$$EFP_j(f_\theta,x) = \sum_{\substack{a\lhd_o^i x\\ a_j=o_j}} \varphi(a,x)\,[f_\theta(a+x_je_j) - f_\theta(a)] + \sum_{\substack{a\lhd_o^i x\\ a_j=o_j}} \varphi(a+x_ie_i,x)\,[f_\theta(a+x_ie_i+x_je_j) - f_\theta(a+x_ie_i)]$$

But for all $a \in X$ such that $a \lhd_o^i x$ and $a_j = o_j$, we have that

$$\varphi(a + x_i e_i, x) = \frac{(|a| + 1)!(|x| - |a| - 2)!}{(|x|)!} = \varphi(a + x_j e_j, x).$$

This implies that:

$$\begin{aligned} EFP_i(f_\theta, x) - EFP_j(f_\theta, x) &= \sum_{\substack{a \lhd_o^i x \\ a_j = o}} \varphi(a, x) \left[f_\theta(a + x_i e_i) - f_\theta(a + x_j e_j) \right] \\ &+ \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \varphi(a + x_j e_j, x) \left[f_\theta(a + x_i e_i) - f_\theta(a + x_j e_j) \right] \\ &= \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \left(\varphi(a, x) + \varphi(a + x_j e_j, x) \right) \left[f_\theta(a + x_i e_i) - f_\theta(a + x_j e_j) \right] \end{aligned}$$

But for all $a \in X$ such that $a \lhd_o^i x$ and $a_j = o_j$,

$$\varphi(a, x) + \varphi(a + x_j e_j, x) = \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} + \frac{(|a| + 1)!(|x| - |a| - 2)!}{(|x|)!} = \frac{(|a|)!(|x| - |a| - 2)!}{(|x| - 1)!}$$

It follows that:

$$EFP_i(f_\theta, x) - EFP_j(f_\theta, x) = \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \frac{(|a|)!(|x| - |a| - 2)!}{(|x| - 1)!} \left[f_\theta(a + x_i e_i) - f_\theta(a + x_j e_j) \right] \tag{2.10}$$

On the other hand, we have that:

$$\begin{aligned} EFP_i(f_\theta, (o_j, x_{-j})) &= \sum_{a \lhd_o^i (o_j, x_{-j})} \varphi(a, (o_j, x_{-j})) \left[f_\theta(a + x_i e_i) - f_\theta(a) \right] \\ &= \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \varphi(a, (o_j, x_{-j})) \left[f_\theta(a + x_i e_i) - f_\theta(a) \right] \end{aligned}$$

Similarly, we have that:

$$EFP_j(f_\theta, (o_i, x_{-i})) = \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \varphi(a, (o_i, x_{-i})) \left[f_\theta(a + x_j e_j) - f_\theta(a) \right]$$

But for all $a \in X$ such that $a \lhd_o^i x$ and $a_j = o_j$, we have that:

$$\varphi(a, (o_j, x_{-j})) = \frac{(|a|)!(|x| - |a| - 2)!}{(|x| - 1)!} = \varphi(a, (o_i, x_{-i})).$$

It follows that:

$$EFP_i(f_\theta,(o_j,x_{-j})) - EFP_j(f_\theta,(o_i,x_{-i})) = \sum_{\substack{a \lhd_o^i x \\ a_j = o_j}} \frac{(|a|)!(|x|-|a|-2)!}{(|x|-1)!} \left[f_\theta(a+x_ie_i) - f_\theta(a+x_je_j)\right] \tag{2.11}$$

From Equations 2.10 and 2.11, it follows that Equation 2.9 holds. Hence, the ex-post fair pay satisfies the EBF.

Necessity: *In this part of the proof, I prove the uniqueness of the ex-post fair pay.* $\mathcal{B} = \langle N,(\Theta_i),(X_i),(o_i),(p_i),f,\Phi,(u_i)\rangle$ *be a Myerson economy. Let* $\theta \in \Theta$ *and* $x \in X$*. I want to show that* $\Phi_i(f_\theta,x) = EFP_i(f_\theta,x)$ *for all agents* $i \in N$*. I proceed by induction on the number* $|x|$ *of active agents at* x*. If* $|x| = 0$*, then all agents are inactive at* x*. By EI, it follows that* $\Phi_i(f_\theta,x) = 0 = EFP_i(f_\theta,x)$ *for all* $i \in N$*. So, the proof is verified for* $|x| = 0$*. Now, assume that the two pay schemes are identical for all action profiles with less than* n *active agents and let* x *be an action profile with* n *active agents. By induction, we have that* $EFP_i(f_\theta,(o_j,x_{-j})) = \phi_i(f_\theta,(o_j,x_{-j}))$ *for any* $i,j \in N$*. As a result, for all* $i,j \in N$*, EBF implies that:*

$$\begin{aligned} EFP_i(f_\theta,x) - EFP_j(f_\theta,x) &= EFP_i(f_\theta,(o_j,x_{-j})) - EFP_j(f_\theta,(o_i,x_{-i})) \\ &= \phi_i(f_\theta,x) - \phi_j(f_\theta,(o_i,x_{-i})) \\ &= \phi_i(f_\theta,x) - \phi_j(f_\theta,x) \end{aligned}$$

That is:

$$EFP_i(f_\theta,x) - EFP_j(f_\theta,x) = \phi_i(f_\theta,x) - \phi_j(f_\theta,x) \text{ for all } i,j \in N.$$

Fixing i *and summing over* $j \in N$ *yields:*

$$EFP_i(f_\theta,x) - \sum_{j\in N} EFP_j(f_\theta,x) = \phi_i(f_\theta,x) - \sum_{j\in N} \phi_j(f_\theta,x) \text{ for all } i \in N.$$

EE applies and we get:

$$EFP_i(f_\theta,x) - f_\theta(x) = \phi_i(f_\theta,x) - f_\theta(x) \text{ for all } i \in N.$$

It follows that:

$$EFP_i(f_\theta,x) = \phi_i(f_\theta,x) \text{ for all } i \in N.$$

Hence, the ex-post fair pay, EFP*, is the unique pay scheme of a Myerson economy.*

In addition, let us show that the ex-post fair pay satisfies EU, EA, and EM.

Ex-post Unproductivity: *Let* $\theta \in \Theta$, $x \in X$ *and* $i \in N$ *an unproductive agent in the state of the economy* θ. *We have that* $mc(i, f\theta, a, x) = 0$ *for all* $a \lhd_o^i x$. *It follows that* $EFP_i(f_\theta, x) = \sum_{a \lhd_o^i x} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} mc(i, f_\theta, a, x) = 0$. *Hence, the ex-post fair pay satisfies EU.*

Ex-post Anonymity. *Let* $\theta \in \Theta$, $x \in X$, $\pi \in \mathcal{S}_n$, *and* $i \in N$.

$$EFP_{\pi(i)}(\pi f_\theta, \pi x) = \sum_{a \lhd_o^{\pi(i)} \pi x} \frac{(|a|)!(|\pi x| - |a| - 1)!}{(|\pi x|)!} [\pi f_\theta(x_{\pi(i)}, a_{-\pi(i)}) - \pi f_\theta(a)]$$

For all $a \lhd_o^{\pi(i)} \pi x$, *there exists a unique 0* $b \lhd_o^i x$ *such that* $a = \pi b$. *Moreover, we have* $|a| = |b|$ *and* $|x| = |\pi(x)|$. *It follows that*

$$\begin{aligned} EFP_{\pi(i)}(\pi f_\theta, \pi x) &= \sum_{b \lhd_o^i x} \frac{(|b|)!(|x| - |b| - 1)!}{(|x|)!} [\pi f_\theta(\pi(x_i, b_{-i})) - \pi f_\theta(\pi b)] \\ &= \sum_{b \lhd_o^i x} \frac{(|b|)!(|x| - |b| - 1)!}{(|x|)!} [f_\theta(x_i, b_{-i}) - f_\theta(b)] \\ &= EFP_i(f, x). \end{aligned}$$

Hence, the ex-post fair pay satisfies EA.

(EM) Ex-post Marginality. *Let* $\theta, \theta' \in \Theta$, $x \in X$, *and* $i \in N$ *such that* $mc(i, f_\theta, a, x) \geq mc(i, f_{\theta'}, a, x)$ *for all* $a \lhd_o^i x$. *By definition of the ex-post fair pay, it is immediate that* $\Phi_i(f_\theta, x) \geq \Phi_i(f_{\theta'}, x)$. *Hence, the ex-post fair pay satisfies EM.* ∎

2.9 Appendix B. Proofs from Section 2.4

2.9.1 Proof of Theorem 5

The proof of Theorem 5 relies on the theory of potential games (Monderer and Shapley, 1996) and the type-agent representation of Bayesian games (Harsanyi, 1967; Harsanyi, 1968).

The type-agent representation is a convenient way to transform any Bayesian game into a strategic form game. The resulting strategic form game is called type-agent representation (Myerson, 1997) or the Selten game (Harsanyi, 1967; Harsanyi, 1968). In the type-agent representation, there is one player or agent for every possible type of every player in the given Bayesian game. By relabeling types if necessary, we may assume without loss of generality that the sets Θ_i are disjoint, so that $\Theta_i \cap \Theta_j = \varnothing$ if $i \neq j$. Given a Bayesian game

$\Gamma = \langle N, (\Theta_i), (X_i), (p_i), (u_i)\rangle$, the Selten game is an equivalent strategic form game

$$\Gamma^s = \langle N^s, (X_{\theta_i})_{\substack{\theta_i \in \Theta_i \\ i \in N}}, (U_{\theta_i})_{\substack{\theta_i \in \Theta_i \\ i \in N}} \rangle$$

where $N^s = \bigcup_{i \in N} \Theta_i$, $X_{\theta_i} = X_i$ for every $\theta_i \in \Theta_i$ and $i \in N$, for every $x = (x_{\theta_i}) \in \prod_{\theta_i \in N^s} X_{\theta_i}$ and $\theta_i \in N^s$

$$U_{\theta_i}(x) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\; u_i(\theta_i, t_{-i}, x_{\theta_i}, x_{t_{-i}})$$

It follows that $(s_1^*, s_2^*, ..., s_n^*)$ is a pure strategy Bayesian Nash equilibrium of the Bayesian game Γ if and only if $(s_i^*(\theta_i))_{\substack{\theta_i \in \Theta_i \\ i \in N}}$ is a pure strategy Nash equilibrium of the Selten game Γ^s. Therefore, the existence of a Bayesian Nash equilibrium of a Bayesian game is guaranteed by the existence of a Nash equilibrium of the corresponding Selten game, and vice versa.

In order to state lemmas that prepare the proof of Theorem 5, I need to introduce some definitions.

Definition 17 : *Let $h = (h_1, h_2, ..., h_n) : \mathbb{R} \longrightarrow \mathbb{R}^n$ be a function. h is strictly monotonic if $\forall i \in \{1, 2, ..., n\}$, $\forall x, y \in \mathbb{R}$, $x < y \Leftrightarrow h_i(x) < h_i(y)$.*

Definition 18 : *Let $\mathcal{G} = (N, (X_i)_{i\in N}, (u_i)_{i \in N})$ be a strategic form game and $h = (h_1, h_2, ..., h_n) : \mathbb{R} \longrightarrow \mathbb{R}^n$ a function. The transform of $\mathcal{G}$ by h is naturally the strategic form game denoted by $\mathcal{G}_h = (N, (X_i)_{i\in N}, (h_i \circ u_i)_{i\in N})$ with $h_i \circ u_i(x) = h_i(u_i(x))$ $\forall x \in X = \prod_i^n X_i$, and $i \in N$. If h is strictly monotonic, then $\mathcal{G}_h$ will be referred to as the strictly monotonic transform of $\mathcal{G}$ by h.*

The following lemma establishes a relation between the Nash equilibrium of a game and that of its strictly monotonic transforms.

Lemma 1 : *Let $\mathcal{G} = (N, (X_i)_{i\in N}, (u_i)_{i\in N})$ be a strategic form game and $h = (h_1, h_2, ..., h_n) : \mathbb{R} \longrightarrow \mathbb{R}^n$ a strictly monotonic function. $\mathcal{G}$ admits a Nash equilibrium if and only if $\mathcal{G}_h$ admits a Nash equilibrium. Moreover, the sets of Nash equilibria of both games coincide.*

The proof follows from the definitions and is therefore omitted.

The next lemma provides a sufficient condition for the existence of a Nash equilibrium of a strategic form game. It is derived from corollaries 2.2 and 2.9 of the paper by Monderer and Shapley (1996).

Lemma 2 : *A finite strategic form game $\mathcal{G} = (N, (X_i)_{i\in N}, (u_i)_{i\in N})$ admits a pure Nash equilibrium if for every $i, j \in N$, for every $a = a_{-ij} \in \prod_{k \in N \setminus \{i,j\}} X_k$, and for every $x_i, y_i \in X_i$ and*

$x_j, y_j \in X_j$,

$$u_i(B) - u_i(A) + u_j(C) - u_j(B) + u_i(D) - u_i(C) + u_j(A) - u_j(D) = 0 \tag{2.12}$$

where $A = (x_i, x_j, a)$, $B = (y_i, x_j, a)$, $C = (y_i, y_j, a)$, *and* $D = (x_i, y_j, a)$.

Proof (Theorem 5) : *Let* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ *be a fair Bayesian economy. Let* $\Gamma^{\mathcal{B}}$ *be the Bayesian game associated to* $\mathcal{B}$ *and* $\Gamma^{\mathcal{B}s}$ *its corresponding Selten game. Let* $\mathbb{P}$ *be the prior probability distribution of the Bayesian economy. Assume* $\Theta_i = \{\theta_{i1}, \theta_{i2}, ..., \theta_{ik_i}\}$ *for any* $i \in N$. *For any* $i \in N$ *and* $k \in \{1, 2, ..., k_i\}$, *let* $\beta_{\theta_{ik}} = \sum_{\theta_{-i}\in\Theta_{-i}} \mathbb{P}(\theta_{ik}, \theta_{-i})$. *Consider the strictly monotonic application* $h = (h_{\theta_{ik}}) : \mathbb{R} \to \mathbb{R}^{N^s}$ *defined for all* $i \in N$, $k \in \{1, 2, ..., k_i\}$ *and* $x \in \mathbb{R}$ *by* $h_{\theta_{ik}}(x) = \beta_{\theta_{ik}} x$. *Let* $\Gamma_h^{\mathcal{B}s}$ *be the strictly monotonic transform of* $\Gamma^{\mathcal{B}s}$ *by* h.

From Lemmas 1 and 2, it suffices to show that Equation 2.12 of Lemma 2 is satisfied for the strategic form game $\Gamma_h^{\mathcal{B}s}$. *Let* $\theta_{ik}, \theta_{jm} \in N^s$, *with* $i, j \in N$, $k \in \{1, 2, ..., k_i\}$ *and* $m \in \{1, 2, ..., k_j\}$. *Let* $x_{\theta_{ik}}, y_{\theta_{ik}} \in X_{\theta ik} = X_i$, $x_{\theta_{jm}}, y_{\theta_{jm}} \in X_{\theta_{jm}} = X_j$ *and* $a \in \prod_{\theta_{lt}\in N^s\setminus\{\theta_{ik},\theta_{jm}\}} X_{\theta_{lt}}$. *Let* A, B, C *and* D *be strategies of the strategic form game* $\Gamma_h^{\mathcal{B}s}$ *as defined in Lemma 2 with* i *and* j *being replaced by* θ_{ik} *and* θ_{jm} *respectively. I want to show that the following equation is satisfied*

$$\begin{aligned} &h_{\theta_{ik}}(U_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U_{\theta_{ik}}(A)) + h_{\theta_{jm}}(U_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U_{\theta_{jm}}(B)) + \\ &h_{\theta_{ik}}(U_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U_{\theta_{ik}}(C)) + h_{\theta_{jm}}(U_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U_{\theta_{jm}}(D)) = 0 \end{aligned} \tag{2.13}$$

Let $K_1 = h_{\theta_{ik}}(U_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U_{\theta_{ik}}(A))$, $K_2 = h_{\theta_{jm}}(U_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U_{\theta_{jm}}(B))$,
$K_3 = h_{\theta_{ik}}(U_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U_{\theta_{ik}}(C))$ *and* $K_4 = h_{\theta_{jm}}(U_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U_{\theta_{jm}}(D))$. *The proof can be broken down into two cases.*

Case 1: $i = j$, *then* $\theta_{ik}, \theta_{im} \in \Theta_i$. *We have:*

$$\begin{aligned} K_1 &= h_{\theta_{ik}}(U_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U_{\theta_{ik}}(A)) \\ &= \beta_{\theta_{ik}} U_{\theta_{ik}}(y_{\theta_{ik}}, x_{\theta_{im}}, a) - \beta_{\theta_{ik}} U_{\theta_{ik}}(x_{\theta_{ik}}, x_{\theta_{im}}, a) \\ &= \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\, u_i(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\, u_i(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \\ &= \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik},t_{-i})}, (y_{\theta_{ik}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{ik},t_{-i})}, (x_{\theta_{ik}}, a_{t_{-i}}))\right) \end{aligned}$$

$$\begin{aligned} K_3 &= h_{\theta_{ik}}(U_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U_{\theta_{ik}}(C)) \\ &= \beta_{\theta_{ik}} U_{\theta_{ik}}(x_{\theta_{ik}}, y_{\theta_{im}}, a) - \beta_{\theta_{ik}} U_{\theta_{ik}}(y_{\theta_{ik}}, y_{\theta_{im}}, a) \\ &= \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik},t_{-i})}, (x_{\theta_{ik}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{ik},t_{-i})}, (y_{\theta_{ik}}, a_{t_{-i}}))\right) \end{aligned}$$

It follows that $K_1 + K_3 = 0$. *Similarly, we have*

$$\begin{aligned} K_2 &= h_{\theta_{im}}(U_{\theta_{im}}(C)) - h_{\theta_{im}}(U_{\theta_{im}}(B)) \\ &= \beta_{\theta_{im}} U_{\theta_{im}}(y_{\theta_{ik}}, y_{\theta_{im}}, a) - \beta_{\theta_{im}} U_{\theta_{im}}(y_{\theta_{ik}}, x_{\theta_{im}}, a) \\ &= \beta_{\theta_{im}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{im}) \left(EFP_i(f_{(\theta_{im}, t_{-i})}, (y_{\theta_{im}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{im}, t_{-i})}, (x_{\theta_{im}}, a_{t_{-i}})) \right) \end{aligned}$$

$$\begin{aligned} K_4 &= h_{\theta_{im}}(U_{\theta_{im}}(A)) - h_{\theta_{im}}(U_{\theta_{im}}(D)) \\ &= \beta_{\theta_{im}} U_{\theta_{im}}(x_{\theta_{ik}}, x_{\theta_{im}}, a) - \beta_{\theta_{im}} U_{\theta_{im}}(x_{\theta_{ik}}, y_{\theta_{im}}, a) \\ &= \beta_{\theta_{im}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{im}) \left(EFP_i(f_{(\theta_{im}, t_{-i})}, (x_{\theta_{im}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{im}, t_{-i})}, (y_{\theta_{im}}, a_{t_{-i}})) \right) \end{aligned}$$

It follows that $K_2 + K_4 = 0$.

