# Coordination, Cooperation, and Collective Decision

Satoshi Nakada

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in
Economics

Graduate School of Economics
Hitotsubashi University

2018

**Summary**

One of the main goals of modern economics is to understand what mechanism enhances self-interested agents' coordination and cooperation to achieve a good economic outcome. This dissertation consists of five essays on studies of this theme. In the first part, from Chapter 1 to Chapter 3, we consider a problem of how we can construct an information structure or a mechanism to achieve coordinative outcomes. In the second part, from Chapter 4 to Chapter 5, we consider desirable allocation rules of cooperative games and voting rules from normative perspectives.

The first essay studies how risk and ambiguity of each player's information affect the possibility of the twin crises, that is, a bank run and a currency crisis simultaneously occur. We consider this question by constructing a simple global game model motivated by Goldstein (2005). In contrast to a standard global game model applied to the single financial crisis, we assume that two kinds of strategic complementarities exist: (1) strategic complementarity among players within each market and (2) strategic complementarity among players through the two markets. When each financial crisis is considered separately, more ambiguous information triggers the bank run, whereas less ambiguous information triggers the currency crisis (Ui, 2015). In contrast to the single financial crisis case, when the two financial markets are linked, we show that an effect of the additional risk and ambiguity on the possibility of the crises is not monotone but depends on the degree of a feed-back effect through the two markets. We characterize all patterns of the effect of the additional risk and ambiguity on the possibility of the crises. We also characterize an equilibrium outcome in the limit as the risk and ambiguity vanish.

The second essay studies network games by Jackson and Wolinsky (1996) and characterizes the class of games that have a network potential. We show that there exists a network potential if and only if each player's payoff function can be represented as the Shapley value of a special class of cooperative games indexed by the networks. We also show that a network potential coincides with a potential of the same class of cooperative games.

The third essay studies two-sided one-to-one matching models where we do not assume either complete nor transitive preferences. In this environment, we cannot

guarantee even the existence of a stable matching, which is one of the fundamental desiderata for a matching problem. We show that, if each agent's preference is acyclic, the DA algorithm with any order extension/completion rule always induces a stable matching and is strategy-proof. Despite of such good properties, there is no algorithm in the irrational preference domain such that it is strategy-proof and one-side optimal when we use an order extension/completion rule. Therefore, our result clarifies the trade-off between strategy-proofness and one-side optimality beyond the rational preference domain.

The forth essay studies a new class of allocation rules for cooperative games with transferable utility (TU-games), weighted egalitarian Shapley values, where each rule in this class takes into account each player's contributions and heterogeneity among players to determine each player's allocation. We provide an axiomatic foundation for rules with a given weight profile (i.e, exogenous weights) and the class of rules (i.e., endogenous weights). The axiomatization results illustrate the differences among our class of rules, the Shapley value, the egalitarian Shapley values, and the weighted Shapley values.

The fifth essay studies normative properties of voting rules. Notably, we propose a new normative consequentialist criterion for voting rules under Knightian uncertainty about individuals' preferences to characterize a weighted majority rule (WMR). This criterion, which is referred to as robustness, stresses the significance of responsiveness, i.e., the probability that the social outcome coincides with the realized individual preferences. A voting rule is said to be robust if, for any probability distribution of preferences, it avoids the following worst-case scenario: the responsiveness of every individual is less than or equal to one-half. Our main result establishes that a voting rule is robust if and only if it is a WMR without any ties. This characterization of a WMR avoiding the worst possible outcome complements the well-known characterization of a WMR achieving the optimal outcome, i.e., efficiency regarding responsiveness.

# Acknowledgments

First of all, I thank my advisor Takashi Ui for his continuous and patient guidance. Impressed by his studies and sincere attitude for researches, one of the goals of my Ph.D study has been to write papers that he admits me as an independent researcher. I hope that this dissertation meets his criteria.

I am also indebted to my dissertation committee members Reiko Goto, Daisuke Hirata, Koichi Tadenuma, and Norio Takeoka. They gave me fruitful advice that has significantly improved my works. Their generous support has been also encouraging me.

Some chapters in my dissertation are collaborated with Takaaki Abe, Shmuel Nitzan, Kohei Sashida, and Takashi Ui. I appreciate these opportunities. Also, this dissertation is completed during my visiting at Princeton University. I thank Mark Fleurbaey for his suggestion and hospitality during this opportunity.

I also have benefited from many people through discussion, interaction, and collaborated works. Their encouragement is helpful to my mentality throughout my whole Ph.D life. I cannot include all of them, but the partial list includes Yukihiko Funaki, Akifumui Ishihara, Hideshi Itoh, Atsushi Iwasaki, Yusuke Kasuya, Yoshimasa Katayama, Takashi Kunimoto, Hitoshi Matsushima, Shintaro Miura, Kota Murayama, Nozomu Muto, Daisuke Oyama, Hiroyuki Ozaki, Taishi Sassano, Etsuro Shioji, Yasuhiro Shirata, Takuro Yamashita, and Kazuyoshi Yamamoto.

Finally, I acknowledge the financial support from Japan Society for Promotion of Science (JSPS).

February, 20, 2018

To my love

# Contents

# Chapter 1 Risk and Ambiguity in the Twin Crises

This chapter is based on the same title of the joint work with Kohei Sashida.

## 1.1 Introduction

A panic-based financial crisis cannot be explained by the underlying fundamental value alone. In economics, this type of event can be seen as a coordination game in which such a panic-based crisis is an equilibrium outcome realized from coordination motives. If two different financial markets are connected, more complicated coordination motives arise than in the single market case and, as a result, two different financial crises often occur with a positive correlation. This kind of financial crisis referred to as *twin crises* is empirically documented by Kaminsky and Reinhart (1999) and theoretically studied by Goldstein (2005).[1]

In the financial market, market participants make decisions based on accessible information about the underlying fundamental value. Thus, information quality affects what actions are taken by market participants and, in turn, affects the economic outcome. We usually measure information quality as the precision of information: we consider a signal with lower variance as the high quality information.[2] In contrast to this view, there is another type of information quality, which is referred to as a degree of unknown precision of information. As experimental studies suggest, these two types of information qualities have different implications to a decision maker's

[1] The studies of financial crises as coordination games typically face a multiple equilibria problem. To avoid it, incomplete information games which are called *global games* introduced by Carlsson and van Damme (1993) are often used. Morris and Shin (2000, 2004) and Goldstein and Pauzner (2005) use a global game to study a bank run problem based on Diamond and Dybvig (1983). Morris and Shin (1998) use a global game to study a currency crisis based on Obstfeld (1996). Goldstein (2005) uses a global game to study twin crises. See Morris and Shin (2003) for a survey of the literature.

[2] Effects of information quality on economic outcomes in this sense are extensively studied. Morris and Shin (2002) consider how the precision of public and private information affects the social welfare in the beauty contest games. Angeletos and Pavan (2007) and Ui and Yoshizawa (2015) consider the same question in the more general quadratic payoff games. Iachan and Nenov (2015) consider how the information quality affects the probability of a financial crisis in the regime change games.

action when a decision maker is ambiguity averse.[3] We call the former information quality as *risk* and the latter information quality as *ambiguity*. Ui (2015) shows that the risk and ambiguity have opposite effects on the possibility of a financial crisis. In particular, the more ambiguous information triggers the bank run and the less ambiguous information triggers the currency crisis when we consider these financial crises separately. In this chapter, we reconsider how the risk and ambiguity affect the possibility of the financial crises when the two markets are linked.

We answer this question by characterizing how the risk and ambiguity affect the probability of the twin crises in a simple global game model. Suppose that there are two financial markets, a commercial banking sector (market B) and a currency sector (market C). In this economy, an exchange rate is fixed and the government commits itself to offer financial assistance to the bank to prevent a bank run. Players of the game are foreign creditors in the market B and speculators in the market C. In the market B, each creditor can withdraw his own money from the bank in the short term. In this case, he can obtain no additional dividends. On the other hand, if he does not withdraw the money from the bank, he may get additional dividends: the bank invests the amount of money creditors deposit, so if the fundamental of the economy is good, they can obtain high returns from the investment. However, if many creditors withdraw the money from the bank, the return will be smaller because the amount of investment is decreased. In the market C, each speculator wants to attack the current fixed exchange rate to obtain an arbitrage gain. If he does not attack the exchange rate, he does not either pay or obtain additional money. On the other hand, if he attacks the exchange rate and the speculative attack succeeds, he can obtain the arbitrage gain.