Hence, $K_1 + K_2 + K_3 + K_4 = 0$ *if* $i = j$. *That is, Equation* (2.13) *is satisfied if the two players of the strictly monotonic transform of the Selten game* $\Gamma^{\mathcal{B}s}$ *by* h *are drawn from the same type set. Let us see what happens when they are drawn from different type sets.*

Case 2: $i \neq j$, *then* $\theta_{ik} \in \Theta_i$ *and* $\theta_{jm} \in \Theta_j$. *We have:*

$$\begin{aligned} K_1 &= h_{\theta_{ik}}(U_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U_{\theta_{ik}}(A)) \\ &= \beta_{\theta_{ik}} U_{\theta_{ik}}(y_{\theta_{ik}}, x_{\theta_{jm}}, a) - \beta_{\theta_{ik}} U_{\theta_{ik}}(x_{\theta_{ik}}, x_{\theta_{jm}}, a) \\ &= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\, u_i(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\, u_i(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \\ &= \beta_{\theta_{ik}} \sum_{(\theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left(u_i(\theta_{ik}, \theta_{jm}, t_{-ij}, y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - u_i(\theta_{ik}, \theta_{jm}, t_{-ij}, x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \right) \\ &+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left(u_i(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - u_i(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \right) \\ &= \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}, (y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}})) - EFP_i(f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}, (x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}})) \right) \\ &+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik}, t_{-i})}, (y_{\theta_{ik}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{ik}, t_{-i})}, (x_{\theta_{ik}}, a_{t_{-i}})) \right) \\ &= K_{11} + K_{12}, \text{ where} \end{aligned}$$

$$K_{11} = \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}, (y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}})) - EFP_i(f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}, (x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}})) \right)$$

$$\text{and } K_{12} = \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left(EFP_i(f_{(\theta_{ik}, t_{-i})}, (y_{\theta_{ik}}, a_{t_{-i}})) - EFP_i(f_{(\theta_{ik}, t_{-i})}, (x_{\theta_{ik}}, a_{t_{-i}})) \right).$$

Viewing all players but players i and j as one player and setting $f_{(\theta_{ik},\theta_{jm},t_{-ij})} = f$ yields:

$$\begin{aligned}
K_{11} &= \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \Big[\frac{1}{3}\big(f(y_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(0,x_{\theta_{jm}},a_{t_{-ij}}) + f(y_{\theta_{ik}},0,0) - f(0,0,0)\big) \\
&+ \frac{1}{6}\big(f(y_{\theta_{ik}},x_{\theta_{jm}},0) - f(0,x_{\theta_{jm}},0) + f(y_{\theta_{ik}},0,a_{t_{-ij}}) - f(0,0,a_{t_{-ij}})\big) \\
&- \frac{1}{3}\big(f(x_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(0,x_{\theta_{jm}},a_{t_{-ij}}) + f(x_{\theta_{ik}},0,0) - f(0,0,0)\big) \\
&- \frac{1}{6}\big(f(x_{\theta_{ik}},x_{\theta_{jm}},0) - f(0,x_{\theta_{jm}},0) + f(x_{\theta_{ik}},0,a_{t_{-ij}}) - f(0,0,a_{t_{-ij}})\big)\Big] \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \big[\tfrac{1}{3}\big(f(y_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) + f(y_{\theta_{ik}},0,0) - f(x_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(x_{\theta_{ik}},0,0)\big) \\
&+ \frac{1}{6}\big(f(y_{\theta_{ik}},x_{\theta_{jm}},0) + f(y_{\theta_{ik}},0,a_{t_{-ij}}) - f(x_{\theta_{ik}},x_{\theta_{jm}},0) - f(x_{\theta_{ik}},0,a_{t_{-ij}})\big)\Big]
\end{aligned}$$

$$\begin{aligned}
K_3 &= h_{\theta_{ik}}(U_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U_{\theta_{ik}}(C)) \\
&= \beta_{\theta_{ik}} U_{\theta_{ik}}(x_{\theta_{ik}},y_{\theta_{jm}},a) - \beta_{\theta_{ik}} U_{\theta_{ik}}(y_{\theta_{ik}},y_{\theta_{jm}},a) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i(\theta_{ik},t_{-i},x_{\theta_{ik}},a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i}\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i(\theta_{ik},t_{-i},y_{\theta_{ik}},a_{t_{-i}}) \\
&= \beta_{\theta_{ik}} \sum_{(t_j=\theta_{jm},t_{-ij})\in\Theta_{-i}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \big(u_i(\theta_{ik},\theta_{jm},t_{-ij},x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - u_i(\theta_{ik},\theta_{jm},t_{-ij},y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})\big) \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i}=(t_j\neq\theta_{jm},t_{-ij})\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \big(u_i(\theta_{ik},t_{-i},x_{\theta_{ik}},a_{t_{-i}}) - u_i(\theta_{ik},t_{-i},y_{\theta_{ik}},a_{t_{-i}})\big) \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \big(EFP_i(f_{(\theta_{ik},\theta_{jm},t_{-ij})},(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})) - EFP_i(f_{(\theta_{ik},\theta_{jm},t_{-ij})},(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}))\big) \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i}=(t_j\neq\theta_{jm},t_{-ij})\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \big(EFP_i(f_{(\theta_{ik},t_{-i})},(x_{\theta_{ik}},a_{t_{-i}})) - EFP_i(f_{(\theta_{ik},t_{-i})},(y_{\theta_{ik}},a_{t_{-i}}))\big) \\
&= K_{31} + K_{32},\ \textit{where}
\end{aligned}$$

$$K_{31} = \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \big(EFP_i(f_{(\theta_{ik},\theta_{jm},t_{-ij})},(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})) - EFP_i(f_{(\theta_{ik},\theta_{jm},t_{-ij})},(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}))\big)$$

$$\textit{and } K_{32} = \beta_{\theta_{ik}} \sum_{t_{-i}=(t_j\neq\theta_{jm},t_{-ij})\in\Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \big(EFP_i(f_{(\theta_{ik},t_{-i})},(x_{\theta_{ik}},a_{t_{-i}})) - EFP_i(f_{(\theta_{ik},t_{-i})},(y_{\theta_{ik}},a_{t_{-i}}))\big).$$

Again, viewing all players but players i and j as one player and setting $f_{(\theta_{ik},\theta_{jm},t_{-ij})} = f$ yields:

$$\begin{aligned}
K_{31} &= \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \Big[\frac{1}{3}\big(f(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(0,y_{\theta_{jm}},a_{t_{-ij}}) + f(x_{\theta_{ik}},0,0) - f(0,0,0)\big) \\
&+ \frac{1}{6}\big(f(x_{\theta_{ik}},y_{\theta_{jm}},0) - f(0,y_{\theta_{jm}},0) + f(x_{\theta_{ik}},0,a_{t_{-ij}}) - f(0,0,a_{t_{-ij}})\big) \\
&- \frac{1}{3}\big(f(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(0,y_{\theta_{jm}},a_{t_{-ij}}) + f(y_{\theta_{ik}},0,0) - f(0,0,0)\big) \\
&- \frac{1}{6}\big(f(y_{\theta_{ik}},y_{\theta_{jm}},0) - f(0,y_{\theta_{jm}},0) + f(y_{\theta_{ik}},0,a_{t_{-ij}}) - f(0,0,a_{t_{-ij}})\big)\Big] \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij}\in\Theta_{-ij}} p_i((\theta_{jm},t_{-ij})|\theta_{ik}) \big[\tfrac{1}{3}\big(f(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) + f(x_{\theta_{ik}},0,0) - f(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(y_{\theta_{ik}},0,0)\big) \\
&+ \frac{1}{6}\big(f(x_{\theta_{ik}},y_{\theta_{jm}},0) + f(x_{\theta_{ik}},0,a_{t_{-ij}}) - f(y_{\theta_{ik}},y_{\theta_{jm}},0) - f(y_{\theta_{ik}},0,a_{t_{-ij}})\big)\Big]
\end{aligned}$$

From case 1 we have that $K_{12} + K_{32} = 0$. It follows that:

$$\begin{aligned}
K_1 + K_3 &= K_{11} + K_{31} \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \Big[\frac{1}{3}\big(f(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - f(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) + f(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \\
&- f(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}})\big) + \frac{1}{6}\big(f(y_{\theta_{ik}}, x_{\theta_{jm}}, 0) - f(x_{\theta_{ik}}, x_{\theta_{jm}}, 0) + f(x_{\theta_{ik}}, y_{\theta_{jm}}, 0) - f(y_{\theta_{ik}}, y_{\theta_{jm}}, 0)\big)\Big]
\end{aligned}$$

But $\beta_{\theta_{ik}} = \sum_{\theta_{-i} \in \Theta_{-i}} \mathbb{P}(\theta_{ik}, \theta_{-i})$ and $p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) = \frac{\mathbb{P}(\theta_{ik}, \theta_{jm}, t_{-ij})}{\sum_{\theta_{-i} \in \Theta_{-i}} \mathbb{P}(\theta_{ik}, \theta_{-i})}$. This implies

$$\begin{aligned}
K_1 + K_3 &= \sum_{t_{-ij} \in \Theta_{-ij}} \mathbb{P}(\theta_{ik},\, \theta_{jm},\, t_{-ij}) \Big[\frac{1}{3}\big(f(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - f(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) + f(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \\
&- f(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}})\big) + \frac{1}{6}\big(f(y_{\theta_{ik}}, x_{\theta_{jm}}, 0) - f(x_{\theta_{ik}}, x_{\theta_{jm}}, 0) + f(x_{\theta_{ik}}, y_{\theta_{jm}}, 0) - f(y_{\theta_{ik}}, y_{\theta_{jm}}, 0)\big)\Big]
\end{aligned}$$

Similarly, we show that:

$$\begin{aligned}
K_2 + K_4 &= \beta_{\theta_{jm}} \sum_{t_{-ij} \in \Theta_{-ij}} p_j((\theta_{ik}, t_{-ij})|\theta_{jm}) \Big[\frac{1}{3}\big(f(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - f(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) + f(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \\
&- f(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}})\big) + \frac{1}{6}\big(f(y_{\theta_{ik}}, y_{\theta_{jm}}, 0) - f(y_{\theta_{ik}}, x_{\theta_{jm}}, 0) + f(x_{\theta_{ik}}, x_{\theta_{jm}}, 0) - f(x_{\theta_{ik}}, y_{\theta_{jm}}, 0)\big)\Big]
\end{aligned}$$

But $\beta_{\theta_{jm}} = \sum_{\theta_{-j} \in \Theta_{-j}} \mathbb{P}(\theta_{jm}, \theta_{-j})$ and $p_j((\theta_{ik}, t_{-ij})|\theta_{jm}) = \frac{\mathbb{P}(\theta_{ik}, \theta_{jm}, t_{-ij})}{\sum_{\theta_{-j} \in \Theta_{-j}} \mathbb{P}(\theta_{jm}, \theta_{-j})}$. This implies

$$\begin{aligned}
K_2 + K_4 &= \sum_{t_{-ij} \in \Theta_{-ij}} \mathbb{P}(\theta_{ik},\, \theta_{jm},\, t_{-ij}) \Big[\frac{1}{3}\big(f(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - f(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) + f(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \\
&- f(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}})\big) + \frac{1}{6}\big(f(y_{\theta_{ik}}, y_{\theta_{jm}}, 0) - f(y_{\theta_{ik}}, x_{\theta_{jm}}, 0) + f(x_{\theta_{ik}}, x_{\theta_{jm}}, 0) - f(x_{\theta_{ik}}, y_{\theta_{jm}}, 0)\big)\Big]
\end{aligned}$$

It follows that $K_1 + K_2 + K_3 + K_4 = 0$. So, equation 2.13 is satisfied. Hence, the Bayesian game $\Gamma^{\mathcal{B}}$ associated to the Bayesian economy $\mathcal{B}$ admits a pure strategy Bayesian Nash equilibrium. ■

Proof (Theorem 6) : *Let $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, EFP, (u_i)\rangle$ be an ex-post weakly monotonic fair Bayesian economy with agents' beliefs drawn from uniform distributions. Let $\Gamma^{\mathcal{B}} = \langle N, (\Theta_i), (X_i), (p_i), (u_i)\rangle$ be its associated Bayesian game. By definition, for all $i \in N$ and $\theta_i \in \Theta_i$, there exists a weak order $\succeq_{\theta_i}$ over X_i. For all $i \in N$ and $\theta_i \in \Theta_i$, let $\overline{x}_{\theta_i} \in X_i$ such that $\overline{x}_{\theta_i} \succeq_{\theta_i} x_i$ for all $x_i \in X_i$. Such an action exists since X_i is finite and $\succeq_{\theta_i}$ is complete and transitive. Consider the strategy profile $\overline{s} = (\overline{s}_1, \overline{s}_2, \ldots, \overline{s}_n)$, where $\overline{s}_i(\theta_i) = \overline{x}_{\theta_i}$ for all $\theta_i \in \Theta_i$. First, let us show that $\overline{s}$ is a pure strategy Bayesian Nash equilibrium of the game $\Gamma^{\mathcal{B}}$. Assume that an agent i deviates from $\overline{s} = (\overline{s}_i, \overline{s}_{-i})$ to $s =$*

$(s_i, \overline{s}_{-i})$. I want to show that: $U_i((\overline{s}_i, \overline{s}_{-i})|\theta_i) \geq U_i((s_i, \overline{s}_{-i})|\theta_i)$ for all $\theta_i \in \Theta_i$. Let $\theta_i \in \Theta_i$. By definition, we have that $U_i((\overline{s}_i, \overline{s}_{-i})|\theta_i) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\ EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right)$ and $U_i((s_i, \overline{s}_{-i})|\theta_i) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\ EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), \overline{x}_{t_{-i}})\right)$. So, it is sufficient to prove that $EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right) \geq EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), \overline{x}_{t_{-i}})\right)$ for all $t_{-i} \in \Theta_{-i}$. Let $t_{-i} \in \Theta_{-i}$. By definition of $\overline{x}_{\theta_i}$, it holds that $\overline{x}_{\theta_i} \succeq_{\theta_i} s_i(\theta_i)$. The weakly monotonic condition implies that $f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, a_{-i}) \geq f_{(\theta_i, t_{-i})}, (s_i(\theta_i), a_{-i})$ for all $a \in X$ such that $a_i = o_i$. This implies that, $f_{(\theta_i, t_{-i})}(\overline{x}_{\theta_i}, a_{-i}) - f_{(\theta_i, t_{-i})}(a) \geq f_{(\theta_i, t_{-i})}(s_i(\theta_i), a_{-i}) - f_{(\theta_i, t_{-i})}(a)$ for all $a \in X$ such that $a_i = o_i$. It follows that $EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right) \geq EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), \overline{x}_{t_{-i}})\right)$. Hence, $\overline{s}$ is a pure strategy Bayesian Nash equilibrium of the game $\Gamma^{\mathcal{B}}$.

Second, let us show that $\overline{s}$ is Pareto-efficient. Assume it is not. Then, there exits a strategy profile $s \in S$ such that for all $i \in N$ and $\theta_i \in \Theta_i$, $U_i(s|\theta_i) \geq U_i(\overline{s}|\theta_i)$, with the inequality being strict for some i. This implies that $\sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\ EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), s_{-i}(t_{-i}))\right) \geq \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\ EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right)$. That is, $\sum_{t_{-i} \in \Theta_{-i}} (\frac{1}{|\Theta_{-i}|})\ EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), s_{-i}(t_{-i}))\right) \geq \sum_{t_{-i} \in \Theta_{-i}} (\frac{1}{|\Theta_{-i}|})\ EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right)$. This implies that $\sum_{t_{-i} \in \Theta_{-i}} EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), s_{-i}(t_{-i}))\right) \geq \sum_{t_{-i} \in \Theta_{-i}} EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right)$ for all $i \in N$ and $\theta_i \in \Theta_i$, with the inequality being strict for some i. It follows that

$$\sum_{i \in N} \sum_{\theta_i \in \Theta_i} \sum_{t_{-i} \in \Theta_{-i}} EFP_i\left(f_{(\theta_i, t_{-i})}, (s_i(\theta_i), s_{-i}(t_{-i}))\right) > \sum_{i \in N} \sum_{\theta_i \in \Theta_i} \sum_{t_{-i} \in \Theta_{-i}} EFP_i\left(f_{(\theta_i, t_{-i})}, (\overline{x}_{\theta_i}, \overline{x}_{t_{-i}})\right)$$

That is

$$\sum_{i \in N} \sum_{\theta \in \Theta} EFP_i\left(f_\theta, s(\theta)\right) > \sum_{i \in N} \sum_{\theta \in \Theta} EFP_i\left(f_\theta, \overline{s}(\theta)\right)$$

This implies that $\sum_{\theta \in \Theta} \sum_{i \in N} EFP_i\left(f_\theta, s(\theta)\right) > \sum_{\theta \in \Theta} \sum_{i \in N} EFP_i\left(f_\theta, \overline{s}(\theta)\right)$. Ex-post efficiency applies and we get that $\sum_{\theta \in \Theta} f_\theta(s(\theta)) > \sum_{\theta \in \Theta} f_\theta(\overline{s}(\theta))$. This is a contradiction since ex-post monotonicity implies $\sum_{\theta \in \Theta} f_\theta(s(\theta)) \leq \sum_{\theta \in \Theta} f_\theta(\overline{s}(\theta))$. Hence, the equilibrium $\overline{s}$ is Pareto-efficient.

Last, if the technology in each state of the economy is strictly monotonic, then the equilibrium $\overline{s}$ Pareto-dominates any strategy profile $s \in S$, since the ex-post fair pay is increasing in marginal contributions and the technology f_θ admits a unique maximum at $\overline{s}(\theta)$ for all $\theta \in \Theta$. Hence, the strategy profile $\overline{s}$ is the unique Pareto-efficient equilibrium of the economy. ■

2.10 Appendix C. Proofs from Section 2.5

Proof (Theorem 7) : *Let* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, (c_i), EFP_i, (u_i)\rangle$ *be a fair extended Bayesian economy. For all* $i \in N$ *and* $(\theta, x) \in \Theta \times X$*, agent* i*'s utility function is given by:*

$$u_i(\theta, x) = EFP_i(f_\theta, x) - c_i(\theta, x_i)$$

Let the production function F *be defined for all* $\theta \in \Theta$ *and* $x \in X$ *by:*

$$F_\theta(x) = f_\theta(x) - \sum_{i \in N} c_i(\theta, x_i)$$

Calculations show that:

$$EFP_i(F_\theta, x) = EFP_i(f_\theta, x) - c_i(\theta, x_i) = u_i(\theta, x)$$

As a result, the fair extended Bayesian economy $\mathcal{B} = \langle N, (\Theta_i), (X_i), (o_i), (p_i), f, (c_i), EFP_i, (u_i)\rangle$ *is equivalent to the fair Bayesian economy* $\langle N, (\Theta_i), (X_i), (o_i), (p_i), F, EFP_i, (u_i)\rangle$*. From Theorem 5, we deduce that the extended fair Bayesian economy,* $\mathcal{B}$*, admits an equilibrium. In addition, if for all* $\theta \in \Theta$ *the production function* F *is weakly monotonic and the agents' beliefs,* p_i *for all* $i \in N$*, are drawn from uniform distributions, then it follows from Theorem 6 that the economy admits a Pareto-efficient equilibrium, which is unique if* F *is ex-post strictly monotonic.* ■

Proof (Theorem 8) : *The proof follows the same pattern as the proof of Theorem 5. Let* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, \Phi, (u_i)\rangle$ *be an inclusive Myerson economy. For all* $i \in N$ *and* $(\theta, x) \in \Theta \times X$*, agent* i*'s utility is given by:*

$$u_i(\theta, x) = (1 - T(\theta))EFP_i(f_\theta, x) + T(\theta)\frac{f_\theta(x)}{n} = u_i^1(\theta, x) + u_i^2(\theta, x),$$

where $u_i^1(\theta, x) = (1 - T(\theta))EFP_i(f_\theta, x)$ *and* $u_i^2(\theta, x) = T(\theta)\frac{f_\theta(x)}{n}$*. Then the expected utility of player* i *associated to a strategy profile* $s \in S$ *and conditional on his type* θ_i *is given by:*

$$U_i((s_i, s_{-i})|\theta_i) \;=\; \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\, u_i(\theta_i, t_{-i}, s_i(\theta_i), s_{-i}(t_{-i})) = U_i^1((s_i, s_{-i})|\theta_i) + U_i^2((s_i, s_{-i})|\theta_i)$$

where for all $k = 1, 2$,

$$U_i^k((s_i, s_{-i})|\theta_i) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i)\; u_i^k(\theta_i, t_{-i}, s_i(\theta_i), s_{-i}(t_{-i})).$$

Let $\Gamma^{\mathcal{B}}$ *be the Bayesian game associated to* $\mathcal{B}$ *and* $\Gamma^{\mathcal{B}s}$ *its corresponding Selten game. Let* $\mathbb{P}$ *be the prior probability distribution of the Bayesian economy. Assume* $\Theta_i = \{\theta_{i1}, \theta_{i2}, ..., \theta_{ik_i}\}$ *for any* $i \in N$. *For any* $i \in N$ *and* $k \in \{1, 2, ..., k_i\}$, *let* $\beta_{\theta_{ik}} = \sum_{\theta_{-i} \in \Theta_{-i}} \mathbb{P}(\theta_{ik}, \theta_{-i})$. *Consider the strictly monotonic application* $h = (h_{\theta_{ik}}) : \mathbb{R} \to \mathbb{R}^{N^s}$ *defined for all* $i \in N$, $k \in \{1, 2, ..., k_i\}$ *and* $x \in \mathbb{R}$ *by* $h_{\theta_{ik}}(x) = \beta_{\theta_{ik}} x$. *Let* $\Gamma_h^{\mathcal{B}s}$ *be the strictly monotonic transform of* $\Gamma^{\mathcal{B}s}$ *by* h.

From Lemmas 1 and 2 it suffices to show that Equation 2.12 of Lemma 2 is satisfied for the strategic form game $\Gamma_h^{\mathcal{B}s}$. *Let* $\theta_{ik}, \theta_{jm} \in N^s$, *with* $i, j \in N$, $k \in \{1, 2, ..., k_i\}$ *and* $m \in \{1, 2, ..., k_j\}$. *Let* $x_{\theta_{ik}}, y_{\theta_{ik}} \in X_{\theta ik} = X_i$, $x_{\theta_{jm}}, y_{\theta_{jm}} \in X_{\theta_{jm}} = X_j$ *and* $a \in \prod_{\theta_{lt} \in N^s \setminus \{\theta_{ik}, \theta_{jm}\}} X_{\theta_{lt}}$. *Let* A, B, C *and* D *be strategies of the strategic form game* $\Gamma_h^{\mathcal{B}s}$ *as defined in Lemma 2 with* i *and* j *being replaced by* θ_{ik} *and* θ_{jm} *respectively. I want to show that the following equation is satisfied*

$$: \begin{array}{l} h_{\theta_{ik}}(U_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U_{\theta_{ik}}(A)) + h_{\theta_{jm}}(U_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U_{\theta_{jm}}(B)) + \\ h_{\theta_{ik}}(U_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U_{\theta_{ik}}(C)) + h_{\theta_{jm}}(U_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U_{\theta_{jm}}(D)) = 0 \end{array} \tag{2.14}$$

Denote by $\mathcal{T}$ *the left-hand side of Equation 2.14. It can be written as* $\mathcal{T} = \mathcal{T}^1 + \mathcal{T}^2$, *where:*

$$\begin{aligned} \mathcal{T}^1 &= h_{\theta_{ik}}(U^1_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U^1_{\theta_{ik}}(A)) + h_{\theta_{jm}}(U^1_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U^1_{\theta_{jm}}(B)) \\ &+ h_{\theta_{ik}}(U^1_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U^1_{\theta_{ik}}(C)) + h_{\theta_{jm}}(U^1_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U^1_{\theta_{jm}}(D)) \end{aligned}$$

$$\begin{aligned} \mathcal{T}^2 &= h_{\theta_{ik}}(U^2_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(A)) + h_{\theta_{jm}}(U^2_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U^2_{\theta_{jm}}(B)) \\ &+ h_{\theta_{ik}}(U^2_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(C)) + h_{\theta_{jm}}(U^2_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U^2_{\theta_{jm}}(D)) \end{aligned}$$

From the proof of Theorem 5, it can be easily shown that $\mathcal{T}^1 = 0$. *As a result, it remains to show that* $\mathcal{T}^2 = 0$ *to complete the proof.*

Let $K_1 = h_{\theta_{ik}}(U^2_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(A))$, $K_2 = h_{\theta_{jm}}(U^2_{\theta_{jm}}(C)) - h_{\theta_{jm}}(U^2_{\theta_{jm}}(B))$, $K_3 = h_{\theta_{ik}}(U^2_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(C))$ *and* $K_4 = h_{\theta_{jm}}(U^2_{\theta_{jm}}(A)) - h_{\theta_{jm}}(U^2_{\theta_{jm}}(D))$. *The proof can be broken down into two cases.*

Case 1: *$i = j$, then $\theta_{ik}, \theta_{im} \in \Theta_i$. We have:*

$$\begin{aligned}
K_1 &= h_{\theta_{ik}}(U^2_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(A)) \\
&= \beta_{\theta_{ik}} U^2_{\theta_{ik}}(y_{\theta_{ik}}, x_{\theta_{im}}, a) - \beta_{\theta_{ik}} U^2_{\theta_{ik}}(x_{\theta_{ik}}, x_{\theta_{im}}, a) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u^2_i(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u^2_i(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left[\frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}})\right]
\end{aligned}$$

$$\begin{aligned}
K_3 &= h_{\theta_{ik}}(U^2_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(C)) \\
&= \beta_{\theta_{ik}} U^2_{\theta_{ik}}(x_{\theta_{ik}}, y_{\theta_{im}}, a) - \beta_{\theta_{ik}} U^2_{\theta_{ik}}(y_{\theta_{ik}}, y_{\theta_{im}}, a) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik}) \left[\frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}})\right]
\end{aligned}$$

It follows that $K_1 + K_3 = 0$. Similarly, we have

$$\begin{aligned}
K_2 &= h_{\theta_{im}}(U^2_{\theta_{im}}(C)) - h_{\theta_{im}}(U^2_{\theta_{im}}(B)) \\
&= \beta_{\theta_{im}} U^2_{\theta_{im}}(y_{\theta_{ik}}, y_{\theta_{im}}, a) - \beta_{\theta_{im}} U^2_{\theta_{im}}(y_{\theta_{ik}}, x_{\theta_{im}}, a) \\
&= \beta_{\theta_{im}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{im}) \left[\frac{1}{n} T(\theta_{im}, t_{-i}) f_{(\theta_{im}, t_{-i})}(y_{\theta_{im}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{im}, t_{-i}) f_{(\theta_{im}, t_{-i})}(x_{\theta_{im}}, a_{t_{-i}})\right]
\end{aligned}$$

$$\begin{aligned}
K_4 &= h_{\theta_{im}}(U^2_{\theta_{im}}(A)) - h_{\theta_{im}}(U^2_{\theta_{im}}(D)) \\
&= \beta_{\theta_{im}} U^2_{\theta_{im}}(x_{\theta_{ik}}, x_{\theta_{im}}, a) - \beta_{\theta_{im}} U^2_{\theta_{im}}(x_{\theta_{ik}}, y_{\theta_{im}}, a) \\
&= \beta_{\theta_{im}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{im}) \left[\frac{1}{n} T(\theta_{im}, t_{-i}) f_{(\theta_{im}, t_{-i})}(x_{\theta_{im}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{im}, t_{-i}) f_{(\theta_{im}, t_{-i})}(y_{\theta_{im}}, a_{t_{-i}})\right]
\end{aligned}$$

It follows that $K_2 + K_4 = 0$.

Hence, $\mathcal{T}^2 = K_1 + K_2 + K_3 + K_4 = 0$ if $i = j$. That is, $\mathcal{T}^2 = 0$ if the two players of the strictly monotonic transform of the Selten game $\Gamma^{\mathcal{B}s}$ by h are drawn from the same type set. Let us see what happens when they are drawn from different type sets.

case 2: *$i \neq j$, then $\theta_{ik} \in \Theta_i$ and $\theta_{jm} \in \Theta_j$. We have:*

$$
\begin{aligned}
K_1 &= h_{\theta_{ik}}(U^2_{\theta_{ik}}(B)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(A)) \\
&= \beta_{\theta_{ik}} U^2_{\theta_{ik}}(y_{\theta_{ik}}, x_{\theta_{jm}}, a) - \beta_{\theta_{ik}} U^2_{\theta_{ik}}(x_{\theta_{ik}}, x_{\theta_{jm}}, a) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i^2(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i^2(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \\
&= \beta_{\theta_{ik}} \sum_{(\theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left(u_i^2(\theta_{ik}, \theta_{jm}, t_{-ij}, y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - u_i^2(\theta_{ik}, \theta_{jm}, t_{-ij}, x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \right) \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left(u_i^2(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) - u_i^2(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) \right) \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left[\tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - \tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \right] \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left[\frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}}) \right] \\
&= K_{11} + K_{12},\ \textit{where}
\end{aligned}
$$

$$K_{11} = \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left[\tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - \tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \right]$$

and $K_{12} = \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left[\tfrac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}}) - \tfrac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}}) \right]$.

$$
\begin{aligned}
K_3 &= h_{\theta_{ik}}(U^2_{\theta_{ik}}(D)) - h_{\theta_{ik}}(U^2_{\theta_{ik}}(C)) \\
&= \beta_{\theta_{ik}} U^2_{\theta_{ik}}(x_{\theta_{ik}}, y_{\theta_{jm}}, a) - \beta_{\theta_{ik}} U^2_{\theta_{ik}}(y_{\theta_{ik}}, y_{\theta_{jm}}, a) \\
&= \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i^2(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) - \beta_{\theta_{ik}} \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ u_i^2(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) \\
&= \beta_{\theta_{ik}} \sum_{(t_j = \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left(u_i^2(\theta_{ik}, \theta_{jm}, t_{-ij}, x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - u_i^2(\theta_{ik}, \theta_{jm}, t_{-ij}, y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \right) \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left(u_i^2(\theta_{ik}, t_{-i}, x_{\theta_{ik}}, a_{t_{-i}}) - u_i^2(\theta_{ik}, t_{-i}, y_{\theta_{ik}}, a_{t_{-i}}) \right) \\
&= \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left[\tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - \tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \right] \\
&+ \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left[\frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}}) - \frac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}}) \right] \\
&= K_{31} + K_{32},\ \textit{where}
\end{aligned}
$$

$$K_{31} = \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) \left[\tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - \tfrac{1}{n} T(\theta_{ik}, \theta_{jm}, t_{-ij}) f_{(\theta_{ik}, \theta_{jm}, t_{-ij})}(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \right]$$

and $K_{32} = \beta_{\theta_{ik}} \sum_{t_{-i} = (t_j \neq \theta_{jm}, t_{-ij}) \in \Theta_{-i}} p_i(t_{-i}|\theta_{ik})\ \left[\tfrac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(x_{\theta_{ik}}, a_{t_{-i}}) - \tfrac{1}{n} T(\theta_{ik}, t_{-i}) f_{(\theta_{ik}, t_{-i})}(y_{\theta_{ik}}, a_{t_{-i}}) \right]$.

From case 1 we have that $K_{12} + K_{32} = 0$. Setting $f_{(\theta_{ik}, \theta_{jm}, t_{-ij})} = f$, it follows that:

$$
\begin{aligned}
K_1 + K_3 &= K_{11} + K_{31} \\
&= \frac{1}{n} \beta_{\theta_{ik}} \sum_{t_{-ij} \in \Theta_{-ij}} p_i((\theta_{jm}, t_{-ij})|\theta_{ik}) T(\theta_{ik}, \theta_{jm}, t_{-ij}) \left[f(y_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) - f(x_{\theta_{ik}}, x_{\theta_{jm}}, a_{t_{-ij}}) \right. \\
&+ \left. f(x_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) - f(y_{\theta_{ik}}, y_{\theta_{jm}}, a_{t_{-ij}}) \right]
\end{aligned}
$$

But $\beta_{\theta_{ik}} = \sum_{\theta_{-i}\in\Theta_{-i}} \mathbb{P}(\theta_{ik},\theta_{-i})$ *and* $p_i((\theta_{jm},t_{-ij})|\theta_{ik}) = \frac{\mathbb{P}(\theta_{ik},\theta_{jm},t_{-ij})}{\sum_{\theta_{-i}\in\Theta_{-i}}\mathbb{P}(\theta_{ik},\theta_{-i})}$. *This implies that*

$$\begin{aligned} K_1 + K_3 &= \frac{1}{n}\sum_{t_{-ij}\in\Theta_{-ij}} \mathbb{P}(\theta_{ik},\,\theta_{jm},\,t_{-ij})T(\theta_{ik},\theta_{jm},t_{-ij})\left[f(y_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(x_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}})\right. \\ &+ \left. f(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})\right] \end{aligned}$$

Similarly, I show that:

$$\begin{aligned} K_2 + K_4 &= \frac{1}{n}\beta_{\theta_{jm}}\sum_{t_{-ij}\in\Theta_{-ij}} p_j((\theta_{ik},t_{-ij})|\theta_{jm})T(\theta_{ik},\theta_{jm},t_{-ij})\left[f(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(y_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}})\right. \\ &+ \left. f(x_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})\right] \end{aligned}$$

But $\beta_{\theta_{jm}} = \sum_{\theta_{-j}\in\Theta_{-j}} \mathbb{P}(\theta_{jm},\theta_{-j})$ *and* $p_j((\theta_{ik},t_{-ij})|\theta_{jm}) = \frac{\mathbb{P}(\theta_{ik},\theta_{jm},t_{-ij})}{\sum_{\theta_{-j}\in\Theta_{-j}}\mathbb{P}(\theta_{jm},\theta_{-j})}$. *This implies that:*

$$\begin{aligned} K_2 + K_4 &= \frac{1}{n}\sum_{t_{-ij}\in\Theta_{-ij}} \mathbb{P}(\theta_{ik},\,\theta_{jm},\,t_{-ij})T(\theta_{ik},\theta_{jm},t_{-ij})\left[f(y_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}}) - f(y_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}})\right. \\ &+ \left. f(x_{\theta_{ik}},x_{\theta_{jm}},a_{t_{-ij}}) - f(x_{\theta_{ik}},y_{\theta_{jm}},a_{t_{-ij}})\right] \end{aligned}$$

It follows that $\mathcal{T}^2 = K_1 + K_2 + K_3 + K_4 = 0$.

Hence, the Bayesian game $\Gamma^{\mathcal{B}}$ *associated to the inclusive and fair Bayesian economy* $\mathcal{B}$ *admits a pure strategy Bayesian Nash equilibrium.*

The existence and uniqueness of a Pareto-efficient equilibrium are established using the same reasoning as in the proof of Theorem 6. Therefore, the detailed proof is omitted. ■

Proof (Theorem 9) : *Let* $\mathcal{B} = \langle N,(\Theta_i),(X_i),(o_i),(p_i),f,T,(c_i),EFP_i,(u_i)\rangle$ *be a fair Comprehensive Bayesian economy. For all* $i \in N$ *and* $(\theta,x) \in \Theta \times X$, *agent* i*'s utility function is given by:*

$$u_i(\theta,x) = (1 - T(\theta))EFP_i(f_\theta,x) + T(\theta)\frac{f_\theta(x)}{n} - c_i(\theta,x_i).$$

Let the production function F *be defined for all* $\theta \in \Theta$ *and* $x \in X$ *by:*

$$F_\theta(x) = f_\theta(x) - \frac{1}{1 - T(\theta)}\sum_{i\in N} c_i(\theta,x_i)$$

A calculation shows that:

$$u_i(\theta,x) = (1 - T(\theta))EFP_i(F_\theta,x) + T(\theta)\frac{f_\theta(x)}{n}$$

Once the agents' utility functions are expressed in this form, the proof proceeds along the same line of reasoning as in Theorem 8. ■

2.11 Appendix D. Proofs from Section 2.6

Proof (Proposition 1) : *By contradiction, suppose that the mechanism* $\mathcal{M}^*$ *is not Pareto-efficient. Then, there exists a fair Bayesian economy* $\mathcal{B}$ *in which there is a type profile* θ *and an action profile* x_0 *such that:*

$$u_i(\theta, x) \geq u_i(\theta, \mathcal{M}^*(\theta)) \text{ for all } i \in N, \text{ and } u_i(\theta, x_0) > u_i(\theta, \mathcal{M}^*(\theta)) \text{ for some } i \in N$$

This implies that $\sum_{i \in N} u_i(\theta, x_0) > \sum_{i \in N} u_i(\theta, \mathcal{M}^*(\theta)$. *This is equivalent to:* $\sum_{i \in N} EFP_i(f_\theta, x_0) > \sum_{i \in N} EFP_i(f_\theta, \mathcal{M}^*(\theta)$. *Since the ex-post fair pay satisfies ex-post efficiency, then it follows that* $f_\theta(x_0) > f_\theta(\mathcal{M}^*(\theta)) = \max_{x \in X} f_\theta(x)$, *which is a contradiction. Hence, the mechanism* $\mathcal{M}^*$ *is Pareto-efficient.* ■

Proof (Theorem 10) : *Let* $\mathcal{B} = \langle N, (\Theta_i), (X_i), (p_i), f, EFP, (u_i) \rangle$ *be a fair Bayesian economy endowed with an ex-post weakly monotonic production function. By definition, for all* $i \in N$ *and* $\theta_i \in \Theta_i$, *there exists a weak order* $\succeq_{\theta_i}$ *over* X_i. *For all* $i \in N$ *and* $\theta_i \in \Theta_i$, *let* $\overline{x}_{\theta_i} \in X_i$ *such that* $\overline{x}_{\theta_i} \succeq_{\theta_i} x_i$ *for all* $x_i \in X_i$. *Such an action exists since* X_i *is finite and* $\succeq_{\theta_i}$ *is complete and transitive. Given that* f *is ex-post weakly monotonic, from the definition of the mechanism* $\mathcal{M}^*$, *it holds that for all* $\theta \in \Theta$,

$$\mathcal{M}^*(\theta) = (\overline{x}_{\theta_i})_{i \in N} = \overline{x}_\theta, \text{ i.e., } \mathcal{M}_i^*(\theta) = \overline{x}_{\theta_i} \text{ for all } i \in N.$$

First, let us show that $\mathcal{M}^*$ *is strategy-proof in the fair economy* $\mathcal{B}$. *Let* $\theta \in \Theta$ *be a realized type profile and let* $i \in N$ *be an agent. If agent* i *is truthful and reports* θ_i, *then he would get a utility level equal to*

$$v_i(\mathcal{M}^*(\theta)) = u_i(\theta, \mathcal{M}^*(\theta)) = EFP_i(f_\theta, \overline{x}_\theta)$$

If agent i *is untruthful and reports the type* $\widehat{\theta_i}$, *then he would get a utility level equal to*

$$v_i(\mathcal{M}^*(\widehat{\theta_i}, \theta_{-i})) = u_i(\theta, (\mathcal{M}_i^*(\widehat{\theta_i}, \theta_{-i}), \mathcal{M}_{-i}^*(\theta)) = EFP_i(f_\theta, (x_i, \overline{x}_{\theta_{-i}})) \text{ for some } x_i \in X_i$$

By definition of $\overline{x}_{\theta_i}$*, it holds that* $\overline{x}_{\theta_i} \succeq_{\theta_i} x_i$*. The ex-post weakly monotonic condition implies that* $f_\theta(\overline{x}_{\theta_i}, a_{-i}) \geq f_\theta(x_i, a_{-i})$ *for all* $a \in X$ *such that* $a_i = o_i$*. This implies that* $f_\theta(\overline{x}_{\theta_i}, a_{-i}) - f_\theta(a) \geq f_\theta(x_i, a_{-i}) - f_\theta(a)$ *for all* $a \in X$ *such that* $a_i = o_i$*. It follows that:*

$$EFP_i(f_\theta, \overline{x}_\theta) \geq EFP_i(f_\theta, (x_i, \overline{x}_{\theta_{-i}})) \text{ i.e., } v_i(\mathcal{M}^*(\theta)) \geq v_i(\mathcal{M}^*(\widehat{\theta}_i, \theta_{-i})) \quad (2.15)$$

Hence, the mechanism $\mathcal{M}^*$ *is strategy-proof. In addition, Equation 2.15 above would continue to hold if we replace the action* x_i *by the inaction point* o_i*. This implies that* $v_i(\mathcal{M}^*(\theta)) \geq v_i(o_i, \mathcal{M}^*_{-i}(\theta))$*. Hence, the mechanism* $\mathcal{M}^*$ *is individually rational.* ■

Chapter 3

Valuing Matching Opportunities: Imperfect Information, Roth's Risk Neutrality, and Preferences

If you were looking for a full-time job and you received offers from several companies, which would you accept if you are unsure about the abilities or types of co-workers? If offered the opportunity to participate in two games, which would you choose to play? Classical utility theory addresses a class of choice problems that involve only one agent facing choice alternatives that consist of objects (e.g., consumption bundles, projects, car brand, etc.) not making any decision (Neumann and Morgenstern, 1944). Numerous real-life choice problems, however, involve choice alternatives that consist of sets of agents making rational decisions and whose types are not perfectly known (Blair and McLean, 1990). Examples of such choice alternatives may include companies with several employees, marriage partners, friends, sport teams, collaborators on an investment project, and so on. Such choice situations where multiple agents interact are generally called *games*. Choosing between two games to participate in requires evaluating the utility of playing each game; because such a choice also requires making predictions on other players' actions in each game, it is not clear how the tools of traditional utility theory can be extended to deal with this problem (Roth, 1977; Roth, 2019). In this paper, we address the problem of evaluating the utility of participating in organizations involving several interacting agents and characterized by *imperfect information* in the sense that agents' types (or actions) are uncertain ex-ante.[1] We also examine the implications of our analysis for equilibrium behaviors.