A key point in the model is a strategic feed-back effect through the two markets. In the market B, the return from the investment by the bank is financed by the home currency. However, the bank must pay the foreign currency to each creditor. Therefore, if more speculators attack the exchange rate and, as a result, the value of the home currency decreases, then the return on the bank also decreases. As a result, the returns on the creditors decrease as well. Conversely, if more creditors withdraw their money from the bank, then the government must offer financial assistance to

[3] The seminal example is Ellsberg (1961). See Gilboa and Marinacci (2013) for a recent survey of the literature.

the bank. This financial support will cause a downward pressure on the current exchange rate, which makes the speculative attack successful easier. Therefore, the more creditors withdraw the money from the bank, the higher payoff speculators can obtain from the currency attack. Another key point in the model is that we assume that each player obtains private information about the fundamental of the economy and he faces ambiguity about the true signal distribution. Our model is a stripped down version of Goldstein (2005) with an additional assumption that each player faces ambiguity about the fundamental and is a maximin expected utility decision maker (Gilboa and Schmeidler, 1989).[4]

We first show that this model is dominance solvable and has a unique equilibrium.[5] The equilibrium is characterized by the cutoff points, that is, each player takes an action to induce the crises if and only if his signal is less than the cutoff point in each market. Since we can calculate these cutoff points explicitly, we can provide comparative statics results of how the risk and ambiguity affect the cutoff points. Our main result characterizes when the more/less risky and ambiguous information can reduce the possibility of the crises. The key channel is an influence from one market to the other market. This power of propagation is captured by what we call attraction values. We show that if the attraction value of the market B is higher enough than that of the market C, then the more risky and less ambiguous information can reduce the possibility of both crises. Conversely, if the attraction value of the market C is higher enough than that of the market B, then the less risky and more ambiguous information can reduce the possibility of both crises. In contrast to the single financial crisis case, these results show that the more ambiguous information may reduce the possibility of the bank run and the less ambiguous information may reduce the possibility of the currency crisis depending on the attraction values. Thus, these results have different implications of the effect of the risk and ambiguity on the possibility of the single financial crisis from Ui (2015). Summarizing these observations, we show that all possible patterns of the effect on the possibility of each crisis are classified into twelve types.

[4]Laskar (2014) considers a similar linear global game with a maximin expected utility decision maker where a single financial market is considered. Our model is an extension of his model to accommodate a link between the two financial markets.

[5]Without ambiguity, Goldstein (2005) shows that his model has a unique equilibrium, but he does not show that it is survived by the iterative deletion of interim dominated strategies.

We are also interested in outcomes in the limit as the noise vanishes. The outcome in the limit can be seen as an approximation of an outcome in the complete information model. First, we show that two cutoff points have a positive correlation with each other as demonstrated by Goldstein (2005). In some cases, this correlation is perfect in the sense that either both crises occur or none of them occurs. Moreover, the cutoff point in the limit is higher than that of the single market case. Even in other cases, the cutoff point in the limit in each market is higher than that of the single market case. A difference from the no-ambiguity case is that relative uncertainty between the risk and ambiguity also affects the prediction of the outcome in the limit. That is, what limit case we consider depends on the relative convergence speed of the risk and ambiguity. We show how the outcome in the limit depends on the relative uncertainty between the risk and ambiguity.

This chapter contributes to the recent growing literature on incomplete information games with ambiguity averse players. Epstein (1997) and Epstein and Wang (1996) consider general incomplete information games with ambiguity averse players including the maximin decision makers.[6] Kajii and Ui (2005) consider the existence of equilibria and an implication of the strategic ambiguity in the game with the maximin decision makers. Stauber (2011) considers the robustness of Bayesian Nash equilibria to small amount of ambiguous information with another type of ambiguity averse decision makers in the sense of Bewly (2002). Our analysis is consistent with the framework of Kajii and Ui (2005). Although most papers consider mechanism designs as applications,[7] we consider an impact of the ambiguity on the strategic situation in global games. In particular, our study complements the finding by Ui (2015) as an application to the twin crises.[8]

The rest of this chapter is organized as follows. In section 1.2, we introduce the model. In section 1.3, we derive the unique equilibrium. In section 1.4, we provide our main characterization results. In section 1.5, we discuss the outcome in the case of the small risk and ambiguity. Section 1.6 is the conclusion of this chapter. All omitted proofs are relegated to Appendix A.

---

[6]See also Ahn (2007) and Azrieli and Teper (2011).

[7]See Salo and Weber (1995), Lo (1998), Bose et al. (2006), Turocy (2008), Bose and Daripa (2009), Bodoh-Creed (2012), Lopomo et al. (2009), Di Tillio et al. (2012), Bose and Renou (2014), and Chiesa et al. (2015).

[8]Kawagoe and Ui (2015) study a two-player global game experimentally and obtain data supporting the finding by Ui (2015).

## 1.2 Model

### 1.2.1 Setup

We consider a country where there are two sectors of the financial markets, a banking sector (market B) and a currency sector (market C). We assume that the exchange rate is fixed and the government commits itself to offer financial assistance to the bank to prevent a bank run. In each market, there are continuum of players indexed by $i \in [0,1]$. In the market B, each player corresponds to a foreign creditor who holds claims on the commercial bank in this country. Each player decides whether to withdraw (W) his money from the bank or not (NW). In the market C, each player corresponds to a speculator to the fixed exchange rate. Each player decides whether to attack (A) a fixed exchange rate or not (NA). We denote $n \in [0,1]$ by the fraction of players who play W in the market B and denote $m \in [0,1]$ by the fraction of players who play A in the market C. Let $\theta \in \mathbb{R}$ be a common fundamental value in the two markets.

We then describe each player's payoff. In the market B, if a player chooses W, he obtains a constant payoff normalized to 0 regardless of $\theta$ and other players' actions. On the other hand, if he chooses NW, he obtains $\alpha_B\theta - \beta_B n - \gamma_B m$ where $\alpha_B, \beta_B, \gamma_B > 0$. In the market C, if a player chooses NA, he obtains a constant payoff normalized to 0 regardless of $\theta$ and other players' actions. On the other hand, if he chooses A, he obtains $-\alpha_C\theta + \beta_C m + \gamma_C n$ where $\alpha_C, \beta_C, \gamma_C > 0$. Note that payoffs from W and NA do not depend on both the state and other players' actions. We call such an action a *safe* action.

In the above payoff specification, the parameters $\alpha_B$ and $\alpha_C$ correspond to the marginal payoffs with respect to the fundamental value $\theta$. The parameters $\beta_B$ and $\beta_C$ are the degrees of strategic complementarity within each market, which describes marginal payoffs with respect to an increase of $n$ and $m$, respectively. The key ingredients of the model are the parameters $\gamma_B$ and $\gamma_C$, which are the degrees of strategic complementarity through the two markets.[9] That is, $\gamma_B$ describes a degree of the strategic effect from the market C to the market B and $\gamma_C$ describes a degree of the strategic effect from the market B to the market C.

Each player simultaneously chooses his action in each market. No player can observe the true value of $\theta$, but each player receives a signal about $\theta$. We assume

[9] See Goldstein (2005) for a detailed discussion of these parameters.

that $\theta$ is (improperly) uniformly distributed over the real line. Each player $i$ receives private information $x_i = \theta + \varepsilon_i$ where $\varepsilon_i$ is independently and identically distributed according to the normal distribution $N(0, \sigma^2)$. As $\sigma$ decreases, each agent can obtain the more precise information about $\theta$. The cumulative distribution function of $N(0, 1)$ is denoted by $\Phi(x) : \mathbb{R} \to [0, 1]$.

This simple model corresponds to that of Goldstein (2005) in the following sense. In the market B, each creditor can withdraw his own money from the bank in the short term. In this case, he can obtain no additional dividends, so that the payoff is 0. On the other hand, if he does not withdraw the money from the bank, he may get additional dividends: the bank invests the amount of money creditors deposit, so if the fundamental of the economy $\theta$ is good, they can obtain high returns from the investment. Therefore, an additional payoff depends on $\theta$ and its magnitude is described by $\alpha_B$. However, if many creditors withdraw the money from the bank, the return will be smaller because the amount of investment is decreased. This is the source of the strategic complementarity within the market B and its magnitude is described by $\beta_B$.

In the market C, each speculator wants to attack the fixed exchange rate to obtain an arbitrage gain. If he does not attack the exchange rate, he does not either pay or obtain additional money, so that the payoff is 0. Conversely, if he attacks the exchange rate and the speculative attack succeeds, he can obtain the arbitrage gain. Whether the speculative attack is successful or not depends on the two channels. The first one is the value of $\theta$: when the fundamental of the economy is bad, the speculative attack can succeed more easier than when it is good. Therefore, an additional payoff depends on $\theta$ and its magnitude is described by $\alpha_C$. The second one is the fraction of speculators taking A: the more speculators attack the exchange rate, the more successful the speculative attack is. This is the source of the strategic complementarity within the market C and its magnitude is described by $\beta_C$.

A key ingredient is that there is a strategic feed-back effect through the two markets. In the market B, the return from the investment by the bank is financed by the home currency. However, the bank must pay the foreign currency to each creditor. Therefore, if more speculators attack the exchange rate and, as a result, the value of the home currency decreases, then the return on the bank also decreases. As a result, the returns on the creditors decrease as well. This effect is captured by the parameter $\gamma_B$: when more speculators attack the exchange rate the payoff from NW decreases. Conversely, if more creditors withdraw their money from the bank, then the government offers financial assistance to the bank to prevent the bank run. This financial support will cause a downward pressure on the current exchange rate, which makes the speculative attack successful easier. Therefore, the more creditors withdraw the money from the bank, the higher payoff speculators can obtain from the currency attack. This effect is captured by the parameter $\gamma_C$: when more creditors withdraw the money from the bank the payoff from A increases.

### 1.2.2 Knightian uncertainty

We introduce ambiguity about the fundamental to the above setup. In the global game literature, Laskar (2014) and Ui (2015) consider the effect of ambiguity to the equilibrium outcome. We follow the model of Laskar (2014) and assume that each agent receives ambiguous private information as

$$x_i = \theta + \mu_i + \varepsilon_i \text{ where } \mu_i \in [-\eta, \eta] \text{ and } \eta > 0.$$

Each player knows that $\mu_i \in [-\eta, \eta]$ but does not know what is the true parameter. That is, each agent knows that his private information is drawn from the true distribution in the set $\left\{N(\mu_i, \sigma^2)|\mu_i \in [-\eta, \eta]\right\}$, but does not know what distribution is true. The parameter $\eta$ describes the degree of ambiguity: the higher value $\eta$ takes, the more ambiguity each player faces. We regard $(\sigma, \eta)$ as information quality.

We assume that each player is the maximin decision maker (Gilboa and Schmeidler, 1989). Under this hypothesis, each player $i$ is ambiguity averse and evaluates the payoff from each action in terms of the worst-case scenario over $[-\eta, \eta]$ depending on his private signal $x_i$. Our equilibrium analysis conducted in the next section is consistent with Kajii and Ui (2005) where they consider an incomplete information game with the maximin decision makers.[10]

One might argue why each player faces the ambiguity about the fundamental. An interpretation is that the policy maker sometimes does not report information explicitly and sometimes uses ambiguous language strategically, so that the market participants may interpret the information subjectively.[11] Another interpretation is

[10] Lasker (2014) and Ui (2015) can be considered as the specific applications of Kajii and Ui (2007). Moreover, Ui (2015) considers the more general framework about ambiguous information than that of Lasker (2014). Hence, Lasker (2014) can be seen as the subclass of Ui (2015).

[11] See Halpern and Kets (2015) for the discussion of this interpretation.

that each player considers multiple scenarios based on news. This interpretation may be relevant when we consider a central bank's report which is called the *fan chart*. Since February 1996, the financial report from the Bank of England has included the multiple scenarios for the forecast of the inflation rate.[12] The fan chart is also used to forecast other variables like the GDP growth rate and the unemployment rate and is also used by other central banks like the European Central Bank and the Bank of Japan. If the market participants are ambiguity averse as in our model, they will consider the worst-case scenario when they take an action.

## 1.3 Equilibrium

Following the standard analysis in the global game literature, we focus on monotone strategies called switching strategies. We show that such a strategy profile constitutes a unique equilibrium in our model.[13] The following calculation is similar to Laskar (2014).

Let $\theta_B$ and $\theta_C$ be the cutoff values in the market B and C, respectively. The switching strategies $s(\theta_B)$ and $s(\theta_C)$ are defined as follows.

$$\text{Market B}: \quad s_i(x_i \mid \theta_B) = \begin{cases} W & \text{if } x_i \leq \theta_B, \\ NW & \text{if } x_i > \theta_B. \end{cases}$$

$$\text{Market C}: \quad s_i(x_i \mid \theta_C) = \begin{cases} A & \text{if } x_i \leq \theta_C, \\ NA & \text{if } x_i > \theta_C. \end{cases}$$

Following this strategy, each creditor withdraws his money from the bank only when he obtains a bad signal for the fundamental $\theta$. Similarly, each speculator attacks the exchange rate only when he obtains a bad signal for the fundamental $\theta$. For each player $i$, we denote $E[\theta \mid x_i]$ by the conditional expectation of $\theta$ on $x_i$. Similarly, we denote $E[n \mid x_i]$ and $E[m \mid x_i]$ by the conditional expectation of $n$ and $m$ on $x_i$, respectively. By the law of large numbers and our information structure, we can calculate each value as follows:

$$E[\theta \mid x_i] \quad = \quad x_i - \mu_i.$$

[12] See Britton et al. (1998).

[13] Goldstein (2005) shows that the monotone strategy is the only equilibrium strategy under the no ambiguity case.

$$
\begin{aligned}
E[n \mid x_i] &= Prob(x_j \leq \theta_B \mid x_i) \\
&= Prob(x_i - \varepsilon_i - \mu_i + \varepsilon_j + \mu_j \leq \theta_B \mid x_i) \\
&= Prob(\varepsilon_j - \varepsilon_i \leq \theta_B - x_i + \mu_i - \mu_j \mid x_i) \\
&= \Phi(\frac{\theta_B - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}).
\end{aligned}
$$

$$
\begin{aligned}
E[m \mid x_i] &= Prob(x_j \leq \theta_C \mid x_i) \\
&= Prob(x_i - \varepsilon_i - \mu_i + \varepsilon_j + \mu_j \leq \theta_C \mid x_i) \\
&= Prob(\varepsilon_j - \varepsilon_i \leq \theta_C - x_i + \mu_i - \mu_j \mid x_i) \\
&= \Phi(\frac{\theta_C - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}).
\end{aligned}
$$

where we denote $j \neq i$ by a representative player other than $i$ in view of player $i$.

We denote $\pi_B^W(\theta_B, \theta_C, x_i, \mu_i, \mu_j)$ and $\pi_B^{NW}(\theta_B, \theta_C, x_i, \mu_i, \mu_j)$ by the conditional payoff on $x_i$ of a player in the market B when he takes W and NW, respectively. Similarly, we denote $\pi_C^A(\theta_B, \theta_C, x_i, \mu_i, \mu_j)$ and $\pi_C^{NA}(\theta_B, \theta_C, x_i, \mu_i, \mu_j)$ by the conditional payoff on $x_i$ of a player in the market C when he takes A and NA, respectively. By the above calculation, we obtain the following:

$$\pi_B^W(\theta_B, \theta_C, x_i, \mu_i, \mu_j) = 0.$$

$$
\begin{aligned}
\pi_B^{NW}(\theta_B, \theta_C, x_i, \mu_i, \mu_j) &= \alpha_B E[\theta \mid x_i] - \beta_B E[n \mid x_i] - \gamma_B E[m \mid x_i] \\
&= \alpha_B (x_i - \mu_i) - \beta_B \Phi(\frac{\theta_B - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}) - \gamma_B \Phi(\frac{\theta_C - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}).
\end{aligned}
$$

$$
\begin{aligned}
\pi_C^A(\theta_B, \theta_C, x_i, \mu_i, \mu_j) &= -\alpha_C E[\theta \mid x_i] + \beta_C E[m \mid x_i] + \gamma_C E[n \mid x_i] \\
&= -\alpha_C (x_i - \mu_i) + \beta_C \Phi(\frac{\theta_C - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}) + \gamma_C \Phi(\frac{\theta_B - x_i}{\sqrt{2}\sigma} + \frac{\mu_i - \mu_j}{\sqrt{2}\sigma}).
\end{aligned}
$$

$$\pi_C^{NA}(\theta_B, \theta_C, x_i, \mu_i, \mu_j) = 0.$$

If $\mu_i$ is high, player $i$ can more likely to obtain a good signal, so that the conditional probability of that other players obtain signals less than the cutoff points is high. This implies that, in the market B, the expected payoff of $i$ from NW is decreasing with respect to $\mu_i$, but in the market C, the expected payoff of $i$ from A is increasing with respect to $\mu_i$. Conversely, if $\mu_j$ is high, player $j$ can more likely to obtain a good signal, so that the conditional probability (in view of player $i$) of that player $j$ obtains a signal less than the cutoff points is low. This implies that, in the market B, the expected payoff of $i$ from NW is increasing with respect to $\mu_j$, but in the market C, the expected payoff of $i$ from A is decreasing with respect to $\mu_j$. Since each player is the maximin decision maker, he evaluates the payoff from each action by the values $\mu_i$ and $\mu_j$ which minimize the expected utility. Therefore, the above arguments imply that, for players in the market B, they evaluate the expected utility by $\mu_i = \eta$ and $\mu_j = -\eta$. Similarly, for players in the market C, they evaluate the expected utility by $\mu_i = -\eta$ and $\mu_j = \eta$. Thus, we have the payoff from each action under the worst-case scenario as follows:

$$\min_{\mu_i,\mu_j} \pi_B^W(\theta_B, \theta_C, x_i, \mu_i, \mu_j) \quad = \quad 0.$$

$$\begin{aligned} V_B(x_i, (\theta_B, \theta_C)) \quad &\equiv \quad \min_{\mu_i,\mu_j} \pi_B^{NW}(\theta_B, \theta_C, x_i, \mu_i, \mu_j) \\ &= \quad \alpha_B(x_i - \eta) - \beta_B \Phi\left(\frac{\theta_B - x_i}{\sqrt{2}\sigma} + \sqrt{2}\frac{\eta}{\sigma}\right) - \gamma_B \Phi\left(\frac{\theta_C - x_i}{\sqrt{2}\sigma} + \sqrt{2}\frac{\eta}{\sigma}\right). \end{aligned}$$

$$\begin{aligned} V_C(x_i, (\theta_B, \theta_C)) \quad &\equiv \quad \min_{\mu_i,\mu_j} \pi_C^A(\theta_B, \theta_C, x_i, \mu_i, \mu_j) \\ &= \quad -\alpha_C(x_i + \eta) + \beta_C \Phi\left(\frac{\theta_C - x_i}{\sqrt{2}\sigma} - \sqrt{2}\frac{\eta}{\sigma}\right) + \gamma_C \Phi\left(\frac{\theta_B - x_i}{\sqrt{2}\sigma} - \sqrt{2}\frac{\eta}{\sigma}\right). \end{aligned}$$

$$\min_{\mu_i,\mu_j} \pi_C^{NA}(\theta_B, \theta_C, x_i, \mu_i, \mu_j) \quad = \quad 0.$$