The question of how to evaluate the prospect (or utility) of playing a game is old, but this research agenda is still in its infancy (Roth, 2019). This question was originally introduced by

[1] As it will become evident, our model of organization can be applied to joint ventures, firms, some two-sided markets, and so on.

Shapley (1953) as follows:

> At the foundation of the theory of games is the assumption that the players of a game can evaluate, in their utility scales, every "prospect" that might arise as a result of a play. In attempting to apply the theory to any field, one would normally expect to be permitted to include, in the class of "prospects," the prospect of having to play a game. The possibility of evaluating games is therefore of critical importance. So long as the theory is unable to assign values to the games typically found in application, only relatively simple situations— where games do not depend on other games— will be susceptible to analysis and solution. p1

Emphasizing the importance of this question for market design, Roth (2019) writes:

> The question that Shapley posed, of how to evaluate the prospect of playing a game, is as important as ever, particularly because market designers have to take into account that participants in a market have the opportunity to choose which marketplaces to participate in ... Roth (2019), pxx

Shapley (1953) originally addressed this question for the class of transferable-utility games following a normative approach, obtaining a solution that has become known as the Shapley value. Roth (2019) argues that Shapley's approach fails to capture the kinds of preferences over risky outcomes that describe expected utility theory. He writes:

> Shapley proposed not to predict the outcome of the game, but instead to formulate a function that could be interpreted along the lines of the expected utility of playing a game, from each of its positions. However, the formal approach he considered did not involve the kinds of preferences over risky outcomes that von Neumann and Morgenstern had modeled with expected utility. Roth (2019), pxix

Early on, this observation motivated Roth (1977) to develop a VNM utility approach to the original question asked by Shapley. Roth (1977) considers positions in a game as objects for which individuals have preferences and elicits conditions under which the cardinal utility representing these preferences coincides with the Shapley value. Consistent with the theory of expected utility theory (Neumann and Morgenstern, 1944), Roth (1977) introduces the notions of strategic and ordinary risk and shows that the Shapley value of a game equals its utility if and only if the underlying preferences satisfy reasonable properties and are neutral to both strategic and ordinary risks. He defines strategic risk as the uncertainty arising from the interaction in a game, while ordinary risk involves the uncertainty arising from the chance mechanism involved in lotteries.

While a satisfactory answer has been obtained for transferable-utility (TU) games, this class of games is too simplistic to capture features of real-world markets. Indeed, Roth (2019) writes:

> One important area of future research will involve how to extend the insights gained, and the tools developed primarily from TU games to the wide class of other kinds of models that economists and game theorists now explore, as well as those explored by other social scientists. Roth (2019), pxx

Our paper contributes to this research agenda by extending Roth's approach to a different class of environments much wider than TU games.[2] Our environment is modeled as an **organization with imperfect information** with respect to agents' types. It describes an environment in which each agent is unsure about his type and that of other agents prior to their realization. It is formalized as a tuple $\langle N, X, o, f, \sigma\rangle$, where N is the set of agents, and $X = \prod_{i \in N} X_i$ is the set of type profiles. An element of X_i specifies a possible **type** of agent, with o being the unproductive type profile; importantly, the elements of X_i may also be interpreted as the different **actions** of agent i or the different **states** of the world to which agent i is subjugated. These latter interpretations imply that the realization of a given type or state of the world fully determines the action of agent i. f is a function that maps each type profile to an outcome in the set of real numbers.[3] Finally, $\sigma = (\sigma_i)_{i \in N}$ is a probability distribution over the set X. Note that there is no secret or private information in this setting. However, information is imperfect because the results of chance in the type of each agent are not known prior to them occurring. The informational setting is therefore that of perfect uncertainty.

The model of an organization with imperfect information generalizes the notion of TU games in several respects, and applies to a wider range of real-life settings. In a TU game, a player has only two possible types—*participate* and *not participate*. So, all players have the same type set, which significantly limits the application of this model to real-life situations. In our model, an agent can have more than two types, and the type set may be different for each agent. Our model also differs in assuming that each agent's type is subject to uncertainty (σ_i), where the nature of uncertainty may again vary across agents. For example, an agent's type

[2] A TU game is a couple (N, v), where N is the set of players and $v : 2^N \longrightarrow \mathbb{R}$ is a function from the set of all subsets of N to the set of real numbers.

[3] As we will see later, the function f can be diversely interpreted depending on the context to which it applies; f can be viewed as the technology of a firm (or its surplus function), a measure of systemic risk in a financial market, aggregate income in an economy, the prevalence of information (or infection) in a social network, or simply the sum of individual utilities in a strategic form game or an economy.

may be determined by weather conditions, health conditions, or any other life circumstance that may not be known in advance and therefore subject to uncertainty. This is a more realistic assumption than the one made in Shapley (1953) and Roth (1977) where it is assumed that the so-called grand coalition forms, meaning that all players choose to participate in the game with certainty. Our model acknowledges that in real-life settings, people are subject to many kinds of uncertainty that may constrain or determine their participation in social interactions. These distinct features of our model make it applicable to a wide range of environments. Indeed, we show applications to firms, matchings, and networks.[4]

We identify an organization with imperfect information by its production function and assume that individuals have preferences over the set of strategic positions, where a strategic position is a pair (f, i) that specifies an organization with imperfect information f and a position i within the organization f. An individual prefers a strategic position (f, i) over a strategic position (g, j) if she prefers to be at position i in the organization f than to be at position j in the organization g. For instance, an individual might prefer to be an assistant professor at the University of Ottawa than to be a consultant at the World Bank. Another individual might prefer to be a substitute player at Real de Madrid than to be a starting player at Olympique de Marseille. These are the kinds of comparisons that we are interested in in this study. Our main goal consists of determining conditions under which these preferences can be represented by a unique cardinal utility. To achieve this goal, we generalize to our setting the notions of ordinary and strategic risks introduced by Roth (1977) and show that a preference relation that is neutral to ordinary and strategic risks and that satisfies some regularity conditions has a unique cardinal representation, called the ex-ante Shapley value (Theorem 11). This result generalizes Roth's result describing the Shapley value as a von Neumann-Morgenstern utility. Indeed, the ex-ante Shapley value defined for organizations with imperfect information generalizes the Shapley value defined for TU games.

In the treatment above, uncertainty is given ex-ante. A second question we answer is whether uncertainty can arise endogenously as an equilibrium play. We show that there always exists a profile of probability distributions that maximize the utility functions of agents when they are neutral to both ordinary and strategic risks. More precisely, we associate to any tuple $\langle N, X, o, f \rangle$ a strategic form game $\langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where for each player $i \in N$, the set of strategies S_i is the set of all probability distributions over X_i, and u_i is the utility function that maps each profile of probability distributions $s \in S = \prod_{i \in N} S_i$ into a real number given by the ex-ante Shapley value associated to the organization with imperfect information $\langle N, X, o, f, s \rangle$.

[4]In the context of the firm, another interpretation of our model is that σ_i is the probability distribution from which the ability of worker i is drawn.

We find that there always exists a profile of probability distributions $s^* \in S$ that maximizes the players' utilities, resulting in a Nash equilibrium (Theorem 12). In this case, the organization $\langle N, X, o, f, s^* \rangle$ is said to be rationalizable. In addition, we prove the existence of rationalizable organizations $\langle N, X, o, f, s^* \rangle$ where s^* is a pure-strategy Nash equilibrium.

We develop applications of our model to firms and two-sided economies. First, we consider a production environment where workers supply costly efforts that are converted by a technology into a monetary output. We address the issue of a firm's informational challenge regarding the individual costs of workers. Our findings reveal that under Roth's risk neutrality, the net utilities of workers remain unchanged, regardless of whether the firm possesses cost information and hence pools all individual costs or not (Proposition 2). In other words, workers' incentives and equilibrium behaviors do not vary depending on whether the firm pools all the costs or lets each worker incur his own (private) cost. We also show that neutrality to both ordinary and strategic risks is sufficient to guarantee the existence of a pure strategy Nash equilibrium in organizations where workers act strategically (Proposition 3). In our second application, we consider an assignment market consisting of sellers and buyers as in Shapley and Shubik (1971). There are as many sellers as there are units of an indivisible good. Each seller owns one unit of the good, for which he has a valuation. Each buyer also has a valuation for each good. The main goal consists of trading these goods based on agents' valuations. We model this market as an organization with imperfect information and determine the utility that an outsider agent can derive from each position in the market when his preferences satisfy reasonable axioms and are neutral to both strategic and ordinal risks. Furthermore, we show that, under certain conditions, there exists a pricing mechanism that achieves the utility of every agent in the economy (Proposition 4).

Our paper contributes to several strands of the literature. Owing to its desirable properties, the classical Shapley value, which is defined under the assumption of full information, has been applied to various classes of problems including operations management and supply chains (Gopalakrishnan et al., 2021; Kemahlıoğlu-Ziya and Bartholdi III, 2011), cost allocations (Littlechild and Owen, 1973; Dubey, 1982), fair division (Moulin, 1992), contract design (Winter, 2002), political power measurement (Shapley and Shubik, 1954; Freixas, Marciniak, and Pons, 2012; Pongou and Tchantcho, 2021), bankruptcy problem (Aumann and Maschler, 1985), network centrality measurement (Grofman and Owen, 1982; Pongou and Tondji, 2018), queueing problems (Maniquet, 2003), input valuation in discrete settings (Pongou and Tondji, 2018), unfairness and income inequality (Aguiar, Pongou, and Tondji, 2018; Aguiar et al., 2019), among others.[5] However, the extension of the Shapley value to study problems involving imperfect information has received less attention. Pongou and Tondji (2018) is the only other paper that studies this problem, but this paper follows Shapley's normative approach to characterize the ex-ante Shapley value. We differ in that we follow Roth's approach and so our focus is on evaluating the utility of participating in an organization where members lack information about each other. We also differ in our applications. Indeed, our applications to profit-maximizing firms and two-sided assignments market are novel and lead to new findings.

Our paper also relates to studies that have followed Roth (1977) 's pioneering approach. Indeed, Roth (1977) 's work has inspired several other studies which apply the theory of expected utility to multi-agent economic models such as social networks (Brink and Rusinowska, 2022), strategic situations (Agastya, 2008), games (Feltovich and Swierzbinski, 2011; Blair and McLean, 1990), and, more recently, rationing problems (Chatterjee, Ertemel, and Kumar, 2023). Our work departs significantly from these studies as we are examining a model that involves *imperfect information* regarding agents' types.

Finally, this chapter complements the recent work of Klein (2015), who investigates the determinants of fairness perceptions, particularly in competitive environments. Klein (2015) explores how the source of inequality–effort versus luck–influences individuals' perception of what is a socially just distribution of income. Using an experimental design, he finds that individuals who earn low incomes due to low effort tend to support greater redistribution than those who achieve high incomes through effort. This suggests that fairness views are endogenously shaped by individuals' own experiences with performance-based inequality. This chapter complements this study by offering a theoretical foundation for how such fairness perceptions may arise from deeper attitudes toward uncertainty. We model environments in which agents must evaluate payoff-relevant positions under both ordinary and strategic risks, and characterize the conditions under which their fairness views remain stable. In particular, we show that an individual who is neutral to both ordinary and strategic risks is more likely to allocate a larger share of the output to high-productivity workers relative to their low-productivity counterparts. Thus, while Klein (2015) identifies empirically that fairness views depend on whether inequalities stem from effort or luck, our analysis explains how individual attitudes toward risk shape those perceptions in the first place. Together, these studies bridge normative and behavioral approaches, highlighting that fairness judgments are not only context-dependent but also mediated by deeper risk postures.

[5] For authoritative volumes on the Shapley value and its various applications, see Roth (1988) and Algaba, Fragnelli, and Sánchez-Soriano (2019). In particular, there are various applications of the Shapley value, along with other rules such as the proportional rule (Csóka and Herings, 2021) and the equal-split rule (Bergantinos and Moreno-Ternero, 2020), to classical problems in management and operations research.

The remainder of this paper is organized as follows. Section 3.1 introduces the model of an organization with imperfect information. In Section 3.2, we characterize the utility function for organizations with imperfect information based on ordinal preferences and prove the existence of rationalizable probability distributions of agents' actions. Section 3.3 shows some applications of our theory. Section 3.4 concludes.

3.1 A model of organizations with imperfect information

An organization with imperfect information is defined as a tuple $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$ where:

- $N = \{1, 2, \dots, n\}$ is a finite set of agents or positions.
- $X = \prod_{i \in N} X_i$ is a finite set of agents' type profiles, X_i being the set of **types** of agent i. Importantly, X_i may also be interpreted as the different **action** (e.g., ability) of agent i, with the implication that agents are truthful and an agent's action is completely determined by his type. X_i may also be interpreted as the set of the different **states** (e.g., weather, health) of the world that determine agent i's action. While we will keep all these interpretations in this paper, we will still be calling X_i a set of types for simplicity.[6]
- $o = (o_1, o_2, \dots, o_n) \in X$ represents the unproductive type.
- $f : X \to \mathbb{R}$ is a production (or output) function. It maps a type profile $x \in X$ to a real number output $f(x)$. $f(x)$ is interpreted as the outcome of the profile of types x or the wealth generated by x. For simplicity, we assume that the unproductive type is worth zero, $f(o) = 0$.
- σ is a probability distribution function over X. For a type profile $x \in X$, $\sigma(x) > 0$ represents the probability of the realization of x.

The model of an organization with imperfect information formalizes a wide range of real-life situations. We show a few examples below.

Example 25 (Firm) : *In a firm $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$, N consists of the set of workers. The set of types, X_i, available to a worker i consists of the different abilities that the worker possesses. This set can also be interpreted as the set of the different states of nature. Each profile of workers' ability is converted by the firm's technology f into an output.*

[6] In some of the applications we will show later, actions are chosen, whereas types and states are exogenously given.

Example 26 (Agrarian economy) : *In an agrarian economy, N can consist of a landowner (agent 1) and a finite number of laborers (agent 2 up to agent n) who supply labor to the landowner (Shapley and Shubik, 1967). The landowner has a parcel of land available for various agricultural purposes. The set X_1 is described by the quality of land, rated as not fertile, less fertile, fertile, and very fertile. Land quality is subject to uncertainty as it depends on weather conditions. For $i > 1$, $X_i = \{0, 1, 2, 3, 4\}$ describes the health state of laborer i, rated as poor, fair, good, very good, and excellent. Here, each laborer's type (or productivity) is entirely determined by the state of his health, and health state is uncertain ex-ante. The production function f maps each combination of land quality and laborers' health conditions to the amount of suitable crops that can be produced.*

Example 27 (Financial market and systemic risk) : *In a financial market, N represents a finite set of interconnected banks that are linked through financial relationships such as investment decisions. While this interconnectivity can have its advantages in promoting financial stability and economic growth, it also poses risks such as the possibility of contagion, where the failure of one bank can have a ripple effect on others. The types available to an individual bank are determined by its investment decisions in a variety of assets such as stocks, bonds, and real estate. More precisely, the types set, X_i, of a bank $i \in N$ consists of vectors $a_i = (a_{i1}, a_{i2}, a_{i3})$, where a_{i1}, a_{i2}, and a_{i3} stand respectively for the levels of investment in stocks, bonds, and real estate. Such investment decisions are drawn from a probability distribution over the set of past investment decisions. In this case, the production function f maps each profile of banks' investment decisions to a level of system-wide risk.*

Another application can be designed along the lines of Roth (2019)'s suggestion. Here, N could be the set of marketplaces, X is a set describing the actions which are likely to be available in each marketplace, and f is a function that determines the global market output.

Note that in each example above, agents are supposed to derive utility from their types. However, individual utility functions are not given. In the section below, we characterize the utility of agents along the lines of expected utility theory as in Roth (1977).

3.2 Utility functions for organizations with imperfect information

3.2.1 Utility existence

In this section, we recall the axiomatization of utility presented in Herstein and Milnor (1953). Following Herstein and Milnor (1953), a set M is a mixture set if for any elements $a, b \in M$, and for any number $p \in [0,1]$, we can associate another element of M denoted by $p \circ a \oplus (1-p) \circ b$ called a lottery between a and b. (Henceforth, the letters p and q will be reserved for elements of $[0,1]$.) We assume that the lotteries have the following properties for all $a, b \in M$:

$$\begin{gathered} 1 \circ a \oplus 0 \circ b = a, \quad p \circ a \oplus (1-p) \circ b = (1-p) \circ b \oplus p \circ a, \quad \textit{and} \\ q \circ (p \circ a \oplus (1-p) \circ b) \oplus (1-q) \circ b = pq \circ a \oplus (1-pq) \circ b \end{gathered} \tag{3.1}$$

The latter assumption is generally called "reduction of compound lotteries."

Definition 19 : *A binary relation, $\succeq$, defined on a set M is said to be:*

- *Reflexive if for all $a \in M$, $a \succeq a$.*
- *Transitive if for all $a, b, c \in M$, if $a \succeq b$ and $b \succeq c$, then $a \succeq c$.*
- *Complete if for all $a, b \in M$, either $a \succeq b$ or $b \succeq a$.*

We write $a \succ b$ if $a \succeq b$ and $b \not\succeq a$, and $a \sim b$ if $a \succeq b$ and $b \succeq a$.

A binary relation is a preference on M if it is reflexive, transitive, and complete.

A real valued function u defined on a mixture set M is a utility function for the preference $\succeq$ if it is order preserving (i.e., if $\forall a, b \in M, u(a) > u(b)$ if and only if $a \succ b$), and if

$$u(p \circ a \oplus (1-p) \circ b) = pu(a) + (1-p)u(b). \tag{3.2}$$

That is, u satisfies the expected utility property.

If $\succeq$ is a preference ordering on a mixture set M, then the following axioms guarantee the existence of a utility function:

Axiom 1 : *For any $a, b, c \in M$, the sets $\{p \mid p \circ a \oplus (1-p) \circ b \succeq c\}$ and $\{p \mid c \succeq p \circ a \oplus (1-p) \circ b\}$ are closed.*

Axiom 2 : *If $a, a' \in M$ and $a \sim a'$, then for any $b \in M$, $\frac{1}{2} \circ a \oplus \frac{1}{2} \circ b \sim \frac{1}{2} \circ a' \oplus \frac{1}{2} \circ b$.*

The utility function is unique up to an affine transformation. For any element $x \in M$, the utility of x can be given by

$$u(x) = \frac{p_{ab}(x) - p_{ab}(r_0)}{p_{ab}(r_1) - p_{ab}(r_0)}$$

where a, b, r_1, and r_0 are elements of M such that $a \succeq x \succeq b$ and $a \succeq r_1 \succ r_0 \succeq b$, and for any $y \in M$ such that $a \succeq y \succeq b$, $p_{ab}(y)$ is defined by

$$y \sim p_{ab}(y) \circ a \oplus (1 - p_{ab}(y)) \circ b. \tag{3.3}$$

The numbers $p_{ab}(\cdot)$ are well defined, and the function $u(\cdot)$ is independent of the choice of a and b. The fixed elements r_1 and r_0 determine the origin and the scale of the utility function.

3.2.2 Utility characterization: An axiomatic foundation

In this section, we generalize the analysis in Roth (1977) to characterize the utility derived from participating in organizations with imperfect information. Some preliminary definitions will be needed.

Definition 20 : *Let $x \in X$ be a type profile. An agent $i \in N$ is active at x if $x_i \neq o_i$. If $x_i = o_i$, then we say that agent i is inactive at x. We denote by $|x| = |\{i \in N : x_i \neq o_i\}|$ the number of active agents at x.*

Definition 21 : *Let $x, a \in X$ be two type profiles and $i \in N$ an agent. We say that:*

- *a is induced by x and we denote $a \trianglelefteq x$ if $\forall j \in N$, $a_j \neq x_j \implies a_j = o_j$. So, a is induced by x if it is obtained from x by making some agents inactive. For instance we have $(o_1, x_2, o_3, ..., x_{n-1}, o_n) \trianglelefteq (x_1, x_2, x_3 ..., x_{n-1}, x_n)$.*
- *a is strictly induced by x and we denote $a \triangleleft x$ if $a \trianglelefteq x$ and $a \neq x$*
- *a is strictly induced by x via agent i and we denote $a \triangleleft_o^i x$ if $a \triangleleft x$ and $a_i = 0$.*

Definition 22 : *Let $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$ be an organization with imperfect information, $i \in N$ an agent, and $a, x \in X$ two type profiles such that $a \triangleleft_o^i x$. The marginal contribution of the agent i relative to the types profiles a and x is given by:*

$$mc(i, f, a, x) = f(x_i, a_{-i}) - f(a),$$

where $(x_i, a_{-i}) \in X$ *is the type profile in which agent* i*'s type is* x_i*, and every other agent* j*'s type is* a_j*.*

The production function of an organization with imperfect information exhibits **complementarity** if the level of output produced increases with the number of active agents, that is, for all $a, b \in X$ such that $a \trianglelefteq b$, $f(a) \leq f(b)$. This condition requires that being active is more desirable for the organization than shirking. From now on, we confine our attention to organizations with imperfect information endowed with a production function that exhibits complementarity.