Then, we define the functions $b_B : \mathbb{R}^2 \to \{W, NW\}$ and $b_C : \mathbb{R}^2 \to \{A, NA\}$ such that $V_B(b_B(\theta_B, \theta_C), (\theta_B, \theta_C)) = 0$ and $V_C(b_C(\theta_B, \theta_C), (\theta_B, \theta_C)) = 0$. Note that, given $\theta_C$, (i) $V_B(x_i, (\theta_B, \theta_C))$ is continuous with respect to $(x_i, \theta_B)$, increasing with respect to $x_i$, and decreasing with respect to $\theta_B$, and (ii) $V_B(k, (k, \theta_C))$ is increasing with respect to $k$. This implies that the function $b_B(\cdot, \cdot)$ is well-defined. The symmetric argument shows that $b_C(\cdot, \cdot)$ is also well-defined. By the above argument, if opponents follow the switching strategies $s(\theta_B)$ and $s(\theta_C)$, the best response for player $i$ is the switching strategy $s(b_B(\theta_B, \theta_C))$ if $i$ is in the market B and $s(b_C(\theta_B, \theta_C))$ if $i$ is in the market C. Therefore, if $b_B(\hat{\theta}_B, \hat{\theta}_C) = \hat{\theta}_B$ and $b_B(\hat{\theta}_B, \hat{\theta}_C) = \hat{\theta}_C$, the switching strategies

$s(\hat{\theta}_B)$ and $s(\hat{\theta}_C)$ constitute an equilibrium. Thus, we can derive the equilibrium cutoff points by considering the following simultaneous equations:

$$\alpha_B\theta_B - \alpha_B\eta - \beta_B\Phi(\sqrt{2}\frac{\eta}{\sigma}) - \gamma_B\Phi(\frac{\theta_C-\theta_B}{\sqrt{2}\sigma} + \sqrt{2}\frac{\eta}{\sigma}) = 0, \tag{1.1}$$

$$-\alpha_C\theta_C - \alpha_C\eta + \beta_C\Phi(-\sqrt{2}\frac{\eta}{\sigma}) + \gamma_C\Phi(\frac{\theta_B-\theta_C}{\sqrt{2}\sigma} - \sqrt{2}\frac{\eta}{\sigma}) = 0. \tag{1.2}$$

Let $\hat{\theta}_B$ and $\hat{\theta}_C$ be the equilibrium cutoff points in the market B and the market C, respectively. The higher the equilibrium cutoff points become, the easier financial crises occur. Therefore, we regard $\hat{\theta}_B$ and $\hat{\theta}_C$ as the indicators of the probability of the financial crises.

We can show that these simultaneous equations have the unique solution $\hat{\theta}_B$ and $\hat{\theta}_C$, which implies the existence of a switching strategy equilibrium. Furthermore, this is the unique equilibrium and this game is dominance solvable.

**Proposition 1.1.** The switching strategies $s(\hat{\theta}_B)$ and $s(\hat{\theta}_C)$ are the unique strategies surviving iterated deletion of interim-dominated strategies. In the market B, there exists the unique cutoff value $\hat{\theta}_B \in [\theta_B^*(\sigma,\eta), \theta_B^{**}(\sigma,\eta)]$ and, in the market C, there exists the unique cutoff value $\hat{\theta}_C \in [\theta_C^*(\sigma,\eta), \theta_C^{**}(\sigma,\eta)]$ where

$$\begin{aligned}
\theta_B^*(\sigma,\sigma) &= \eta + \frac{\beta_B}{\alpha_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}),\\
\theta_B^{**}(\sigma,\eta) &= \frac{\gamma_B}{\alpha_B} + \eta + \frac{\beta_B}{\alpha_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}),\\
\theta_C^*(\sigma,\eta) &= -\eta + \frac{\beta_C}{\alpha_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma}),\\
\theta_C^{**}(\sigma,\eta) &= \frac{\gamma_C}{\alpha_C} - \eta + \frac{\beta_C}{\alpha_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma}).
\end{aligned}$$

Note that $\theta_B^*$ and $\theta_B^{**}$ correspond to the equilibrium cutoff points when players in the market B believe that $m = 0$ ($\theta_C \to -\infty$) and $m = 1$ ($\theta_C \to \infty$), respectively. The former case coincides with the cutoff point where $\gamma_B = 0$ and the latter case coincides with the cutoff point where every creditor knows every speculator attacks the exchange rate. A similar argument is also applied for $\theta_C^*$ and $\theta_C^{**}$. By the effect of the strategic complementarity through the two markets, the equilibrium cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$ are always higher than $\theta_B^*$ and $\theta_C^*$, respectively. This is an implication of a strategic feed-back effect through the two markets.[14]

[14]Goldstein (2005) calls this feed-back effect as the vicious cycle of the twin crises.

In contrast to the no-ambiguity case ($\eta = 0$), the existence of the ambiguity ($\eta > 0$) has a following effect on the equilibrium cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$. Since $\theta_B^*(\sigma, 0) < \theta_B^*(\sigma, \eta)$ and $\theta_B^{**}(\sigma, 0) < \theta_B^{**}(\sigma, \eta)$, the equilibrium cutoff point $\hat{\theta}_B$ may increase. On the contrary, since $\theta_C^*(\sigma, 0) > \theta_C^*(\sigma, \eta)$ and $\theta_C^{**}(\sigma, 0) > \theta_C^{**}(\sigma, \eta)$, the equilibrium cutoff point $\hat{\theta}_C$ may decrease. If $\gamma_B = \gamma_C = 0$, this prediction in each market is consistent with the finding by Ui (2015), who shows that the more ambiguous information makes each player take a safe action. This means that, in the market B, the more ambiguous information makes the financial crisis more probable, but, in the market C, more ambiguous information makes the financial crisis less probable.

However, when $\gamma_B > 0$ and $\gamma_C > 0$, $\hat{\theta}_B$ may decrease and/or $\hat{\theta}_C$ may decrease. This is because the increase of $\hat{\theta}_B$ may induce the increase of $\hat{\theta}_C$ and, similarly, the decrease of $\hat{\theta}_C$ may induce the decrease of $\hat{\theta}_B$ by the feed-back effect through the two markets. In the following section, we characterize how $\hat{\theta}_B$ and $\hat{\theta}_C$ are affected by the increase of $\sigma$ and $\eta$.

## 1.4 Main results

We consider the effect of the increase of the risk $\sigma$ and ambiguity $\eta$ on the equilibrium cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$ to understand how the probability of the financial crises are affected by the information quality.

### 1.4.1 Effect of the risk under no-ambiguity

First, we consider the no-ambiguity case ($\eta = 0$) for a comparison. The following proposition shows that both cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$ are monotonic with respect to $\sigma$.

**Proposition 1.2.** Let $\bar{\theta}_B = \frac{\theta_B^* + \theta_B^{**}}{2}$ and $\bar{\theta}_C = \frac{\theta_C^* + \theta_C^{**}}{2}$. Then, we have the following:

(i) $\frac{\partial \hat{\theta}_B}{\partial \sigma} > 0$ if $\hat{\theta}_B \in [\theta_B^*, \bar{\theta}_B)$ and $\frac{\partial \hat{\theta}_B}{\partial \sigma} \leq 0$ if $\hat{\theta}_B \in [\bar{\theta}_B, \theta_B^{**}]$,

(ii) $\frac{\partial \hat{\theta}_C}{\partial \sigma} > 0$ if $\hat{\theta}_C \in [\theta_C^*, \bar{\theta}_C)$ and $\frac{\partial \hat{\theta}_C}{\partial \sigma} \leq 0$ if $\hat{\theta}_C \in [\bar{\theta}_C, \theta_C^{**}]$.

Note that Goldstein (2005) does not discuss this kind of comparative statics result because of the complexity of his model. Intuition behind this result is as follows. Suppose that $\hat{\theta}_B$ is sufficiently high ($\hat{\theta}_B > \bar{\theta}_B$). In this case, the region in which the fundamental $\theta$ is less than $\hat{\theta}_B$ is also large. Therefore, if each player obtains the more accurate signal, the probability that his private signal is less than the cutoff value is higher, which implies that the bank run occurs with the higher probability. On the other hand, when $\hat{\theta}_B$ is sufficiently low ($\hat{\theta}_B < \bar{\theta}_B$), the region in which the fundamental $\theta$ is less than $\hat{\theta}_B$ is also small. Therefore, if each player obtains the more accurate signal, the probability that his private signal is less than the cutoff value is lower, which implies that the bank run occurs with the lower probability. The same argument holds in the market C.

### 1.4.2 Effect of the risk and ambiguity

As we will see it later, a simple result in the above case does not hold if the ambiguity exists ($\eta > 0$). One reason is that both $[\theta_B^*, \theta_B^{**}]$ and $[\theta_C^*, \theta_C^{**}]$ themselves are changed depending on $\sigma$ and $\eta$. To see this, suppose that $\gamma_B$ and $\gamma_C$ are small enough, so that the two markets are almost separated. In this case, $\hat{\theta}_B \approx \theta_B^* \approx \theta_B^{**}$ and it becomes low (high) as $\sigma$ ($\eta$) increases, whereas $\hat{\theta}_C \approx \theta_C^* \approx \theta_C^{**}$ and it becomes high (low) as $\sigma$ ($\eta$) increases. A similar argument is applied for any value of $\gamma_B$ and $\gamma_C$. However, by the feed-back effect through the two markets, both $\hat{\theta}_B$ and $\hat{\theta}_C$ tend to move in the same direction. These opposite movements play tug of war in direction of $\hat{\theta}_B$ and $\hat{\theta}_C$. Therefore, the final direction of $\hat{\theta}_B$ and $\hat{\theta}_C$ will be determined by the balance of these opposite movements.

We show that this effect is quantified by the parameters of the strategic complementarities in the following sense. Remember that the parameters $\beta_B$ and $\beta_C$ are the coefficients of the strategic complementarity within each market, which means that the higher $\beta_B(\beta_C)$ becomes, the more the payoff is affected by other players' actions in the market B (the market C). Similarly, the parameters $\gamma_B$ and $\gamma_C$ are the coefficients of the strategic complementarity through the two markets, which means that the higher $\gamma_B(\gamma_C)$ becomes, the more the payoff is affected by other players' actions in the market C (the market B). Therefore, the value $\Delta_{BC} = \beta_B\gamma_C$ can be seen as the degree of influence from the market B to the market C. Similarly, the value $\Delta_{CB} = \beta_C\gamma_B$ can be seen as the degree of influence from the market C to the market B. We call these values $\Delta_{BC}$ and $\Delta_{CB}$ *attraction values*. For the additional risk (higher $\sigma$) and the additional ambiguity (higher $\eta$), we will see that the impact on equilibrium cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$ is characterized by the degree of $\Delta_{BC}$ and $\Delta_{CB}$. The following proposition shows the effect of the increase of the risk on the cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$.

**Proposition 1.3.** There exists values $-\infty \leq X_\sigma, Y_\sigma \leq \infty$ such that

(i) $\frac{\partial \hat{\theta}_B}{\partial \sigma} < 0$ if and only if $\Delta_{CB} - \Delta_{BC} < X_\sigma$,

(ii) $\frac{\partial \hat{\theta}_C}{\partial \sigma} < 0$ if and only if $\Delta_{CB} - \Delta_{BC} < Y_\sigma$.

Proposition 1.3 shows that when the effects from the market B to the market C is high enough, the more risky information can reduce the possibility of each crisis. In particular, when $\Delta_{CB} - \Delta_{BC} < \min\{X_\sigma, Y_\sigma\}$, the more risky information can reduce the probability of both crises. Similarly, the following proposition shows the effect of the increase of the ambiguity on the cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$.

**Proposition 1.4.** There exists values $0 < X_\eta$ and $Y_\eta < 0$ such that

(i) $\frac{\partial \hat{\theta}_B}{\partial \eta} < 0$ if and only if $X_\eta < \Delta_{CB} - \Delta_{BC}$,

(ii) $\frac{\partial \hat{\theta}_C}{\partial \eta} < 0$ if and only if $Y_\eta < \Delta_{CB} - \Delta_{BC}$.

Proposition 1.4 implies that, when $\gamma_B = \gamma_C = 0$, $\frac{\partial \hat{\theta}_B}{\partial \eta} > 0$ and $\frac{\partial \hat{\theta}_C}{\partial \eta} < 0$. This is consistent with the finding by Ui (2015). However, in general, both $\frac{\partial \hat{\theta}_B}{\partial \eta}, \frac{\partial \hat{\theta}_C}{\partial \eta} < 0$ and $\frac{\partial \hat{\theta}_B}{\partial \eta}, \frac{\partial \hat{\theta}_C}{\partial \eta} > 0$ will occur depending on the parameters. This is a difference from the individual finial crisis case, where there is no feed-back effect through the two markets.

The implications from Proposition 1.3 and 1.4 are as follows. Since critical values $X_\sigma, Y_\sigma, X_\eta$ and $Y_\eta$ are determined by the parameters of the payoff and information quality, we can know whether the information disclosure policy with the less ambiguous and risky information is effective to prevent the financial crises.[15] Let $T = \Delta_{CB} - \Delta_{BC}$ be the difference of the two attraction values. Depending on the value of $T$, we can classify the all possible patterns into twelve types, which are summarized in Table 1.1.

Goldstein (2005) also discusses some implications focusing on the two points of policy measures that have a direct effect on the possibility of the one crisis and that have a direct effect on the possibility of the twin crises. Our results provide a new insight to the former point. Goldstein (2005) notes that raising the transaction cost

[15]Effects of additional information on the equilibrium cutoff is also studied by Iachan and Nenov (2015) and Inostroza and Pavan (2017) without ambiguity.

| $(\frac{d\hat{\theta}_B}{d\eta}, \frac{d\hat{\theta}_C}{d\eta}, \frac{d\hat{\theta}_B}{d\sigma}, \frac{d\hat{\theta}_C}{d\sigma})$ | $T > X_\sigma, Y_\sigma$ | $X_\sigma > T > Y_\sigma$ | $Y_\sigma > T > X_\sigma$ | $X_\sigma, Y_\sigma > T$ |
|---|---|---|---|---|
| $T > X_\eta$ | $(-,-,+,+)$ | $(-,-,-,+)$ | $(-,-,+,-)$ | $(-,-,-,-)$ |
| $X_\eta > T > Y_\eta$ | $(+,-,+,+)$ | $(+,-,-,+)$ | $(+,-,+,-)$ | $(+,-,-,-)$ |
| $Y_\eta > T$ | $(+,+,+,+)$ | $(+,+,-,+)$ | $(+,+,+,-)$ | $(+,+,-,-)$ |

Table 1.1: Characterization of the effect of noise, where $T = \Delta_{CB} - \Delta_{BC}$.

in the market C, which corresponds to the payoff from taking NA, [16] can reduce the possibility of the currency crisis and hence it also indirectly reduces the possibility of the bank run. By the similar argument, he states that the reduction of the short term return for withdrawing in the market B, which corresponds to the payoff from taking W, can reduce the possibility of the bank run and hence it also indirectly reduces the possibility of the currency crisis. Our results suggest a different policy. Suppose that $T > X_\sigma, Y_\sigma$ and $T > X_\eta$. In this case, we can say that the more ambiguous and less risky information can reduce the possibility of the twin crises because the indirect effect from the market B to the market C and vice versa gradually changes the cutoff points. As described in Table 1.1, we can know a optimal information disclosure policy depending on the situations.

## 1.5 Outcomes in the limit case

As in the usual global game analysis, we investigate what equilibrium outcome arises as $\sigma \to 0$ and $\eta \to 0$. Such an outcome can be seen as an approximation to the outcome in the complete information case. In contrast to the standard analysis, we must determine how the noise parameters $\sigma$ and $\eta$ vanish. Let us define $K \in [0, \infty)$ such that $\frac{\eta}{\sigma} \to K$. When $\sigma$ converges to 0 relatively faster than $\eta$, in which the noise consists almost entirely of the ambiguity, then $K$ can be higher. On the other hand, when $\eta$ converges to 0 relatively faster than $\sigma$, in which the noise consists almost entirely of the risk, then $K$ can be lower. Let $\xi = \lim_{\sigma\to 0, \eta\to 0} \Phi(\sqrt{2}\frac{\eta}{\sigma}) = \Phi(\sqrt{2}K)$ be the value of $\Phi(\sqrt{2}\frac{\eta}{\sigma})$ in the limit. Since $\frac{1}{2} \leq \Phi(x) \leq 1$ for each $x \geq 0$, we see that $\xi \in [\frac{1}{2}, 1]$. Again, when $\sigma$ ($\eta$) converges to 0 relatively faster than $\eta$ ($\sigma$), then $\xi$ will be higher (lower). Therefore, when we see that the complete information game is an

[16] If transaction cost increases, the relative payoff from NA also increases because the opportunity cost of taking A increases.

approximation of the incomplete information game where the risk is relatively higher than the ambiguity, we should consider the outcome in the case of the lower value of $\xi$. Similarly, when we see that the complete information game is an approximation of the incomplete information game where the ambiguity is relatively higher than the risk, we should consider the outcome in the case of the higher value of $\xi$. In this sense, we regard $\xi$ as an indicator of the environment which we want to see approximately.