For simplicity, an organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$ will be represented by its production function f. We denote the collection of such organizations by $\mathcal{E}_\sigma^X$. That means that two elements of $\mathcal{E}_\sigma^X$ only differ in their production functions.

A couple (f, i) such that $f \in \mathcal{E}_\sigma^X$ and $i \in N$ is called a **strategic position**. Therefore, the set $\mathcal{E}_\sigma^X \times N$ represents the set of strategic positions.

Definition 23 : *A **strategic utility** is a function:*

$$\begin{array}{rccc} \theta : & \mathcal{E}_\sigma^X \times N & \longrightarrow & \mathbb{R} \\ & (f, i) & \longmapsto & \theta(f, i) = \theta_i(f) \end{array},$$

$\theta(f, i)$ *represents the utility of agent* $i \in N$ *in the organization with imperfect information* $f \in \mathcal{E}_\sigma^X$*.*

A position $i \in N$ is called dummy for an organization with imperfect information $f \in \mathcal{E}_\sigma^X$ if for all $a, x \in X$ such that $a \lhd_o^i x$, $mc(i, f, a, x) = 0$. We denote by $D_i \subseteq \mathcal{E}_\sigma^X$ the class of organizations with imperfect information for which i is a dummy.

It will be convenient to define the organizations with imperfect information f_0 and f_i given for all $x \in X$ by

$$f_0(x) = 0 \quad and \quad f_i(x) = \begin{cases} 1 & if \quad x_i \neq o_i \\ 0 & if \quad x_i = o_i \end{cases}$$

In the organization with imperfect information f_0, all positions are dummies; in f_i, all positions but i are dummies.

Let π be a permutation of N, that is, a one-to-one mapping from N to itself. For any organization with imperfect information $f \in \mathcal{E}_\sigma^X$, we define the organization with imperfect information $\pi f \in \mathcal{E}_{\pi\sigma}^{\pi X}$ for all $x \in X$ by: $\pi f(\pi x) = f(x)$ and $\pi\sigma(\pi x) = \sigma(x)$, where $\pi X = \prod_{i \in N} X_{\pi(i)}$ and $(\pi x)_i = x_{\pi(i)}$ for all $i \in N$.

We will be interested in the mixture set M generated by the set $\mathcal{E}_\sigma^X \times N$ of strategic positions. Thus, M contains all lotteries of the form $p \circ (f, i) \oplus (1-p) \circ (g, j)$, where (f, i) and (g, j) are elements of $\mathcal{E}_\sigma^X \times N$. We assume that a preference relation $\succeq$ is defined on M and satisfies Axioms 1 and 2. $(f, i) \succeq (g, j)$ means that it is preferred to be at position i in the organization with imperfect information f than to be at position j in the organization with imperfect information g.

We assume the preferences to satisfy the following axioms.

Axiom 3 : *For all $i \in N$, $f \in \mathcal{E}_\sigma^X$ and for any permutation π, $(f, i) \sim (\pi f, \pi(i))$.*

Axiom 4 : *If $f \in D_i$, then $(f, i) \sim (f_0, i)$, and for every $f \in \mathcal{E}_\sigma^X$ and $i \in N$, $(f, i) \succeq (f_0, i)$ and $(f_i, i) \succ (f_0, i)$.*

Axiom 5 : *For any real number $c > 1$, and for every $f \in \mathcal{E}_\sigma^X$, $i \in N$,*

$$(f, i) \sim \frac{1}{c} \circ (cf, i) \oplus (1 - \frac{1}{c}) \circ (f_0, i).$$

Axiom 3 merely says that the names of the positions do not determine their desirability in an organization with imperfect information. Axiom 4 says that being at any position in any organization with imperfect information is at least as desirable as being a dummy in any organization with imperfect information and that there is some strategic position, namely (f_i, i), which is strictly preferable to being a dummy. Axiom 5 says that if two strategic positions are identical except for the fact that the utility obtainable in one is a positive multiple of the utility obtainable in the other, then the first is indifferent to the appropriate gamble between the second, and the prospect of receiving zero.

A preference relation that satisfies axioms 1 through 5 is said to be **regular**.

We are now ready to define a strategic utility representing the preference relation $\succeq$ defined over the mixture set generated by strategic positions. Such a function exists since the preference $\succeq$ satisfies Axioms 1 and 2 by assumption.

The strategic utility of an organization with imperfect information $f \in \mathcal{E}_\sigma^X$ is the vector $\theta(f) = (\theta_1(f), \theta_2(f), \ldots, \theta_n(f))$, where

$$\theta_i(f) \equiv \theta(f, i) = \frac{p_{ab}(f, i) - p_{ab}(r_0)}{p_{ab}(r_1) - p_{ab}(r_0)}$$

for probabilities $p_{ab}(\cdot)$ as in Equation 3.3 and for $a, b, r_1, r_0 \in M$ such that $a \succeq (f, i) \succeq b$, and $a \succeq r_1 \succ r_0 \succeq b$. Fixing $r_1 = (f_i, i)$ and $r_0 = (f_0, i)$, we get $\theta(f_i, i) = 1$ and $\theta_i(f_0, i) = 0$. Axiom 4 insures that we can always take $b = r_0$, so that $p_{ab}(r_0) = 0$ for all $a \in M$.

The following lemmas are required.

Lemma 3 : *For any permutation* π*, and for every strategic position* $(f,i) \in \mathcal{E}_\sigma^X \times N$*,*

$$\theta_i(\pi f) = \theta_{\pi(i)}(f)$$

The proof follows from the order-preserving properties of utility functions, and from Axiom 3.

Lemma 4 : *For any real number* $c \geq 0$*, and for every strategic position* $(f,i) \in \mathcal{E}_\sigma^X \times N$*,*

$$\theta_i(cf) = c\theta_i(f)$$

Proof : *Let us first assume that* $c \geq 1$*. We distinguish two cases.*
Case 1: $(cf,i) \succeq (f_i,i) = r_1$*. Let us take* $a = (cf,i)$ *and* $b = r_0 = (f_0,i)$*. Then*

$$\theta_i(cf) = \frac{p_{ab}((cf,i))}{p_{ab}(r_1)} = \frac{1}{p_{ab}(r_1)}.$$

But by Axiom 5, $(f,i) \sim \frac{1}{c} \circ (cf,i) \oplus (1-\frac{1}{c}) \circ (f_0,i)$*, so* $p_{ab}((f,i)) = \frac{1}{c}$*. Therefore,* $\theta_i(f) = \frac{p_{ab}((f,i))}{p_{ab}(r_1)} = \frac{1}{c}\theta_i(cf)$*.*
Case 2: $r_1 = (f_i,i) \succeq (cf,i)$*. Let us take* $a = r_1$*,* $b = r_0$*. Then* $p_{ab}(r_1) = 1$*, and so* $\theta_i(cf) = p_{ab}(cf,i)$*. But* $(f,i) \sim \frac{1}{c} \circ (cf,i) \oplus (1-\frac{1}{c}) \circ r_0 \sim \frac{1}{c} \circ [p_{ab}((cf,i)) \circ a \oplus (1-p_{ab}((cf,i))) \circ b] \oplus (1-\frac{1}{c}) \circ b$ *by definition of* $p_{ab}(\cdot)$*. But by Equation 3.1, the latter expression is equal to* $\frac{1}{c}p_{ab}((cf,i)) \circ a \oplus (1-\frac{1}{c}p_{ab}((cf,i))) \circ b$*. Hence,* $\theta_i(f) = p_{ab}((f,i)) = \frac{1}{c}p_{ab}((cf,i)) = \frac{1}{c}\theta_i(cf)$*.*

On the other hand, if $c \in [0,1]$*, then* $\frac{1}{c} \geq 1$*, and from the first part, it follows that* $\theta_i(f) = \theta_i(\frac{1}{c}(cf)) = \frac{1}{c}\theta_i(cf)$*. This implies* $\theta_i(cf) = c\theta_i(f)$*.* ■

From Lemma 4, we have that $\theta_i(cf_i) = c$ for all $c \geq 0$ and any strategic position $(cf_i,i) \in \mathcal{E}_\sigma^X \times N$. In order to evaluate θ for other elements of the mixture set M (i.e., for organizations with imperfect information not of the form cf_i) we must investigate the risk posture of the preference relation $\succeq$.

We consider two kinds of risk: **ordinary risk** and **strategic risk**. Ordinary risk involves the uncertainty arising from the chance mechanism involved in lotteries, while strategic risk involves the uncertainty arising from the interaction in an organization with imperfect information of strategic agents.

Definition 24 : *The preference* $\succeq$ *is adverse to strategic risk if for every* $x \in X$ *and all* $i \in N$ *such that* $x_i \neq 0$*,* $(f_i,i) \succ (r_x f_x, i)$*, where* $f_x \in \mathcal{E}_\sigma^X$ *is defined for all* $y \in X$ *by:* $f_x(y) =$

$\begin{cases} 1 & if \quad x \trianglelefteq y \\ 0 & if \quad x \ntrianglelefteq y \end{cases}$, *and* $r_x = \frac{|x|}{\sum_{y \in X, x \trianglelefteq y} \sigma(y)}$. *This means that it is preferable to receive a utility of one for certain (in an organization with imperfect information with no other strategic agents) than to negotiate how to distribute a utility of* r_x *among* $|x|$ *agents.*

If the preference is reversed, we say it is risk preferring to strategic risk.

The preference relation $\geq$ *is* ***neutral to strategic risk*** *if for all* $x \in X$, *and every* $i \in N$ *such that* $x_i \neq 0$,

$$(f_i, i) \sim (r_x f_x, i) \tag{3.4}$$

To illustrate, consider an individual who owns a business that is not working very well. Assume the individual has the possibility to develop or expand her business by making a partnership with new shareholders. If she chooses not to expand her business, then we say that she is averse to strategic risk; if she chooses to expand her business, then we say that she is prone to strategic risk; and if she is indifferent between the two options, then we say that she is neutral to strategic risk.

Definition 25 : *The preference* $\geq$ *is adverse to ordinary risk if for all* $i \in N$, *and* $f, g \in \mathcal{E}_\sigma^X$, $((pf+(1-p)g), i) \succ p \circ (f,i) \oplus (1-p) \circ (g,i)$, *i.e., if it is preferable to be in the organization with imperfect information* $(pf + (1-p)g)$ *than to be in the organization with imperfect information* f *with probability* p *and in the organization with imperfect information* g *with probability* $(1-p)$.

The preference relation $\geq$ *is* ***neutral to ordinary risk*** *if for all* $i \in N$, *and* $f, g \in \mathcal{E}_\sigma^X$,

$$((pf + (1-p)g), i) \sim p \circ (f, i) \oplus (1-p) \circ (g, i). \tag{3.5}$$

To illustrate, consider a working individual who has a permanent position in his company. In the meantime, assume that two other companies call for job applications in comparatively similar positions. Assume that the individual has a probability p to be hired in the first company and a probability $1-p$ to be hired in the second company. If the individual decides not to apply to any company, then we say that the individual is adverse to ordinary risk. If he decides to apply to either of the two companies then we say that the individual is prone to ordinary risk. If he is totally indifferent between the two options then we say that he is neutral to ordinary risk.

We can now prove the following lemmas.

Lemma 5 : *If the preference relation* $\geq$ *is neutral to strategic risk, then for all* $x \in X$ *and* $i \in N$

$$\theta_i(f_x) = \begin{cases} \frac{1}{r_x} & if \quad x_i \neq o_i \\ 0 & if \quad x_i = o_i \end{cases}$$

Proof : *Let* $x \in X$ *and* $i \in N$. *Assume* $x_i = o_i$. *Then for all* $a, y \in X$ *such that* $a \lhd_o^i y$, *we have* $x \unlhd a \Leftrightarrow x \unlhd (y_i, a_{-i})$. *This implies* $mc(i, f_x, a, y) = 0$ *for all* $a, y \in X$ *such that* $a \lhd_o^i y$. *Therefore,* $f_x \in D_i$ *and* $\theta_i(f_x) = 0$ *by Axiom 4. If* $x_i \neq o_i$, *then from Equation 3.4 and Lemma 4 we have* $r_x\theta_i(f_x) = \theta_i(r_x f_x, i) = \theta_i(f_i) = 1$. *Hence,* $\theta_i(f_x) = \frac{1}{r_x}$. ∎

Lemma 6 : *If the preference relation* $\succeq$ *is neutral to ordinary risk, then for all* $f, g \in \mathcal{E}_\sigma^X$ *and* $i \in N$

$$\theta_i(f + g) = \theta_i(f) + \theta_i(g).$$

Proof : *Let* $f, g \in \mathcal{E}_\sigma^X$ *and* $i \in N$. *By Lemma 4 we have* $\theta_i(f+g) = \theta_i(2(\frac{1}{2}f+\frac{1}{2}g)) = 2\theta_i(\frac{1}{2}f+\frac{1}{2}g)$. *But by Equations 3.5 and 3.2 we have that* $\theta_i(\frac{1}{2}f+\frac{1}{2}g) = \theta_i(\frac{1}{2}\circ(f,i)\oplus\frac{1}{2}\circ(g,i)) = \frac{1}{2}\theta_i(f)+\frac{1}{2}\theta_i(g)$. *Hence,* $\theta_i(f + g) = \theta_i(f) + \theta_i(g)$. ∎

We are now ready to state our main result.

Theorem 11 : *A regular preference relation is neutral to both ordinary and strategic risk if and only if it can be represented by the strategic utility* $\mathcal{W}$ *defined for any strategic position* $(f, i) \in \mathcal{E}_\sigma^X \times N$ *by :*

$$\mathcal{W}_i(f) = \sum_{x \in X} \sigma(x) Sh_i(f, x) \tag{3.6}$$

where for all $x \in X$,

$$Sh_i(f, x) = \sum_{a \lhd_o^i x} \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} mc(i, f, a, x)$$

For clarity in the exposition, the proof of Theorem 1 and all the subsequent results are provided in the Appendix.

It can be noticed that the function $Sh_i(f, x)$ is the average of agent i's marginal contributions to the output $f(x)$ generated by the profile x. It can be easily shown that $Sh_i(f, x)$ generalizes the classical Shapley value (Shapley, 1953). In fact, to obtain the classical Shapley value, one only has to assume that each agent's type set is the pair $\{0, 1\}$, where o is the unproductive type and 1 is the productive type. In this case, the classical Shapley value is simply $Sh_i(f, \mathbf{1})$ where $\mathbf{1} = (1, 1, ..., 1)$. For this reason, $Sh_i(f, x)$ will be called the Shapley value at (f, x). Moreover, taking the expectation of the function $Sh_i(f, x)$ over the set of possible type profiles

x yields the function $\mathcal{W}$, which can then be called the ex-ante Shapley value since this is the utility that agent i expects to obtain from participation in the organization f prior to the realization of all agents' types.

Theorem 11 generalizes the result of Roth (1977), which uniquely characterizes, in terms of posture toward risk, the utility players derive from playing a game. In fact, the utility players derive from a TU game (N, v) coincides with the utility they derive from the organization with imperfect information $\langle N, X, o, f_v, \sigma\rangle$, where $X_i = \{0,1\}$ for all $i \in N$, $\sigma(x) = \begin{cases} 1 & \text{if} \quad x_i = 1\ \forall i \in N \\ 0 & \text{otherwise} \end{cases}$ and $f_v(x) = v(\{i \in N : x_i = 1\})$ for all $x \in X$. This observation is valid provided that the preference relation $\succeq$ is regular and neutral to both ordinary and strategic risks.

To parallel our result with that of the expected utility theory (Herstein and Milnor, 1953), observe that a preference relation that is regular and neutral to ordinary risk can be represented by a cardinal utility function that is unique up to affine transformations. The neutrality to strategic risk allows to pin down a unique cardinal utility out of those infinitely many utility functions representing the preference relation, and this utility is nothing but the ex-ante Shapley value.

We illustrate our main result using the following examples. These show that each agent may have a different set of types.

Example 28 : *Consider an agrarian economy with two agents, a landowner (agent 1) who provides the land and a laborer (agent 2) who supplies labor. The type set of the landowner is determined by the quality of his land. For simplicity, we assume that there are two possible qualities for the land, bad and good denoted respectively by a_1 and a_2. So, $X_1 = \{a_1, a_2\}$. Land quality is subject to uncertainty characterized by a probability distribution σ_1 over X_1. The land quality is bad with probability $\frac{1}{5}$ and good with probability $\frac{4}{5}$. The type set of the laborer is determined by his health status. For simplicity, we assume that there are three possible health statuses: poor, good, and excellent denoted respectively by b_1, b_2 and b_3. So, $X_2 = \{b_1, b_2, b_3\}$. The laborer's health status is poor, good, and excellent with probability $\frac{3}{10}, \frac{1}{2}$ and $\frac{1}{5}$, respectively. We denote this probability distribution by σ_2. We assume that land quality and the laborer's health status are independent. The production function f of the economy maps each profile of agents' conditions into an output measuring the monetary value of the yield of a suitable crop.*

It is defined for any $x \in X = X_1 \times X_2$ *by:*

$$f(x) = \begin{cases} 0 & if \quad x = (a_1, b_1) \\ 50 & if \quad x \in \{(a_1, b_2), (a_2, b_1)\} \\ 55 & if \quad x = (a_1, b_3) \\ 65 & if \quad x = (a_2, b_2) \\ 100 & if \quad x = (a_2, b_3) \end{cases} \qquad \textit{(units of \$ 1000)}$$

This agrarian economy can be modeled as an organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$*, where* N, X, f*, and* σ *are described as above and the unproductive type profile* $o = (a_1, b_1)$*. The ex-ante Shapley value of each agent* $i = 1, 2$ *is defined as:*

$$_i(f) = \sum_{x \in X} \sigma(x)\, Sh_i(f, x)$$

where for all $x \in X$, $Sh_1(f, x) = \frac{1}{2}\left(f(x_1, x_2) - f(a_1, x_2) + f(x_1, b_1) - f(a_1, b_1)\right)$ *and* $Sh_2(f, x) = \frac{1}{2}\left(f(x_1, x_2) - f(x_1, b_1) + f(a_1, x_2) - f(a_1, b_1)\right)$*. Table 3.1 provides the detailed information needed to compute the ex-ante Shapley value of each agent of the economy. We deduce that* $\mathcal{W}_1(f) = \$\ 32600$ *and* $\mathcal{W}_2(f) = \$\ 28600$*. It can be concluded that the landowner enjoys a higher utility than the laborer.*

Probability $\sigma(x_1, x_2)$

x_1 \ x_2	b_1	b_2	b_3
a_1	0.06	0.1	0.04
a_2	0.24	0.4	0.16

Technology $f(x_1, x_2)$

x_1 \ x_2	b_1	b_2	b_3
a_1	0	50	55
a_2	50	65	100

Shapley value $(Sh_1(f, x), Sh_2(f, x))$

x_1 \ x_2	b_1	b_2	b_3
a_1	(0, 0)	(0, 50)	(0, 55)
a_2	(50, 0)	(32.5, 32.5)	(47.5, 52.5)

Table 3.1: Information needed to compute $\mathcal{W}_1(f)$ and $\mathcal{W}_2(f)$.