Note that the limit of $\theta_B^*(\sigma,\eta)$, $\theta_B^{**}(\sigma,\eta)$, $\theta_C^*(\sigma,\eta)$, $\theta_C^{**}(\sigma,\eta)$ as $\sigma \to 0$ and $\eta \to 0$ can be written as follows:

$$\begin{aligned}
\theta_B^*(\sigma,\eta) &\to \frac{\beta_B}{\alpha_B}\xi \ (\sigma \to 0, \eta \to 0),\\
\theta_B^{**}(\sigma,\eta) &\to \frac{\gamma_B}{\alpha_B} + \frac{\beta_B}{\alpha_B}\xi \ (\sigma \to 0, \eta \to 0),\\
\theta_C^*(\sigma,\eta) &\to \frac{\beta_C}{\alpha_C} - \frac{\beta_C}{\alpha_C}\xi \ (\sigma \to 0, \eta \to 0),\\
\theta_C^{**}(\sigma,\eta) &\to \frac{\gamma_C}{\alpha_C} + \frac{\beta_C}{\alpha_C} - \frac{\beta_C}{\alpha_C}\xi \ (\sigma \to 0, \eta \to 0).
\end{aligned}$$

Let us denote $\theta_B^*(\xi) = \frac{\beta_B}{\alpha_B}\xi$, $\theta_B^{**}(\xi) = \frac{\gamma_B}{\alpha_B} + \frac{\beta_B}{\alpha_B}\xi$, $\theta_C^*(\xi) = \frac{\beta_C}{\alpha_C} - \frac{\beta_C}{\alpha_C}\xi$, $\theta_C^{**}(\xi) = \frac{\gamma_C}{\alpha_C} + \frac{\beta_C}{\alpha_C} - \frac{\beta_C}{\alpha_C}\xi$ by each value in the limit. Depending on the parameters, the regions $[\theta_B^*(\xi), \theta_B^{**}(\xi)]$ and $[\theta_C^*(\xi), \theta_C^{**}(\xi)]$ may overlap with each other or there is no intersection. The possible patterns are as follows:

$$\begin{aligned}
\text{(i) } &\theta_B^*(\xi) < \theta_C^*(\xi) < \theta_B^{**}(\xi) < \theta_C^{**}(\xi),\\
&\theta_B^*(\xi) < \theta_C^*(\xi) < \theta_C^{**}(\xi) < \theta_B^{**}(\xi),\\
&\theta_C^*(\xi) < \theta_B^*(\xi) < \theta_C^{**}(\xi) < \theta_B^{**}(\xi),\\
&\theta_C^*(\xi) < \theta_B^*(\xi) < \theta_B^{**}(\xi) < \theta_C^{**}(\xi).
\end{aligned}$$

$$\text{(ii) } \theta_B^*(\xi) < \theta_B^{**}(\xi) < \theta_C^*(\xi) < \theta_C^{**}(\xi).$$

$$\text{(iii) } \theta_C^*(\xi) < \theta_C^{**}(\xi) < \theta_B^*(\xi) < \theta_B^{**}(\xi).$$

Depending on each case, we obtain the following result, which corresponds to Proposition 2 of Goldstein (2005).

**Proposition 1.5.** As $\sigma \to 0, \eta \to 0$,
(i) $\hat{\theta}_B, \hat{\theta}_C \to \hat{\theta}(\xi)$ where $\hat{\theta}(\xi) = \frac{\alpha_B\gamma_C}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_B^* + \frac{\alpha_C\gamma_B}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_C^{**} = \frac{\alpha_B\gamma_C}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_B^{**} + \frac{\alpha_C\gamma_B}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_C^*$,
(ii) $\hat{\theta}_B \to \theta_B^{**}(\xi)$ and $\hat{\theta}_C \to \theta_C^*(\xi)$,
(iii) $\hat{\theta}_B \to \theta_B^*(\xi)$ and $\hat{\theta}_C \to \theta_C^{**}(\xi)$.

As Goldstein (2005) shows, when the region of each cutoff point intersects (case (i)), their limits are perfectly correlated with each other so that these values converge to the same value $\hat{\theta}(\xi)$. In other cases, even when there is no intersection, these points approach each other. By Proposition 1.5, we can see how the prediction changes as $\xi$ varies. In case (i), there are only two possibilities: either the two crises occur or none of them occurs. The effect of ambiguity on the cutoff point is $\hat{\theta}(\xi) - \hat{\theta}(\frac{1}{2}) = \frac{\Delta_{BC}-\Delta_{CB}}{\alpha_B\gamma_C+\alpha_C\gamma_B}(\xi - \frac{1}{2})$. This implies that, the ambiguity has a negative effect on the cutoff point ($\hat{\theta}(\xi) - \hat{\theta}(\frac{1}{2}) > 0$) if $\Delta_{BC} - \Delta_{CB} > 0$ and a positive effect on the cutoff point ($\hat{\theta}(\xi) - \hat{\theta}(\frac{1}{2}) < 0$) if $\Delta_{BC} - \Delta_{CB} < 0$.

In cases (ii) and (iii), there are three possibilities: the twin crises occur, the single crisis occurs, or none of them occurs. In case (ii), $\hat{\theta}_B = \theta_B^{**}$ increases and $\hat{\theta}_C = \theta_C^*$ decreases as $\xi$ increases. This implies that, as $\xi$ increases, the probability of the twin crises increases and, finally, the perfect correlation occur like case (i). In this sense, the ambiguity has a negative effect on both the twin crises and the currency crisis. In contrast, in case (iii), $\hat{\theta}_B = \theta_B^*$ increases and $\hat{\theta}_C = \theta_C^{**}$ decreases as $\xi$ increases. This implies that, as $\xi$ increases, the probability of the twin crises decreases and of the bank run increases. In this sense, the ambiguity has a positive effect on the twin crises but a negative effect on the bank run. We summarize such a comparative statics result about outcomes in the limit as follows.

**Proposition 1.6.**
(i) $\frac{d\hat{\theta}(\xi)}{d\xi} < 0$ if and only if $\Delta_{BC} < \Delta_{CB}$,
(ii) $\frac{d\theta_B^{**}(\xi)}{d\xi} > 0$ and $\frac{d\theta_C^*(\xi)}{d\xi} < 0$,
(iii) $\frac{d\theta_B^*(\xi)}{d\xi} > 0$ and $\frac{d\theta_C^{**}(\xi)}{d\xi} < 0$.

## 1.6 Conclusion

In this chapter, we study the following question: how the risky and ambiguous information affects the possibility of the bank run and the currency crisis when two financial markets are linked. We construct a simple global game model inspired by Goldstein (2005) and show that the effect of noise on the possibility of the financial crises can be different from the single financial crisis by Ui (2015). Our characterization of the effect of the risk and ambiguity on the possibility of the financial crises has some implications about an information disclosure policy to prevent the crises.

# Chapter 2 A Shapley Value Representation of Network Potentials

## 2.1 Introduction

A network game is introduced by Jackson and Wolinsky (1996) to study what networks emerge through self-interested agents' strategic interaction.[1] One of the solution concepts in network games is pairwise stability defined by Jackson and Wolinsky (1996). A network is pairwise stable if there is neither a player who prefers to sever a current link with his neighbor nor a pair of players who prefer to make a new link. It is known that a pairwise stable network exists if there exists a *network potential* (Jackson, 2003; Chakrabarti and Gilles, 2007), which can be considered as an imaginary representative player's payoff function.

A network potential is closely related to the Shapley value. Jackson (2003) shows that if each player's payoff function can be represented as the Shapley value of a network value function, which assigns a collective payoff generated by players to each network, then a network game admits a network potential, but it has been an open question whether or not the converse is also true.[2] On the other hand, Chakrabarti and Gilles (2007), who formally define a network potential, show that a network potential exists if a network payoff function satisfies the property called the Shapley value consistency, while they also demonstrate that its converse is not true: it is only a sufficient condition for the existence of the network potential.

This chapter provides a necessary and sufficient condition for the existence of network potentials in terms of the Shapley value and clarify the relationship between network potentials and the Shapley value. To this end, we introduce a collection of characteristic functions indexed by networks such that a collective payoff to a

[1] Examples include supply chain networks among firms (Palsule-Desai et al., 2013), FTAs among several countries (Goyal and Joshi, 2006; Furusawa and Konishi, 2007), and social network services (Farrell and Fudge, 2013). See Jackson (2008) for other examples.

[2] Jackson (2003) does not give the formal definition of network potentials.

coalition is determined by the sub-graph restricted to the coalition. We call it a network characteristic function, which has more degree of freedom than that of a network value function. More specifically, a value function is a function which assigns a real number to each network, whereas a network characteristic function is a function which assigns a real number to each pair of network and coalition. In this sense, it generalizes a network value function; that is, for any network value function, there exists a network characteristic function that represents the network value function, but not vice versa.

Our main result shows that there exists a network potential if and only if each player's payoff function can be represented as the Shapley value of a network characteristic function. This result is shown by the following three steps. First, we show that there exists a network potential if and only if each player's payoff function can be represented as an *interaction network potential*, which is a collection of functions indexed by coalitions such that each function assigns a collective payoff to each network restricted to the coalition. Second, we show that there is a one-to-one correspondence between an interaction network potential and a network characteristic function in the sense that a collection of the Möbius inverses of the characteristic functions corresponds to the interaction network potential. Finally, since the Shapley value of each characteristic function can be represented as the sum of its Möbius inverses proved by Shapley (1953), the above argument shows that the condition for the existence of a network potential is equivalent to the condition that each player's payoff function can be represented as the Shapley value of a network characteristic function. The argument of the proof is an application of Ui (2000) to network games, who shows that there exists a potential of noncooperative games by Monderer and Shapley (1996) if and only if each player's payoff function is represented as the Shapley value of a particular class of cooperative games indexed by strategy profiles.