Example 29 : *Consider a financial market with two banks, 1 and 2, that are interconnected. For simplicity, assume the two banks only invest in a single risky asset and that their investment decisions boil down to three types:* L_i, M_i*, and* H_i *corresponding respectively to a low, medium, and high level of investment for bank* $i = 1, 2$*. So, the type set of each bank is given by* $X_1 = \{L_1, M_1, H_1\}$ *and* $X_2 = \{L_2, M_2, H_2\}$*. The investment decisions* L_1, M_1*, and* H_1 *of Bank 1 are taken with probability* σ_1 *equal respectively to* 0.25*,* 0.5*, and* 0.25*. The investment decisions* L_2, M_2*, and* H_2 *of Bank 2 are taken with probability* σ_2 *equal respectively to* 0.1*,* 0.2*, and* 0.7*. The investment decisions of Bank 1 are independent of those of Bank 2, and vice versa. A function* f *defined on* $X = X_1 \times X_2$ *measures the system-wide risk associated to each profile*

$x \in X$ of banks' investments decisions as follows:

$$f(x) = \begin{cases} 0 & if \quad x = (L_1, L_2) \\ 0.25 & if \quad x \in \{(L_1, M_2), (M_1, L_2)\} \\ 0.5 & if \quad x \in \{(L_1, H_2), (H_1, L_2), (M_1, M_2)\} \\ 0.75 & if \quad x \in \{(H_1, M_2), (M_1, H_2)\} \\ 1 & if \quad x = (H_1, H_2) \end{cases}$$

This financial market can be modeled as an organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$, where N, X, f, and σ are described as above and the unproductive type profile $o = (L_1, L_2)$. The ex-ante Shapley value of each bank $i = 1, 2$ is defined as:

$$\mathcal{W}_i(f) = \sum_{x \in X} \sigma(x)\, Sh_i(f, x)$$

Table 3.2 provides the detailed information needed to compute the ex-ante Shapley value of each bank. We deduce that $\mathcal{W}_1(f) = 0.25$ and $\mathcal{W}_2(f) = 0.4$. It can be concluded that Bank 2 contributes more to the system-wide risk than Bank 1.

Probability $\sigma(x_1, x_2)$

		x_2		
		L_2	M_2	H_2
x_1	L_1	0.025	0.05	0.175
	M_1	0.05	0.1	0.35
	H_1	0.025	0.05	0.175

Technology $f(x_1, x_2)$

		x_2		
		L_2	M_2	H_2
x_1	L_1	0	0.25	0.5
	M_1	0.25	0.5	0.75
	H_1	0.5	0.75	1

Shapley value $(Sh_1(f,x), Sh_2(f,x))$

		x_2		
		L_2	M_2	H_2
x_1	L_1	(0, 0)	(0, 0.25)	(0, 0.5)
	M_1	(0.25, 0)	(0.25, 0.25)	(0.25, 0.5)
	H_1	(0.5,0)	(0.5, 0.25)	(0.5, 0.5)

Table 3.2: Information needed to compute $\mathcal{W}_1(f)$ and $\mathcal{W}_2(f)$.

3.2.3 Rationalization under Roth's risk neutrality

In an organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$, we have so far assumed that the probability distribution σ is given. This is a legitimate assumption as X represented the set of agents' type profiles. In this section, we assume that each agent's type uniquely determines the action chosen by the agent and interpret X as the set of action profiles. We are interested in knowing whether there exists a probability distribution σ that maximizes agents' preferences when they are neutral to both ordinary and strategic risks. In other words, if agents could choose the uncertainty characterizing their participation in an organization with imperfect information, which probability profile σ would they choose assuming Roth's risk neutrality?[7]

We first consider a tuple $\langle N, X, o, f\rangle$, and we associate to it a finite strategic form game $\langle N, (S_i)_{i\in N}, (u_i)_{i\in N}\rangle$, where for each $i \in N$, S_i is the set of all probability distributions over X_i, and u_i is the utility function of player i defined on $S = \prod_{i\in N} S_i$. S_i represents the set of strategies of player i, and S represents the set of the strategy profiles of agents. Given Roth's risk neutrality, each agent derives utility from a strategy profile $s = (s_1, s_2, ..., s_n) \in S$ given by the ex-ante Shapley value associated to the organization with imperfect information $\langle N, X, o, f, s\rangle$. Therefore, the utility of agent i from $s \in S$ is defined by:

$$u_i(s) = \sum_{x\in X} s(x) Sh_i(f, x) = \sum_{x\in X} s(x) \left\{ \sum_{a \lhd_o^i x} \frac{(|a|)!(|x| - |a| - 1)!}{(|x|)!} \left[f(x_i, a_{-i}) - f(a) \right] \right\}$$

where, for each $x \in X$, $s(x) = \prod_{i\in N} s_i(x_i)$.

Definition 26 : *Let $\mathcal{E} = \langle N, X, o, f, s\rangle$ be an organization with imperfect information. We say that $\mathcal{E}$ is rationalizable if s is a Nash equilibrium of the game $\langle N, (S_i)_{i\in N}, (u_i)_{i\in N}\rangle$: for all $i \in N$, $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$, for all $s_i' \in S_i$.*

We have the following result, which states that there always exists a rationalizable profile of probability distributions in any organization.

Theorem 12 : *For any set of agents N, any set of action profile X, and any real-value function f defined on X, there always exists a probability distribution σ over X such that the organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma\rangle$ is rationalizable.*

Since the game associated to an organization is finite, the proof of this result simply follows from the proof of the existence of a mixed Nash equilibrium in finite games (Nash, 1951). Interestingly, a rationalizable organization where the probability distribution σ is a *pure strategy Nash equilibrium* always exists. We will prove this result later when we show how our theory applies to the firm. The following example illustrates this point.

Example 30 : *Consider the tuple $\langle N, X, o, f\rangle$, where $N = \{1, 2\}$, $X_1 = \{a_0, a_1\}$, $X_2 = \{b_0, b_1\}$, $o = (a_0, b_0)$, $X = X_1 \times X_2$, and the production function is given by $f(a_0, b_0) = 0, f(a_0, b_1) = -1, f(a_1, b_0) = -\frac{1}{2}$, and $f(a_1, b_1) = \frac{5}{2}$. Under Roth's risk neutrality, the strategic form game*

[7] By Roth's risk neutrality, we mean our extensions of neutrality to strategic risk and ordinary risk introduced by Roth (1977).

associated to this organization is the tuple $\langle N, (S_i)_{i\in N}, (u_i)_{i\in N}$, *where* $S_1 = S_2 = \{(p_1, p_2) \in \mathbb{R}^2_+ : p_1 + p_2 = 1\}$, *and for all* $s_1 = (p, 1-p) \in S_1$ *and* $s_2 = (q, 1-q) \in S_2$, *we have:*

$$u_i(s_1, s_2) = pqSh_i(f, (a_0, b_0)) + p(1-q)Sh_i(f, (a_0, b_1)) + (1-p)qSh_i(f, (a_1, b_0)) + (1-p)(1-q)Sh_i(f, (a_1, b_1)).$$

For all $x \in X$, *the payoffs* $(Sh_1(f, x), Sh_2(f, x)$ *are given in Table 3.3.*[8]

		agent 2: b_0	agent 2: b_1
agent 1	a_0	$(0, 0)$	$(0, -1)$
	a_1	$(-\frac{1}{2}, 0)$	$(\frac{3}{2}, 1)$

Table 3.3: The agents' payoffs $(Sh_1(f, x), Sh_2(f, x))$ for all $x \in X$.

To find the Nash equilibria, given $q \in [0, 1]$, *agent 1 solves the problem:*

$$\begin{array}{ll} \underset{p}{\textit{Maximize}} & -\frac{1}{2}q(1-p) + \frac{3}{2}(1-q)(1-p) \\ \textit{subject to} & 0 \le p \le 1 \end{array},$$

and given $p \in [0, 1]$, *agent 2 solves the problem*

$$\begin{array}{ll} \underset{q}{\textit{Maximize}} & -p(1-q) + (1-p)(1-q) \\ \textit{subject to} & 0 \le q \le 1 \end{array}.$$

A Nash equilibrium is equivalent to a couple $(p^*, q^*) \in [0, 1]^2$ *that simultaneously solves the above maximization problems. The resolution of these problems yields three solutions. The first solution is* $p^* = q^* = 0$, *which is equivalent to agent 1 and agent 2 choosing respectively actions* a_1 *and* b_1 *with certainty. The second solution is* $p^* = q^* = 1$, *which means that agent 1 and agent 2 respectively choose actions* a_0 *and* b_0 *with certainty. The last solution is* $p^* = \frac{1}{2}$ *and* $q^* = \frac{3}{4}$, *which corresponds to agent 1 choosing the action* a_0 *with probability* $\frac{1}{2}$ *and agent 2 choosing the action* b_1 *with probability* $\frac{3}{4}$. *There are, therefore, three rationalizable profiles of probability distributions in this setting:* $((0, 1), (0, 1)), ((1, 0), (1, 0))$ *and* $((\frac{1}{2}, \frac{1}{2}), (\frac{3}{4}, \frac{1}{4}))$. *This gives rise to three rationalizable organizations associated with the tuple* $\langle N, X, o, f\rangle$.

[8] $s_1 = (p, 1-p)$ means that agent 1 plays the action a_0 with probability p and the action a_1 with probability $1-p$. Likewise, $s_2 = (q, 1-q)$ means that agent 2 plays the action b_0 with probability q and the action b_1 with probability $1-q$.

3.3 Some applications

This section provides two applications of our model. In the first application, we consider the informational problem associated with workers' costs in a firm. We show that, under Roth's risk neutrality, workers' net utilities do not change when the firm has information on workers' private costs and decides to pool them and it does not have this information. The availability of cost information, therefore, does not affect workers' incentives, and so it may be better to keep that information private if acquiring it is costly. In the second application, we consider the classical problem of assigning indivisible goods in a two-sided buyer-seller market.

3.3.1 Private versus common information in the firm: Pooling versus separating cost

Consider a firm consisting of a finite set of workers, $N = \{1, 2, ..., n\}$, who possibly differ in their productivity and ability. Each worker $i \in N$ supplies an effort level $x_i \in X_i = \{0, 1, ..., \overline{x_i}\}$ and incurs a cost $c_i(x_i) \in \mathbb{R}$. Each profile of workers' effort levels $x = (x_1, x_2, ..., x_n) \in X = \prod_{i \in N} X_i$ is converted by a technology F into a monetary output $F(x) \in \mathbb{R}$ of which a share α is distributed among workers as salary. The remaining share, $1 - \alpha$, can be thought of as being distributed among the firm's shareholders or invested in risky and safe assets.[9] The size or use of the latter share is inconsequential for our analysis.

Pooling cost versus separating cost

There are two potential scenarios based on the cost information that is available to the firm. When the firm has full information about the costs incurred by workers, it may decide to pool all the individual costs prior to sharing the firm's surplus among the workers; this scenario is referred to as the "pooling cost" scenario. When that information is not available to the firm, the firm can only share the revenue as salary, and each worker consumes his salary and incurs his own cost; this scenario is referred to as the "separating cost" scenario.

Pooling cost

Under the assumption of pooling cost, the firm incurs all the costs of its workers. Therefore, it can be modeled as an organization with perfect information $\mathcal{E} = \langle N, X, o, f^p \rangle$, where the

[9]We remark that this model can also be thought of as a model of a *joint venture*, where individuals in N are the parties forming the venture. In this case, $F(x)$ can be thought of as the profit, which is entirely redistributed among the different parties.

inaction profile is the profile of nil effort level, $o = (0, 0, ..., 0)$, and the surplus function is $f^p(x) = \alpha F(x) - \sum_{i \in N} c_i(x_i)$ for all $x \in X$. Consistent with our theoretical model, the net utility a potential worker derives at a given position $i \in N$ in the firm when an effort profile x is supplied is given by:

$$u_i(x) = Sh_i(f^p, x) \tag{3.7}$$

Separating cost

Under the separating cost assumption, each worker incurs his own cost. Therefore, the firm can be modeled as the organization with perfect information $\mathcal{E} = \langle N, X, o, f^s \rangle$, where the inaction profile is the profile of nil effort level, $o = (0, 0, ..., 0)$, and the production function is $f^s(x) = \alpha F(x)$ for all $x \in X$. Consistent with our theoretical model, the utility a potential worker derives at a given position $i \in N$ in the firm when an effort profile x is supplied is equal to $Sh_i(f^s, x)$. It follows that his net utility is given by:

$$u_i(x) = Sh_i(f^s, x) - c_i(x_i) \tag{3.8}$$

Interestingly, both approaches result in the same outcome from the perspective of workers, as demonstrated by the following proposition. This is essentially attributed to the appealing properties of Roth's risk neutrality that characterize the Shapley value.

Proposition 2 : *If a potential worker has regular preferences towards various positions within a firm that are neutral to both ordinary and strategic risks, then his net utility at a specific position within the firm, evaluated at a given effort profile, will be the same regardless of whether the cost is pooled or separated.*

This result follows immediately from the definition of the Shapley value and is therefore omitted.

The remaining part of this section deals with the separating cost assumption as it is consistent with most real-life situations.

Assuming each worker $i \in N$ is a rational agent that chooses a level of effort in X_i that maximizes his net utility as defined by equation 3.8, what level of effort will each worker choose? Theorem 12 guarantees the existence of a mixed strategy Nash equilibrium, a probability distribution over X_i, that rationalizes worker i's behavior. While randomizing over a set of actions is interesting and is often observed in practice, it might be more interesting if each worker can find it rational to choose an action with certainty. A refined solution is therefore that of a pure strategy equilibrium which consists of each worker choosing with certainty the action that maximizes his utility. Unlike mixed strategy Nash equilibrium, a pure strategy Nash equilibrium doesn't always exist when the sets of players' strategies are finite. Nonetheless, the following proposition unveils a class of finite strategic form games that always admits a pure strategy Nash equilibrium under Roth's risk neutrality.

Proposition 3 : *Let $\mathcal{E} = \langle N, X, o, f^s \rangle$ be a tuple as defined above. Then, the strategic form game $\langle N, (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where the payoff functions are defined by Equation 3.8 admits a pure strategy Nash equilibrium.*

Remark that an immediate implication of this proposition is that for any set of agents N, any set of action profile X, and any real-value function f defined on X, there always exists a probability distribution σ over X such that the organization with imperfect information $\mathcal{E} = \langle N, X, o, f, \sigma \rangle$ is rationalizable, where σ is a **pure strategy Nash equilibrium**. This is because in the proposition above, for each agent, the cost of any action can be set to zero, and our result will still go through.

To illustrate this result, consider a firm consisting of two workers 1 and 2 who respectively supply effort levels $x_1 \in \{0, 1, 4\}$ and $x_2 \in \{0, 4, 9\}$ measured by the number of tasks they can perform. Workers have different skills for different tasks on which the firm needs expertise. Worker 2 is the most equipped in terms of the number of tasks he can perform, but this makes him much less efficient compared to Worker 1. As a result, the workers' cost functions are given by $c_1(x_1) = 3x_1$ and $c_2(x_2) = 2x_2^2$. The firm's technology is represented by the constant elasticity of substitution production function

$$F(x_1, x_2) = 100\left(0.4x_1^{\frac{1}{2}} + 0.6x_2^{\frac{1}{2}}\right)^2.$$

Assume that $\alpha = \frac{1}{2}$, that is an equal share of the output produced is distributed among the workers and the shareholders. Therefore, the utility a potential worker, with preferences consistent with Theorem 11, derives from a given position $i = 1, 2$ when an effort profile $x \in X = \{0, 1, 4\} \times \{0, 4, 9\}$ is supplied, is given by:

$$u_i(x) = \frac{1}{2} Sh_i(F, x) - c_i(x_i).$$

Table 3.4 provides the information needed to compute the utility of each worker.

Technology $f(x_1, x_2)$

		$x_2 = 0$	$x_2 = 4$	$x_2 = 9$
x_1	0	0	40	0
	1	5	93	77
	4	20	156	164

Workers' utilities $(u_1(x), u_2(x))$

		$x_2 = 0$	$x_2 = 4$	$x_2 = 9$
x_1	0	(0, 0)	(0, 40)	(0, 0)
	1	(5, 0)	(29, 64)	(41, 36)
	4	(20,0)	(68, 88)	(92, 72)

Table 3.4: Firm's technology and workers' utilities.

As a result, there is only one pure strategy Nash equilibrium, which is (4, 4). This equilibrium is achieved when worker 1 performs four tasks and receives a payoff of 68, and worker 2 also performs four tasks but receives a payoff of 88. Let π^* be the probability distribution defined on X by: $\pi^*(x) = \begin{cases} 1 & if \quad x = (4,4) \\ 0 & \text{otherwise} \end{cases}$. It follows that the organization with imperfect information $\langle N, X, o, f^p, \pi^* \rangle$ is rationalizable. Furthermore, the utility a potential worker, with preferences consistent with Theorem 11, derives from a given position $i = 1, 2$ within the firm is given by $\mathcal{W}_1(f^p) = 68$ and $\mathcal{W}_2(f^p) = 88$.

3.3.2 Buyer-seller assignment market: Pricing, optimality, and utility achievability

Consider an economy with a finite set of agents, $N = S \cup B$, partitioned into two disjoint sets of equal size m: the set of sellers (S) and the set of buyers (B). Each seller $i \in S$ owns an indivisible object evaluated at $c_i \in \mathbb{R}$ from his perspective, and all objects belong to the same family of goods, though they might differ in a number of characteristics.[10] Each buyer $j \in B$ wants to acquire one of the objects, and his valuation for each of the m objects is given by $v_j = (v_{ij})_{i \in S} \in \mathbb{R}^m$. An assignment market is defined as a tuple $\mathcal{A} = (S \cup B, (c_i)_{i \in S}, (v_j)_{j \in B})$. Given an assignment market $\mathcal{A}$, a natural question that arises is how to assign the objects to buyers such that each buyer has exactly one object. This problem, known as the assignment problem (Shapley and Shubik, 1971), has been widely studied in the literature (see, e.g., Kelso Jr and Crawford (1982), Demange, Gale, and Sotomayor (1986), Kranton and Minehart (2001), and Azevedo and Leshno (2016)). Below, we model an assignment problem as an organization with imperfect information. Before that, we need to recall some definitions. Let $\mathcal{A} = (S \cup B, (c_i)_{i \in S}, (v_j)_{j \in B})$ be an assignment market. An assignment or a matching

[10] For example, objects might be cars or smartphones of different brands.

is a set $M \subseteq S \times B$ such that $|M| = m$ and $(i_1, j_1), (i_2, j_2) \in M$ if and only if $i_1 \neq i_2$ and $j_1 \neq j_2$. We denote by $\mathcal{M}$ the set of all matchings. The surplus of a matching M is defined as $s(M) = \sum_{(i,j)\in M} s_{ij}$, where $s_{ij} = max(0, v_{ij} - c_i)$ for all $(i,j) \in M$. An **optimal matching** is a matching with the maximum surplus, that is, M is an optimal matching if $s(M) = \max_{L\in\mathcal{M}} s(L)$. We denote by $\mathcal{M}^*$ the set of optimal matchings.

Any assignment market $\mathcal{A} = (S \cup B, (c_i)_{i\in S}, (v_j)_{j\in B})$ can be modeled as an organization with imperfect information $\mathcal{E}^{\mathcal{A}} = \langle N, X, o, f, \sigma \rangle$, where $N = S \cup B$, an action of a seller consists of choosing a buyer with whom he wants to trade, or no buyer at all if he deems the trade less likely after a haggling process. So the action set of a seller $i \in S$ is $X_i = B \cup \{o_i\}$, where o_i is the no-trade action. Similarly, an action of a buyer consists of choosing a seller with whom he wants to trade, or no seller at all, so $X_j = S \cup \{o_j\}$ for all $j \in S$. A trade occurs between a seller and a buyer when both choose each other. The set $X = \prod_{i\in N} X_i$ represents the set of trading profiles. The **surplus function** f is defined for all $x \in X$ by:

$$f(x) = \sum_{(i,j)\in S\times B} \delta_{ij}(x) s_{ij}, \text{ where } \delta_{ij}(x) = \begin{cases} 1 & \text{if } x_i = j \text{ and } x_j = i \\ 0 & \text{otherwise} \end{cases} \quad \forall (i,j) \in S \times B \tag{3.9}$$

The function σ represents a probability distribution over X; it may be inferred from past trading experiences.

Each matching $M \in \mathcal{M}$ gives rise to a unique trading profile denoted by $x(M)$ and defined for all $k \in N$ by $x_k(M) = k'$ if and only if $(k, k') \in M$.

Consistent with our model, we assume that an outsider has preferences on the different positions of an assignment market $\mathcal{A} = (S \cup B, (c_i)_{i\in S}, (v_j)_{j\in B})$; it can be a position of a buyer or a seller. If these preferences are regular and neutral to both ordinary and strategic risk, then the individual's utility at a given position $i \in N$ is given by the ex-ante Shapley value $\mathcal{W}_i(f)$, where f is defined by Equation 3.9. Assuming that only optimal matchings form, then it follows that the individual utility at position $i \in N$ is given by:

$$ESV_k(f) = \sum_{M\in\mathcal{M}^*} \sigma(x(M)) Sh_k(f, x(M)). \tag{3.10}$$

Now that we have assigned to each risk-neutral agent the utility he can expect to achieve in the two-sided economy, it will be interesting to analyze whether such utility can be supported by a **price system**. If the seller $i \in S$ decides to sell his item at the price p_i, then his net gain should be equal to $p_i - c_i$. In this case, a potential buyer $j \in B$ will be confronted by a choice among the m possible net gains $v_{ij} - p_i$, $i \in S$. If these numbers are all negatives, then buyer j will prefer to stay out of the market and consume a reservation utility of zero; otherwise he will choose to trade with the seller that maximizes his net gain, that is the seller $i_j = arg \max_{i \in S} v_{ij} - p_i$.[11]

We say that the ex-ante Shapley value is **achievable** if there exists a price system $(p_i)_{i \in S}$ such that $ESV_i(f) = p_i - c_i$ for all $i \in S$ and for every $j \in B$ there exists a unique $i_j \in S$ such that $ESV_j(f) = v_{i_j j} - p_{i_j}$. In this case, we say that the matching $\{(i_j, j) \in S \times B : j \in B\}$ achieves the ex-ante Shapley value. We derive a sufficient condition under which the ex-ante Shapley value is achievable. When there is only one optimal matching, it is easy to show that the ex-ante Shapley value can be achieved through the price system $p_i = ESV_i(f) + c_i, i \in N$. However, if there are multiple optimal matchings, the ex-ante Shapley value is only partially achievable, meaning that it is only achieved for one side of the market—the sellers' side. Partial achievability of sellers' utilities is less challenging than that of buyers' utilities. For simplicity, we focus on the former case when utilities are given by the ex-ante Shapley value. An example, among many that might exist, of partial achievability of buyers' utilities consists of the following: each seller $i \in N$ sets a price $p_i = ESV_i(f) + c_i, i \in N$ and each buyer $j \in B$ trades with the seller $i_j = arg \max_{i \in S} \min\{ESV_j(f), v_{ij} - p_i\}$. In addition, we assume that any tie is broken in favor of the formation of an optimal matching or simply a matching if the latter is not possible. Unlike full achievability, some buyers don't necessarily achieve their ex-ante Shapley value. However, the mechanism we propose minimizes the discrepancy between the expected utility and the realized utility while maximizing the utilities of those buyers. The previous discussion can be summarized by the following proposition.

Proposition 4 : *Let $\mathcal{A} = (S \cup B, (c_i)_{i \in S}$ be an assignment market. If there is a unique optimal matching, then there exists a price system, $(p_i)_{i \in S}$, that achieves the ex-ante Shapley value of all agents. However, if there are multiple optimal matchings, then the price system only achieves the ex-ante Shapley value of the sellers.*

As an illustration, consider an assignment market for cars, consisting of three car sellers, $S = \{1, 2, 3\}$, and three potential buyers, $B = \{4, 5, 6\}$. Sellers' and buyers' valuations are given by Table 3.5. We have six possible matchings, M_1 to M_6 displayed in Figure 3.1. In matching M_1, for instance, sellers 1, 2, and 3 trade respectively with buyers 4, 5, and 6. In addition, we have three optimal matchings, namely M_1, M_2, and M_4, with a maximum surplus of \$18,000 each. Table 3.6 displays the matrix (s_{ij}) of bilateral trade surplus.

[11] In case there are many sellers that provide the same net gain utility to buyer j, we assume that he will choose to trade with the seller who has the lowest price and that any other ties are broken with an arbitrary tie-breaking rule.

Car sellers (i)	Sellers' valuation (c_i)	Buyers' valuation (v_4)	(v_5)	(v_6)
1	\$25,000	\$30,000	\$27,000	\$34,000
2	\$32,000	\$31,000	\$38,000	\$40,000
3	\$40,000	\$43,000	\$45,000	\$47,000

Table 3.5: A car assignment market with three sellers and three buyers.

Assuming that optimal matchings are equally likely to be formed, the utility an individual with preferences consistent with Theorem 11 derives from a given position $i \in \{1, 2, ..., 6\}$ is given by:

$$\mathcal{W}_i(f) = \frac{1}{3}\left(Sh_i(f, x(M_1) + Sh_i(f, x(M_2) + Sh_i(f, x(M_4)))\right)$$

A simple computation show that $\mathcal{W}_1(f) = \frac{19000}{6}$, $\mathcal{W}_2(f) = \frac{20000}{6}$, $\mathcal{W}_3(f) = \frac{15000}{6}$, $\mathcal{W}_4(f) = \frac{13000}{6}$, $\mathcal{W}_5(f) = \frac{17000}{6}$, $\mathcal{W}_6(f) = \frac{24000}{6}$. It follows that such an individual would prefer to be at the position of seller 2 if he has some selling ability. Otherwise, he would prefer to be at the position of buyer 6.

The price system that supports a partial implementation of the ex-ante Shapley value is given for each seller by: $p_1 = \frac{169000}{6}$, $p_2 = \frac{212000}{6}$, and $p_3 = \frac{255000}{6}$. Given these prices, agent 4 would prefer to trade with seller 1 as $\max_{i \in S} \min\{ESV_4(f), v_{ij} - p_i\}$ is equal to $\max\{\min\{\frac{13000}{6}, \frac{11000}{6}\}, \min\{\frac{13000}{6}, -\frac{26000}{6}\}, \min\{\frac{13000}{6}, 500\}\} = \frac{11000}{6}$. Agent 5 would prefer to trade with seller 2 as $\max_{i \in S} \min\{ESV_5(f), v_{ij} - p_i\}$ is equal to
$\max\{\min\{\frac{17000}{6}, -\frac{7000}{6}\}, \min\{\frac{17000}{6}, \frac{16000}{6}\}, \min\{2500, 500\}\} = \frac{16000}{6}$. Agent 6 is indifferent between trading with either of the seller as everyone provides him with the same maxmin utility of $4000 = ESV_6(f)$. The tie-breaking rule applies and agent 6 trades with seller 3. It follows that the matching M_1 implements partially the ex-ante Shapley value following our proposed mechanism. All sellers and buyer 6 achieve their ex-ante Shapley value, agent 4 achieves a lower utility, while agent 5 achieves a higher utility.

3.4 Conclusion

In this paper, we consider the classical problem of evaluating the cardinal utility of participating in an organization with imperfect information. In such an organization, each agent is unsure about other agents' types or actions, because the latter is subject to uncertainty.

		Buyers 4	5	6
Sellers	1	5	2	9
	2	0	6	8
	3	3	5	7

Table 3.6: Bilateral trade surplus matrix (s_{ij}) (in thousands of dollars) of the car assignment market with three sellers and three buyers.

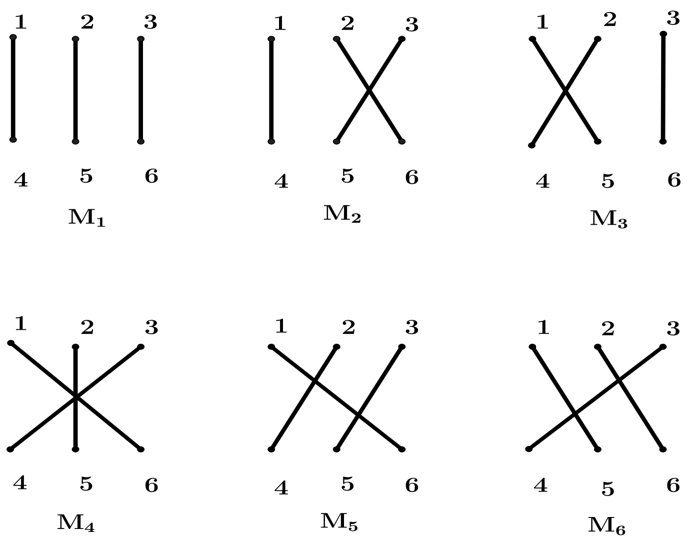

Figure 3.1: Possible matchings of an assignment market with 3 sellers and 3 buyers.

After defining a set of intuitive axioms the preferences of agents should satisfy, we show that these preferences are represented by a unique cardinal utility—called the ex-ante Shapley value—if and only if they are neutral to both strategic and ordinary risks. These notions of neutrality extend the work of Roth (1977) and Roth (2019) to a setting that formalizes a wide range of real-life environments. We also prove the existence of rationalizable organizations, where uncertainty over agents' actions arises as an equilibrium play. We then analyze the implications of Roth's risk neutrality for incentives and equilibrium behaviors. Our analysis has several interesting applications, owing to the fact that our model of an organization can be used to formalize a wide range of economic environments such as joint ventures, firms, and markets. We develop two main applications. Our first application is to a classical informational problem in the firm; we prove that the availability of cost information does not affect workers' incentives under Roth's risk neutrality. This is because the utility derived by a worker from each position in the firm does not change regardless of whether the individual costs incurred by workers are known and thus pooled (as in a joint venture) or unknown and hence incurred privately. Our second application considers two-sided assignment markets involving sellers and buyers of indivisible goods, where the valuation of the indivisible goods owned by the sellers may be different for each agent. Our theory shows how this marketplace can be evaluated by agents prior to them participating in it, and we prove the existence of a pricing mechanism supporting the agents' evaluation. In particular, we show that, under risk neutrality, there is a pricing system that achieves the utility of each seller and each buyer when the optimal matching is unique, and that only achieves the utility of sellers when there are at least two optimal matchings. These findings have important implications for market design, given the high prevalence of assignment markets in the real world.

3.5 Appendix: Proofs of Results

Proof (Theorem 11) : ***First***, *we observe that any production function $f : X \longrightarrow \mathbb{R}$ can be written as a linear decomposition of the basic functions $f_x, x \in X$. That is:*

$$f = \sum_{x\in X;\, x\neq 0} c_x(f) f_x, \quad \text{where } c_x(f) = \sum_{x'\trianglelefteq x} (-1)^{|x|-|x'|} f(x').\text{[12]} \tag{3.11}$$

Second, *let us show that for all $f, g \in \mathcal{E}_\sigma^X$ and $i \in N$,*

$$\mathcal{W}_i(f+g) = \mathcal{W}_i(f) + \mathcal{W}_i(g).$$

By definition, we have $\mathcal{W}_i(f+g) = \sum_{x\in X} \sigma(x) Sh_i(f+g, x)$, where

$$\begin{aligned} Sh_i(f+g,x) &= \sum_{a \triangleleft_o^i x} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} mc(i, f+g, a, x) \\ &= \sum_{a \triangleleft_o^i x} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} [(f+g)(x_i, a_{-i}) - (f+g)(a)] \\ &= \sum_{a \triangleleft_o^i x} \frac{(|a|)!(|x|-|a|-1)!}{(|x|)!} [f(x_i, a_{-i}) - f(a) + g(x_i, a_{-i}) - g(a)] \\ &= Sh_i(f,x) + Sh_i(g,x) \end{aligned}$$

Hence $\mathcal{W}_i(f+g) = \mathcal{W}_i(f) + \mathcal{W}_i(g)$.

Third, *let us show that for all $x \in X$ and $i \in N$, $\mathcal{W}_i(f_x) = \begin{cases} \frac{1}{r_x} & if \quad x_i \neq o_i \\ 0 & if \quad x_i = o_i \end{cases}$.*

Let $x \in X$ and $i \in N$. If $x_i = o_i$, then for all $a, y \in X$ such that $a \triangleleft_o^i y$, we have $x \trianglelefteq a \Leftrightarrow x \trianglelefteq (y_i, a_{-i})$. This implies $mc(i, f_x, a, y) = 0$ for all $a, y \in X$ such that $a \triangleleft_o^i y$. It follows that $Sh_i(f_x, y) = 0$ for all $y \in X$, i.e., $\mathcal{W}_i(f_x) = 0$.

[12] This result follows from simple computations.

Assume now $x_i \neq o_i$. *We have:*

$$\begin{aligned}
\mathcal{W}_i(f_x) &= \sum_{y\in X}\sigma(y)\sum_{a\lhd_o^i y}\frac{(|a|)!(|y|-|a|-1)!}{(|y|)!}(f_x(y_i,a_{-i})-f_x(a)) \\
&= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{a\lhd_o^i y}\frac{(|a|)!(|y|-|a|-1)!}{(|y|)!}(f_x(y_i,a_{-i})-f_x(a)) \\
&= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{a\lhd_o^i y}\frac{(|a|)!(|y|-|a|-1)!}{(|y|)!}f_x(y_i,a_{-i}) \text{ since } a_i = o_i \text{ and } x_i \neq o_i \Rightarrow x \ntrianglelefteq a
\end{aligned}$$

Let $P_x = \{j \in N : x_j \neq o_j\}$. *Let* $a, y \in X$ *such that* $x \unlhd y$ *and* $a \lhd_o^i y$, *we have that* $x \unlhd (y_i, a_{-i}) \Leftrightarrow x_j = a_j = y_j\ \forall j \in P_x \setminus \{i\}$. *This implies that* $x \unlhd (y_i, a{-}i) \Leftrightarrow |x| - 1 \leq |a| \leq |y| - 1$. *But for every* $k \in \{|x|-1, |x|, ..., |y|-1\}$, *there is exactly* $\binom{|y|-|x|}{k-|x|+1}$ *a in* X *such that* $x \unlhd (y_i, a{-}i)$ *and* $|a| = k$. *This implies,*

$$\begin{aligned}
\mathcal{W}_i(f_x) &= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{k=|x|-1}^{|y|-1}\binom{|y|-|x|}{k-|x|+1}\frac{(|a|)!(|y|-|a|-1)!}{(|y|)!}f_x(y_i,a_{-i}) \\
&= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{k=|x|-1}^{|y|-1}\frac{(|y|-|x|)!}{(k-|x|+1)!(|y|-k-1)!}\frac{(|a|)!(|y|-|a|-1)!}{(|y|)!}f_x(y_i,a_{-i}) \\
&= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{k=|x|-1}^{|y|-1}\frac{k!(|y|-|x|)!}{(k-|x|+1)!(|y|)!} \\
&= \sum_{y\in X, x\unlhd y}\sigma(y)\sum_{k=0}^{|y|-|x|}\frac{(k+|x|-1)!(|y|-|x|)!}{k!(|y|)!}
\end{aligned}$$

Let $y \in X$ *such that* $x \unlhd y$ *and set* $A(|y|) = \sum_{k=0}^{|y|-|x|}\frac{(k+|x|-1)!(|y|-|x|)!}{k!(|y|)!}$. *Since* $x_i \neq o_i$, *then* $x \unlhd y$ *implies that* $1 \leq |y| \leq n$. *Let us show by induction on* $|y|$ *that* $A(|y|) = \frac{1}{|x|}$. *If* $|y| = 1$, *then* $|x| = 1$ *and* $A(|y|) = 1 = \frac{1}{|x|}$. *Now, assume* $A(|y|) = \frac{1}{|x|}$ *for all* $|y| < n$, *and let us show that* $A(|y|+1) = \frac{1}{|x|}$. *We have that:*

$$\begin{aligned}
A(|y|+1) &= \sum_{k=0}^{|y|+1-|x|}\frac{(k+|x|-1)!(|y|+1-|x|)!}{k!(|y|+1)!} \\
&= \left(\frac{(|y|+1-|x|)(|y|-|x|)!}{(|y|+1)(|y|)!}\right)\left(\sum_{k=0}^{|y|-|x|}\frac{(k+|x|-1)!}{k!}+\frac{(|y|)!}{(|y|+1-|x|)!}\right) \\
&= \frac{|y|+1-|x|}{|x|(|y|+1)}+\frac{1}{|y|+1} \text{ by the induction assumption} \\
&= \frac{1}{|x|} \text{ as wanted.}
\end{aligned}$$

It follows that $\sum_{k=0}^{|y|-|x|} \frac{(k+|x|-1)!(|y|-|x|)!}{k!(|y|)!} = \frac{1}{|x|}$ *for all* $y \in X$ *such that* $x \trianglelefteq y$. *This implies that* $\mathcal{W}_i(f_x) = \sum_{y \in X, x \trianglelefteq y} \sigma(y)\frac{1}{|x|} = \frac{1}{r_x}$ *if* $x_i \neq o_i$. *Hence,* $\mathcal{W}_i(f_x) = \begin{cases} \frac{1}{r_x} & if \quad x_i \neq o_i \\ 0 & if \quad x_i = o_i \end{cases}$.

Finally*, let us show that* $\mathcal{W}_i(f) = \theta_i(f)$ *for all* $i \in N$ *and* $f \in \mathcal{E}_\sigma^X$. *From Equation 3.11,* f *can be written as* $f = \sum_{x \in X;\, x \neq 0} c_x(f) f_x$. *Therefore,* $\mathcal{W}_i(f) = \sum_{x \in X;\, x \neq 0} c_x(f)\mathcal{W}_i(f_x) = \sum_{x \in X;\, x \neq 0} c_x(f)\theta_i(f_x) = \theta_i(f)$ *by Lemmas 4, 5 and 6. Hence* $\mathcal{W}_i(f) = \theta_i(f)$ *for all* $i \in N$ *and* $f \in \mathcal{E}_\sigma^X$. ■

Proof (Proposition 3) : *We first need to recall a preliminary result that provides a sufficient condition for the existence of a Nash equilibrium in a strategic form game. It is derived from corollaries 2.2 and 2.9 of Monderer and Shapley (1996).*

Lemma 7 : *A finite strategic form game* $\langle N, (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$ *admits a pure Nash equilibrium if for every* $i, j \in N$, *for every* $a \in \prod_{k \in N \setminus \{i,j\}} X_k$, *and for every* $x_i, y_i \in X_i$ *and* $x_j, y_j \in X_j$,

$$u_i(B) - u_i(A) + u_j(C) - u_j(B) + u_i(D) - u_i(C) + u_j(A) - u_j(D) = 0 \quad (3.12)$$

where $A = (x_i, x_j, a)$, $B = (y_i, x_j, a)$, $C = (y_i, y_j, a)$, *and* $D = (x_i, y_j, a)$.