Our result generalizes the result of Jackson (2003). As we mentioned above, Jackson (2003) shows that if each player's payoff function can be represented as the Shapley value of a network value function, then a network game admits a network potential. Because a network value function is a special case of a network characteristic function, our result provides an alternative proof for that of Jackson (2003). Moreover, we can show that the converse of Jackson (2003)'s result is not true by constructing a network characteristic function that cannot be represented as a network value function.

Our result also generalizes the result of Chakrabarti and Gilles (2007). As we mentioned above, Chakrabarti and Gilles (2007) show that a network potential exists if a network payoff function satisfies the property called the Shapley value consistency. If each player's payoff function satisfies the Shapley value consistency, we can show that it can be represented as the Shapley value of a network characteristic function. Therefore, our result also provides an alternative proof for that of Chakrabarti and Gilles (2007).

Except for our result, there is another characterization result for the existence of a network potential by Chakrabarti and Gilles (2007) in terms of potential games. Chakrabarti and Gilles (2007) show that a network potential exists if and only if the corresponding game, called Myerson's consent game (Myerson, 1991), is a potential game.[3] A drawback of this result is that it requires an indirect step to identify the existence of a network potential; that is, we need additional results to check whether the corresponding game is a potential game or not. In contrast, we provide a direct condition to identify the existence of a network potential. Moreover, combining the result of Ui (2000) with ours, we can derive the characterization of Chakrabarti and Gilles (2007) as a byproduct and show the coincidence of potentials in the different class of games: network games, noncooperative games, and cooperative games.

The rest of this chapter is organized as follows. In section 2.2, we define a formal model. In section 2.3, we show our main results. In section 2.4, we discuss some examples to show how our results are useful to find a network potential. Section 2.5 concludes the chapter. Omitted proofs are relegated to Appendix B.

## 2.2 Model

### 2.2.1 Setup

Let $N = \{1, \cdots, n\}$ be a (finite) set of players. A network is described by an undirected graph whose nodes are players. Let $g^N = \{ij | i, j \in N, i \neq j\}$ be a set of all possible links. Then, a network $g$ is a subset of $g^N$. We denote the set of all networks by $\mathbb{G}^N = \{g | g \subset g^N\}$. For each network $g \in \mathbb{G}^N$ and player $i \in N$, let

[3] Myerson's consent game is a network formation game such that each player's action is to choose the set of other players with whom he wants to make links and each player's payoff function depends upon the constructed network. See Section 2.3 for the formal definition and discussion.

$N_i(g) = \{j \in N | i \neq j \text{ and } ij \in g\}$ be the set of $i$'s neighborhood in $g$. For each $S \in 2^N$ and $g \in \mathbb{G}^N$, let $g|_S = \{ij \in g | i \in S \text{ and } j \in S\}$ be a restricted network whose nodes are in $S$. We denote by $\mathbb{G}^S$ the set of networks where the set of players is $S$. For each $ij \in g$, let $g - ij = g \backslash \{ij\}$ be the network which remains after removing a link $ij$ from $g$. Similarly, for each $ij \notin g$, let $g + ij = g \cup \{ij\}$ be the network formed by adding a link $ij$ to $g$. The payoff function for player $i \in N$ is denoted by $\phi_i : \mathbb{G}^N \to \mathbb{R}$. Jackson and Wolinsky (1996) call $\phi = (\phi_i)_{i \in N}$ a *network game*.

A solution concept on network games is *pairwise stability* defined by Jackson and Wolinsky (1996). A network is pairwise stable if there is neither a player who wants to sever the link with his neighbor nor a pair of players who agree to make a new link.

**Definition 2.1.** A network $g$ is pairwise stable if

(i) for all $ij \in g, \phi_i(g) \geq \phi_i(g - ij)$ and $\phi_j(g) \geq \phi_j(g - ij)$, and

(ii) for all $ij \notin g$, if $\phi_i(g) < \phi_i(g + ij)$ then $\phi_j(g) > \phi_j(g + ij)$.

A sufficient condition for the existence of a pairwise stable network is the existence of a network potential function defined by Chakrabarti and Gilles (2007). The function is analogous to the potential function defined by Monderer and Shapley (1996) in noncooperative games.

**Definition 2.2.** A network game $\phi = (\phi_i)_{i \in N}$ admits a *network potential* if there is a function $\omega : \mathbb{G}^N \to \mathbb{R}$ such that, for any $g \in \mathbb{G}^N$ and $ij \in g$,

$$\phi_i(g) - \phi_i(g - ij) = \omega(g) - \omega(g - ij).$$

If a network game $\phi$ admits a network potential function $\omega$, it is known that its maximizers are pairwise stable networks. Since there are a finite number of networks, a maximizer of $\omega$ always exists, which implies the existence of a pairwise stable network. The following result summarizes this observation.

**Proposition 2.1.** Suppose that a network game $\phi = (\phi_i)_{i \in N}$ admits a network potential function $\omega$. Then, there is at least one pairwise stable network. Moreover, a maximizer of $\omega$ is a pairwise stable network.

### 2.2.2 Symmetric interaction on networks

To illustrate how to find a network potential, consider the following network game. Let $N = \{1, 2, 3\}$ be the set of players. We assume that payoff functions are as follows: for each $S \in 2^N$, there exists a function $w_S : \mathbb{G}^S \to \mathbb{R}$ such that

$$\phi_i(g) = w_{\{i\}}(\emptyset) + \sum_{j \neq i} w_{\{i,j\}}(g|_{\{i,j\}}) + w_N(g).$$

We call $\phi$ a *symmetric interaction* network game (SI network game) in the sense that each player's payoff function is described by symmetric bilateral interaction terms and a total interaction term.

Let us define the function $\omega : \mathbb{G}^N \to \mathbb{R}$ such that

$$\omega(g) = \sum_{i \in N} w_{\{i\}}(\emptyset) + \sum_{i<j} w_{\{i,j\}}(g|_{\{i,j\}}) + w_N(g).$$

Then, for each $i, j \in N$ with $ij \in g$,

$$\begin{aligned}\phi_i(g) - \phi_i(g - ij) &= (w_{\{i,j\}}(g|_{\{i,j\}}) - w_{\{i,j\}}((g - ij)|_{\{i,j\}}))) + (w_N(g) - w_N(g - ij)) \\ &= \omega(g) - \omega(g - ij).\end{aligned}$$

Therefore, a SI network game admits a network potential function. We will show that a condition of the existence of a network potential function is equivalent to the condition that each player's payoff function $\phi_i$ can be decomposed into symmetric interaction terms such as a SI network game.

## 2.3 Representation theorem

### 2.3.1 Cooperative games

To state our results, we prepare some concepts of cooperative games. A characteristic function or a TU game is defined as a function $v : 2^N \to \mathbb{R}$ such that $v(\emptyset) = 0$. We denote the set of all TU games by $\mathcal{G}_N$.

For each $T \in 2^N$, a unanimity game $u_T \in \mathcal{G}_N$ is defined as

$$u_T(S) = \begin{cases} 1 & \text{if } T \subset S, \\ 0 & \text{otherwise.} \end{cases}$$

Shapley (1953) shows that any TU game $v \in \mathcal{G}_N$ can be represented as a unique linear combination of a collection of the unanimity games $\{u_T\}_{T \in 2^N}$, i.e., $v(S) = \sum_{T \in 2^N} v^T u_T(S)$ where $v^T = \sum_{R \subset T} (-1)^{|T \setminus R|} v(R)$ is called the Möbius inversion of $v$.

For $v \in \mathcal{G}_N$ and $T \in 2^N$, a restricted game $v|_T \in \mathcal{G}_N$ is defined as

$$v|_T(S) = \begin{cases} v(S \cap T) & \text{if } S \cap T \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

The Shapley value is defined as a map $\psi : \mathcal{G}_N \to \mathbb{R}^N$ such that

$$\psi_i(v) = \sum_{S \in 2^N, i \in S} \frac{(|S|-1)!(|N|-|S|)!}{|N|!} (v(S) - v(S \setminus \{i\}))$$

for each $i \in N$.

It is known that $\psi$ is a linear map which satisfies

$$\psi_i(u_T) = \begin{cases} 1/|T| & \text{if } i \in T, \\ 0 & \text{otherwise.} \end{cases}$$

By the decomposition result of Shapley (1953), we can write $\psi_i(v) = \sum_{T \in 2^N, i \in T} v^T/|T|$ where $v^T/|T|$ is called *Harsanyi's dividend* to the member of $T$.