We now prove Proposition 3. Let $\mathcal{E} = \langle N, X, o, f^s \rangle$ *be a tuple and* $\langle N, (X_i)_{i \in N}, (u_i)_{i \in N} \rangle$ *the strategic form game where*

$$u_i(x) = Sh_i(f^s, x) - c_i(x_i) \text{ for all } x \in X$$

From Proposition 2, $u_i(x)$ *can also be written as* $u_i(x) = Sh_i(f^p, x)$. *Therefore, from Lemma 7, it suffices to show that Equation 3.12 is satisfied. Let* $T_1 = u_i(B) - u_i(A)$, $T_2 = u_j(C) - u_j(B)$, $T_3 = u_i(D) - u_i(C)$ *and* $T_4 = u_j(A) - u_j(D)$. *Viewing all players but players* i *and* j *as one player, we have that:*

$$
\begin{aligned}
T_1 &= u_i(B) - u_i(A)\\
&= Sh_i(f^p,(y_i,x_j,a)) - Sh_i(f^p,(x_i,x_j,a))\\
&= \frac{1}{3}\left(f^p(y_i,x_j,a) - f^p(o_i,x_j,a) + f^p(y_i,o_j,o_{-ij}) - f^p(o_i,o_j,o_{-ij})\right)\\
&+ \frac{1}{6}\left(f^p(y_i,x_j,o_{-ij}) - f^p(o_i,x_j,o_{-ij}) + f^p(y_i,o_j,a) - f^p(o_i,o_j,a)\right)\\
&- \frac{1}{3}\left(f^p(x_i,x_j,a) - f^p(o_i,x_j,a) + f^p(x_i,o_j,o_{-ij}) - f^p(o_i,o_j,o_{-ij})\right)\\
&- \frac{1}{6}\left(f^p(x_i,x_j,o_{-ij}) - f^p(o_i,x_j,o_{-ij}) + f^p(x_i,o_j,a) - f^p(o_i,o_j,a)\right)\\
&= \frac{1}{3}\left(f^p(y_i,x_j,a) + f^p(y_i,o_j,o_{-ij}) - f^p(x_i,x_j,a) - f^p(x_i,o_j,o_{-ij})\right)\\
&+ \frac{1}{6}\left(f^p(y_i,x_j,o_{-ij}) + f^p(y_i,o_j,a) - f^p(x_i,x_j,o_{-ij}) - f^p(x_i,o_j,a)\right)
\end{aligned}
$$

$$
\begin{aligned}
T_3 &= u_i(D) - u_i(C)\\
&= Sh_i(f^p,(x_i,y_j,a)) - Sh_i(f^p,(y_i,y_j,a))\\
&= \frac{1}{3}\left(f^p(x_i,y_j,a) - f^p(o_i,y_j,a) + f^p(x_i,o_j,o_{-ij}) - f^p(o_i,o_j,o_{-ij})\right)\\
&+ \frac{1}{6}\left(f^p(x_i,y_j,o_{-ij}) - f^p(o_i,y_j,o_{-ij}) + f^p(x_i,o_j,a) - f^p(o_i,o_j,a)\right)\\
&- \frac{1}{3}\left(f^p(y_i,y_j,a) - f^p(o_i,y_j,a) + f^p(y_i,o_j,o_{-ij}) - f^p(o_i,o_j,o_{-ij})\right)\\
&- \frac{1}{6}\left(f^p(y_i,y_j,o_{-ij}) - f^p(o_i,y_j,o_{-ij}) + f^p(y_i,o_j,a) - f^p(o_i,o_j,a)\right)\\
&= \frac{1}{3}\left(f^p(x_i,y_j,a) + f^p(x_i,o_j,o_{-ij}) - f^p(y_i,y_j,a) - f^p(y_i,o_j,o_{-ij})\right)\\
&+ \frac{1}{6}\left(f^p(x_i,y_j,o_{-ij}) + f^p(x_i,o_j,a) - f^p(y_i,y_j,o_{-ij}) - f^p(y_i,o_j,a)\right)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
T_1 + T_3 &= \frac{1}{3}\left(f^p(y_i,x_j,a) + f^p(x_i,y_j,a) - f^p(x_i,x_j,a) - f^p(y_i,y_j,a)\right)\\
&+ \frac{1}{6}\left(f^p(y_i,x_j,o_{-ij}) + f^p(x_i,y_j,o_{-ij}) - f^p(x_i,x_j,o_{-ij}) - f^p(y_i,y_j,o_{-ij})\right)
\end{aligned}
$$

Similarly, one shows that

$$\begin{aligned} T_2 + T_4 &= \frac{1}{3}\left(f^p(y_i, y_j, a) + f^p(x_i, x_j, a) - f^p(y_i, x_j, a) - f^p(x_i, y_j, a)\right) \\ &+ \frac{1}{6}\left(f^p(y_i, y_j, o_{-ij}) + f^p(x_i, x_j, o_{-ij}) - f^p(y_i, x_j, o_{-ij}) - f^p(x_i, y_j, o_{-ij})\right) \end{aligned}$$

It follows that $T_1 + T_2 + T_3 + T_4 = 0$*. Therefore, the strategic form game* $\langle N, (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$*, where the payoff functions are defined by Equation 3.8, admits a pure strategy Nash equilibrium.*

■

Proof (Proposition 4) : *We only prove the first part of the proposition concerning the full achievability of the ex-ante Shapley value. The second part can be easily inferred from the argument presented for full achievability. Let* $\mathcal{A} = (S \cup B, (c_i)_{i\in S})$ *be an assignment market. Assume there exists a unique optimal matching* $M \in \mathcal{M}^*$*. Let us show that the price system* $(p_i)_{i\in S}$ *defined for each* $i \in S$ *by* $p_i = ESV_i(f) + c_i$ *achieves the ex-ante Shapley value of each agent in the economy. The net utility of each seller* i *is given by* $p_i - c_i = ESV_i(f)$*. It follows that the price system achieves the utility of sellers. Let us show that it also achieves the utility of buyers. For all* $j \in B$*, let* i_j *be the unique seller such that* $(i_j, j) \in M$*. Since* M *is the unique optimal assignment, it follows that* $ESV_{i_j}(f) = ESV_j(f) = Sh_j(f, x(M)) = \frac{1}{2}s_{i_j j}$*, where* $s_{i_j j} = max(0, v_{i_j j} - c_{i_j})$ *for all* $(i_j, j) \in M$*. If* $v_{i_j j} - c_{i_j} \leq 0$*, then* $ESV_j(f) = 0$ *and* $v_{i_j j} - p_{i_j} \leq 0$*. This implies that no trade occurs between buyer* j *and seller* i_j*. In this case, buyer* j *gets a reservation utility of zero, which coincides with the ex-ante Shapley value. If* $v_{i_j j} - c_{i_j} > 0$*, then* $ESV_j(f) = \frac{1}{2}(v_{i_j j} - c_{i_j})$*. In this case, the trade occurs between seller* i_j *and buyer* j*, and the latter gets a net utility equal to* $v_{i_j j} - p_{i_j} = v_{i_j j} - c_{i_j} - ESV_{i_j} = \frac{1}{2}(v_{i_j j} - c_{i_j}) = ESV_j(f)$*, as wanted.*

■

Bibliography

Adams, J Stacy (1965). “Inequity in social exchange”. In: *Advances in Experimental Social Psychology*. Vol. 2. Elsevier, pp. 267–299.

Agastya, Murali (2008). “On choosing which game to play when ignorant of the rules”. In: *Economic Theory* 34.2, pp. 297–308.

Aguiar, Victor H, Roland Pongou, and Jean-Baptiste Tondji (2018). “A non-parametric approach to testing the axioms of the Shapley value with limited data”. In: *Games and Economic Behavior* 111, pp. 41–63.

Aguiar, Victor H et al. (2019). “An index of unfairness”. In: *Handbook of the Shapley value*. Chapman and Hall/CRC, pp. 31–48.

Akerlof, George A. (1970). “The Market for "Lemons": Quality Uncertainty and the Market Mechanism”. In: *The Quarterly Journal of Economics* 84.3, pp. 488–500.

Alcalde, José and Salvador Barberà (1994). “Top dominance and the possibility of strategy-proof stable solutions to matching problems”. In: *Economic Theory* 4.3, pp. 417–435.

Algaba, Encarnación, Vito Fragnelli, and Joaquín Sánchez-Soriano (2019). *Handbook of the Shapley value*. CRC Press.

Alva, Samson and Vikram Manjunath (2020). “The impossibility of strategy-proof, Pareto efficient, and individually rational rules for fractional matching”. In: *Games and Economic Behavior* 119, pp. 15–29.

Athey, Susan (2001). “Single crossing properties and the existence of pure strategy equilibria in games of incomplete information”. In: *Econometrica* 69.4, pp. 861–889.

Aumann, Robert J and Michael Maschler (1985). “Game theoretic analysis of a bankruptcy problem from the Talmud”. In: *Journal of Economic Theory* 36.2, pp. 195–213.

Azevedo, Eduardo M and Jacob D Leshno (2016). “A supply and demand framework for two-sided matching markets”. In: *Journal of Political Economy* 124.5, pp. 1235–1268.

Bergantinos, Gustavo and Juan D Moreno-Ternero (2020). “Sharing the revenues from broadcasting sport events”. In: *Management Science* 66.6, pp. 2417–2431.

Bimpikis, Kostas, Shayan Ehsani, and Rahmi Ilkılıç (2019). “Cournot competition in networked markets”. In: *Management Science* 65.6, pp. 2467–2481.

Binmore, Ken (2004). "Reciprocity and the social contract". In: *Politics, Philosophy & Economics* 3.1, pp. 5–35.

Blair, Douglas H and Richard P McLean (1990). "Subjective evaluations of n-person games". In: *Journal of Economic Theory* 50.2, pp. 346–361.

Bogomolnaia, Anna and Hervé Moulin (2004). "Random matching under dichotomous preferences". In: *Econometrica* 72.1, pp. 257–279.

Bogomolnaia, Anna, Hervé Moulin, and Richard Stong (2005). "Collective choice under dichotomous preferences". In: *Journal of Economic Theory* 122.2, pp. 165–184.

Bosmans, Kristof, Luc Lauwers, and Erwin Ooghe (2009). "A consistent multidimensional Pigou–Dalton transfer principle". In: *Journal of Economic Theory* 144.3, pp. 1358–1371.

Brink, René van den and Agnieszka Rusinowska (2022). "The degree measure as utility function over positions in graphs and digraphs". In: *European Journal of Operational Research* 299.3, pp. 1033–1044.

Carroll, Gabriel (2015). "Robustness and linear contracts". In: *American Economic Review* 105.2, pp. 536–563.

Casajus, André and Frank Huettner (2014). "Weakly monotonic solutions for cooperative games". In: *Journal of Economic Theory* 154, pp. 162–172.

Chatterjee, Siddharth, Sinan Ertemel, and Rajnish Kumar (2023). "Rationing rules for risky claims". In: *Journal of Mathematical Economics*, p. 102868.

Chen, Yiling et al. (2013). "Truth, justice, and cake cutting". In: *Games and Economic Behavior* 77.1, pp. 284–297.

Csóka, Péter and P Jean-Jacques Herings (2021). "An axiomatization of the proportional rule in financial networks". In: *Management Science* 67.5, pp. 2799–2812.

Dai, Tianjiao and Juuso Toikka (2022). "Robust incentives for teams". In: *Econometrica* 90.4, pp. 1583–1613.

Demange, Gabrielle, David Gale, and Marilda Sotomayor (1986). "Multi-item auctions". In: *Journal of Political Economy* 94.4, pp. 863–872.

Demeze-Jouatsa, Ghislain Herman, Roland Pongou, and Jean-Baptiste Tondji (2024). "Justice, inclusion, and incentives". In: *Journal of Theoretical Politics* 36.2, pp. 101–131.

Dubey, Pradeep (1982). "The shapley value as aircraft landing fees–revisited". In: *Management Science* 28.8, pp. 869–874.

Feltovich, Nick and Joe Swierzbinski (2011). "The role of strategic uncertainty in games: An experimental study of cheap talk and contracts in the nash demand game". In: *European Economic Review* 55.4, pp. 554–574.

Folger, Robert and Russell Cropanzano (2001). "Fairness theory: Justice as accountability". In: *Advances in Organizational Justice* 1.1-55, p. 12.

Freixas, Josep (2005). "The Shapley–Shubik power index for games with several levels of approval in the input and output". In: *Decision Support Systems* 39.2, pp. 185–195.

Freixas, Josep, Dorota Marciniak, and Montserrat Pons (2012). "On the ordinal equivalence of the Johnston, Banzhaf and Shapley power indices". In: *European Journal of Operational Research* 216.2, pp. 367–375.

Gibbard, Allan (1973). "Manipulation of voting schemes: a general result". In: *Econometrica*, pp. 587–601.

Gilliland, Stephen W (1994). "Effects of procedural and distributive justice on reactions to a selection system." In: *Journal of Applied Psychology* 79.5, p. 691.

Gopalakrishnan, Sanjith et al. (2021). "Incentives and emission responsibility allocation in supply chains". In: *Management Science* 67.7, pp. 4172–4190.

Green, Jerry and Jean-Jacques Laffont (1977). "Characterization of satisfactory mechanisms for the revelation of preferences for public goods". In: *Econometrica*, pp. 427–438.

Grofman, Bernard and Guillermo Owen (1982). "A game theoretic approach to measuring degree of centrality in social networks". In: *Social Networks* 4.3, pp. 213–224.

Grossman, Sanford J. and Oliver D. Hart (1983). "An Analysis of the Principal-Agent Problem". In: *Econometrica* 51.1, pp. 7–45.

Harsanyi, John C (1967). "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model". In: *Management Science* 14.3, pp. 159–182.

— (1968). "Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points". In: *Management Science* 14.5, pp. 320–334.

Hart, Oliver and Bengt Holmström (1987). "The theory of contracts". In: *Advances in Economic Theory: Fifth World Congress*. Vol. 1. Cambridge.

He, Wei and Yeneng Sun (2019). "Pure-strategy equilibria in Bayesian games". In: *Journal of Economic Theory* 180, pp. 11–49.

Herstein, Israel N and John Milnor (1953). "An axiomatic approach to measurable utility". In: *Econometrica*, pp. 291–297.

Higgins, E Tory (1987). "Self-discrepancy: a theory relating self and affect." In: *Psychological Review* 94.3, p. 319.

Holmström, Bengt (1979). "Moral hazard and observability". In: *The Bell Journal of Economics*, pp. 74–91.

Holmstrom, Bengt (1982). “Moral hazard in teams”. In: *The Bell Journal of Economics*, pp. 324–340.

Intropi, Pietro (2024). “Reciprocal libertarianism”. In: *European Journal of Political Theory* 23.1, pp. 23–43.

Ismail, Abdussalaam Iyanda, AM Abdul-Halim, and Mohd Hasanur Raihan Joarder (2015). “Mediating role of distributive justice in the relationship between career incentives and employee performance”. In: *Journal of Economics, Business and Management* 3.10, pp. 929–935.

Jing-Yuan, Wei and Yves Smeers (1999). “Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices”. In: *Operations Research* 47.1, pp. 102–112.

Kelso Jr, Alexander S and Vincent P Crawford (1982). “Job matching, coalition formation, and gross substitutes”. In: *Econometrica*, pp. 1483–1504.

Kemahlıoğlu-Ziya, Eda and John J Bartholdi III (2011). “Centralizing inventory in supply chains by using Shapley value to allocate the profits”. In: *Manufacturing & Service Operations Management* 13.2, pp. 146–162.

Klein, Arnd Heinrich (2015). “Optimal incentives and distributive justice in competitive environments”. PhD thesis. University of Zurich.

Klein, Luciana and Romualdo Douglas Colauto (2020). “Perceptions of organizational justice in incentive contracts and their effect on congruence between personal and organizational goals”. In: *Revista Brasileira de Gestão de Negócios* 22, pp. 582–607.

Kranton, Rachel E and Deborah F Minehart (2001). “A theory of buyer-seller networks”. In: *American Economic Review* 91.3, pp. 485–508.

Krasa, Stefan and Nicholas C Yannelis (1994). “The value allocation of an economy with differential information”. In: *Econometrica*, pp. 881–900.

Lind, E Allan and Tom R Tyler (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Lister, Andrew (2013). “Reciprocity, relationships, and distributive justice”. In: *Social Theory and Practice*, pp. 70–94.

— (2020). “Publicity, reciprocity, and incentives”. In: *Canadian Journal of Philosophy* 50.1, pp. 67–82.

Littlechild, Stephen C and Guillermo Owen (1973). “A simple expression for the Shapley value in a special case”. In: *Management Science* 20.3, pp. 370–372.

Maniquet, François (2003). “A characterization of the Shapley value in queueing problems”. In: *Journal of Economic Theory* 109.1, pp. 90–103.

Manjunath, Vikram and Alexander Westkamp (2021). "Strategy-proof exchange under trichotomous preferences". In: *Journal of Economic Theory* 193, p. 105197.

Marku, Keler, Sergio Ocampo, and Jean-Baptiste Tondji (2024). "Robust contracts in common agency". In: *The RAND Journal of Economics* 55.2, pp. 199–229.

Miao, Jianjun and Alejandro Rivera (2016). "Robust contracts in continuous time". In: *Econometrica* 84.4, pp. 1405–1440.

Milgrom, Paul R and Robert J Weber (1985). "Distributional strategies for games with incomplete information". In: *Mathematics of Operations Research* 10.4, pp. 619–632.

Momi, Takeshi (2017). "Efficient and strategy-proof allocation mechanisms in economies with many goods". In: *Theoretical Economics* 12.3, pp. 1267–1306.

Monderer, Dov and Lloyd S Shapley (1996). "Potential games". In: *Games and Economic Behavior* 14.1, pp. 124–143.

Moulin, Hervé (1992). "An application of the Shapley value to fair division with money". In: *Econometrica*, pp. 1331–1349.

— (1995). *Cooperative microeconomics: a game-theoretic introduction*. Princeton University Press.

Myerson, Roger B (1977). "Graphs and cooperation in games". In: *Mathematics of Operations Research* 2.3, pp. 225–229.

— (1980). "Conference structures and fair allocation rules". In: *International Journal of Game Theory* 9, pp. 169–182.

— (1997). *Game theory: analysis of conflict*. Harvard University Press.

Nash, John (1951). "Non-cooperative games". In: *Annals of Mathematics*, pp. 286–295.

Nesterov, Alexander S (2017). "Fairness and efficiency in strategy-proof object allocation mechanisms". In: *Journal of Economic Theory* 170, pp. 145–168.

Neumann, J von and O Morgenstern (1944). *Theory of Games and Economic Behavior.-Princeton: Princeton Univ. Press.*

Pongou, Roland and Bertrand Tchantcho (2021). "Round-robin political tournaments: Abstention, truthful equilibria, and effective power". In: *Games and Economic Behavior* 130, pp. 331–351.

Pongou, Roland and Jean-Baptiste Tondji (2018). "Valuing inputs under supply uncertainty: The Bayesian Shapley value". In: *Games and Economic Behavior* 108, pp. 206–224.

Pycia, Marek and M Utku Ünver (2017). "Incentive compatible allocation and exchange of discrete resources". In: *Theoretical Economics* 12.1, pp. 287–329.

Radner, Roy and Robert W Rosenthal (1982). “Private information and pure-strategy equilibria”. In: *Mathematics of Operations Research* 7.3, pp. 401–409.

Rawls, John (1971). *A Theory of Justice*. Harvard U Press, Belknapp Press.

Roth, Alvin E (1977). “The Shapley value as a von Neumann-Morgenstern utility”. In: *Econometrica*, pp. 657–664.

— (1982). “The economics of matching: Stability and incentives”. In: *Mathematics of Operations Research* 7.4, pp. 617–628.

— (1988). *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press.

— (2019). “The Shapley Value, a Giant Legacy, and Ongoing Research Agenda”. In: *Handbook of the Shapley value*. Ed. by Encarnación Algaba, Vito Fragnelli, and Joaquín Sánchez-Soriano. Chapman and Hall/CRC, pp. xvii –xxii.

Rothschild, Michael and Joseph Stiglitz (1978). “Equilibrium in competitive insurance markets: An essay on the economics of imperfect information”. In: *Uncertainty in Economics*. Elsevier, pp. 257–280.

Satterthwaite, Mark Allen (1975). “Strategy-proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions”. In: *Journal of Economic Theory* 10.2, pp. 187–217.

Serizawa, Shigehiro (2002). “Inefficiency of strategy-proof rules for pure exchange economies”. In: *Journal of Economic Theory* 106.2, pp. 219–241.

Shapley, Lloyd S (1953). *A value for n-person games, Contributions to the Theory of Games, 2, 307–317*.

Shapley, Lloyd S and Martin Shubik (1954). “A method for evaluating the distribution of power in a committee system”. In: *American Political Science Review* 48.3, pp. 787–792.

— (1967). “Ownership and the production function”. In: *The Quarterly Journal of Economics* 81.1, pp. 88–111.

— (1971). “The assignment game I: The core”. In: *International Journal of game theory* 1.1, pp. 111–130.

Shorrocks, Anthony F and James E Foster (1987). “Transfer sensitive inequality measures”. In: *The Review of Economic Studies* 54.3, pp. 485–497.

Singh, Nirvikar and Xavier Vives (1984). “Price and quantity competition in a differentiated duopoly”. In: *The Rand Journal of Economics*, pp. 546–554.

Sinha, Abhinav and Achilleas Anastasopoulos (2017). "Incentive mechanisms for fairness among strategic agents". In: *IEEE Journal on Selected Areas in Communications* 35.2, pp. 288–301.

Spence, Michael (1974). "Competitive and optimal responses to signals: An analysis of efficiency and distribution". In: *Journal of Economic theory* 7.3, pp. 296–332.

Sung, Sun Young, Jin Nam Choi, and Sung-Choon Kang (2017). "Incentive pay and firm performance: Moderating roles of procedural justice climate and environmental turbulence". In: *Human Resource Management* 56.2, pp. 287–305.

Takeng, Rodrigue Tido (2022). "Uncertain production environment and communication structure". In: *Journal of Mathematical Economics* 102, p. 102744.

Tijs, Stef H and Theo SH Driessen (1986). "Game theory and cost allocation problems". In: *Management Science* 32.8, pp. 1015–1028.

Varian, Hal R (1992). *Microeconomic analysis.* Vol. 3. Norton New York.

Velez, Rodrigo A (2011). "Are incentives against economic justice?" In: *Journal of Economic Theory* 146.1, pp. 326–345.

Widerquist, Karl (1999). "Reciprocity and the guaranteed income". In: *Politics & Society* 27.3, pp. 387–402.

Winter, Eyal (2002). "The Shapley value". In: *Handbook of game theory with economic applications* 3, pp. 2025–2054.

Young, H Peyton (1985). "Monotonic solutions of cooperative games". In: *International Journal of Game Theory* 14.2, pp. 65–72.