In TU games, Hart and Mas-Colell (1989) define a potential function. For a function $P : \mathcal{G}_N \to \mathbb{R}$, the marginal contribution of player $i$ to $P$ is defined as

$$DP_i(v) = P(v) - P(v|_{N \setminus \{i\}}).$$

Then, $P$ is a *potential* if it satisfies

$$\sum_{i \in N} DP_i(v) = v(N).$$

Hart and Mas-Colell (1989) show that the potential $P$ is uniquely given by $P(v) = \sum_{T \in 2^N} v^T/|T|$. Thus, each player's marginal contribution satisfies

$$DP_i(v) = \sum_{T \in 2^N, i \in T} v^T/|T| = \psi_i(v).$$

We consider a collection of TU games indexed by networks $\{v_g\}_{g \in \mathbb{G}^N}$ such that $v_g(S) = v_{g'}(S)$ if $g|_S = g'|_S$. We call $\{v_g\}_{g \in \mathbb{G}^N}$ a network characteristic function or a *TU game on networks*. Note that the value of a coalition $S \in 2^N$ is determined by the network structure in $S$ and not by the network structure of $N\backslash S$. We denote by $\mathcal{G}_{N,\mathbb{G}^N} = \{\{v_g\}_{g\in\mathbb{G}^N} | v_g(S) = v_{g'}(S) \text{ if } g|_S = g'|_S\}$ the set of all TU games on networks. The following lemma shows a property of Harsanyi's dividends of $\{v_g\}_{g\in\mathbb{G}^N}$.

**Lemma 2.1.** $\{v_g\}_{g\in\mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$ if and only if $g|_S = g'|_S$ implies $v_g^S = v_{g'}^S$ for any $S \in 2^N$.

### 2.3.2 Main results

The goal of this section is to show a relationship between network potentials and the Shapley value. Let us consider a collection of functions $\{\zeta_S\}_{S\in 2^N}$ such that $\zeta_S : \mathbb{G}^S \to \mathbb{R}$. We call a collection of functions an *interaction network potential*, which is analogous to the definition of an *interaction potential* defined by Ui (2000). Our main result is the following representation theorem.

**Theorem 2.1.** For any network game $\phi = (\phi_i)_{i\in N}$, the following statements are equivalent:

(i) The network game $\phi$ admits a network potential.

(ii) There exists a TU game on networks $\{v_g\}_{g\in\mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$ such that

$$\phi_i(g) = \psi_i(v_g) \text{ for all } i \in N.$$

(iii) There exists an interaction network potential $\{\zeta_S\}_{S\in 2^N}$ such that

$$\phi_i(g) = \sum_{S\in 2^N, i\in S} \zeta_S(g|_S) \text{ for all } i \in N.$$

Furthermore, a network potential function $\omega$ is given by

$$\omega(g) = P(v_g) = \sum_{S\in 2^N} \zeta_S(g|_S).$$

Theorem 2.1 shows that a SI network game in the three player case in section 2.2 is a special case of the network games which have a network potential. A family of the functions $w_S : \mathbb{G}^S \to \mathbb{R}$ for each $S \in 2^N$ is an interaction network potential.

### 2.3.3 Relation with other results

In this subsection, we discuss the relation between our results and previous results by Jackson (2003) and Chakrabarti and Gilles (2007). Jackson (2003) and Chakrabarti and Gilles (2007) show a sufficient condition for the existence of network potentials.

We show that both of results are corollaries of Theorem 2.1. Moreover, we provide an alternative simple proof of a characterization result of the existence of network potentials by Chakrabarti and Gilles (2007) in terms of potential games.

**Myerson-Jackson-Wolinsky value**. Jackson and Wolinsky (1996) consider the following network game. Let $\tilde{v} : \mathbb{G}^N \to \mathbb{R}$ be a network value function. We denote by $\tilde{\mathcal{G}}_N$ the set of all network value functions. Unlike a TU-game, the value of $\tilde{v}$ depends on the networks rather than coalitions. We say that $\tilde{v}$ is component additive if $\sum_{h \in C(g)} \tilde{v}(h) = \tilde{v}(g)$ where $C(g)$ is the set of a connected components of $g$. Let $\Pi(g) \subset 2^N$ be the set of coalitions such that each player is connected in $g$. A mapping $f : \tilde{\mathcal{G}}_N \times \mathbb{G}^N \to \mathbb{R}^N$ is called an *allocation rule*. Jackson and Wolinsky (1996) define the allocation rule, which we call the *Myerson-Jackson-Wolinsky value*, such that, for each $i \in N$,

$$f_i^{MJW}(\tilde{v}, g) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (\tilde{v}(g|_{S \cup \{i\}}) - \tilde{v}(g|_S)).$$

Like Myerson (1977), Jackson and Wolinsky (1996) show that $f^{MJW}(\tilde{v}, g)$ is the unique allocation rule satisfying following properties.

Component balance (CB): for any component additive $\tilde{v}$, $g \in \mathbb{G}^N$, and $S \in \Pi(g)$, $\sum_{i \in S} f_i(\tilde{v}, g) = \tilde{v}(g|_S)$.

Equal bargaining power (EBP): for any component additive $\tilde{v}$, for any $g \in \mathbb{G}^N$ and any $ij \in g$, $f_i(\tilde{v}, g) - f_i(\tilde{v}, g - ij) = f_j(\tilde{v}, g) - f_j(\tilde{v}, g - ij)$.

In this setup, Jackson (2003) shows the following result, which is implied by Theorem 2.1.

**Corollary 2.1.** (Proposition 2 of Jackson, 2003). Suppose that a network game $\phi = (\phi_i)_{i \in N}$ satisfies $\phi_i(g) = f_i^{MJW}(\tilde{v}, g)$ for some $\tilde{v} \in \tilde{\mathcal{G}}_N$. Then, a network game $\phi$ admits a network potential $\omega(g) = \sum_{S \subset N} \frac{(|S|-1)!(|N|-|S|)!}{|N|!} (\tilde{v}(g|_S))$.

This result states that network games where each player's payoff function is given by $f_i^{MJW}(\tilde{v}, g)$ admit network potentials. A key idea behind Corollary 2.1 is that, for each network value function $\tilde{v}$, there is a network characteristic function $\{\tilde{v}_g\}_{g\in\mathbb{G}^N}$ such that $\tilde{v}_g(S) = \tilde{v}(g|_S)$ for all $S \in 2^N$ and $g \in \mathbb{G}^N$. However, the converse of the result is not true in general. To see this, consider the following network game. Each player's payoff function is given in Table 2.1.

| Network | $\phi_1(g)$ | $\phi_2(g)$ | $\phi_3(g)$ | $\omega(g)$ |
|---|---|---|---|---|
| $g_0 = \emptyset$ | 0 | 0 | 0 | 0 |
| $g_1 = \{12\}$ | 5/6 | 5/6 | 2/6 | 5/6 |
| $g_2 = \{13\}$ | 1 | 0 | 1 | 1 |
| $g_3 = \{23\}$ | $-1$ | 9/6 | 9/6 | 9/6 |
| $g_4 = \{12, 13\}$ | 11/6 | 5/6 | 8/6 | 11/6 |
| $g_5 = \{12, 23\}$ | 7/6 | 22/6 | 19/6 | 22/6 |
| $g_6 = \{13, 23\}$ | $-2/6$ | 7/6 | 13/6 | 13/6 |
| $g_7 = g^N$ | 13/6 | 22/6 | 25/6 | 28/6 |

Table 2.1: Payoff functions of the network game.

Note that this game admits a network potential function $\omega$. By Theorem 2.1, there is a TU game on networks $\{v_g\}_{g\in\mathbb{G}^N}$ such that $\phi_i(g) = \psi_i(v_g)$. By *efficiency* of the Shapley value, $v_g(N)$ must satisfy the following: $v_{g_0}(N) = 0, v_{g_1}(N) = 2, v_{g_2}(N) = 2, v_{g_3}(N) = 2, v_{g_4}(N) = 4, v_{g_5}(N) = 8, v_{g_6}(N) = 3, v_{g_7}(N) = 10$.[4] An example of $\{v_g\}_{g\in\mathbb{G}^N}$ is given in Table 2.2.

| $v_g \backslash S$ | $\{1\}$ | $\{2\}$ | $\{3\}$ | $\{1,2\}$ | $\{1,3\}$ | $\{2,3\}$ | $N$ |
|---|---|---|---|---|---|---|---|
| $v_{g_0}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{g_1}$ | 0 | 0 | 0 | 1 | 0 | 0 | 2 |
| $v_{g_2}$ | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| $v_{g_3}$ | 0 | 0 | 0 | 0 | 0 | 5 | 2 |
| $v_{g_4}$ | 0 | 0 | 0 | 1 | 2 | 0 | 4 |
| $v_{g_5}$ | 0 | 0 | 0 | 1 | 0 | 5 | 8 |
| $v_{g_6}$ | 0 | 0 | 0 | 0 | 2 | 5 | 3 |
| $v_{g_7}$ | 0 | 0 | 0 | 1 | 2 | 5 | 10 |

Table 2.2: An example of $\{v_g\}_{g\in\mathbb{G}^N}$ such that $\phi_i(g) = \psi_i(v_g)$.

If there is a network value function $\tilde{v}$ which satisfies $\phi_i(g) = f_i^{MJW}(\tilde{v}, g)$, then

[4] Shapley value satisfies the efficiency such that $\sum_{i\in N} \psi_i(v) = v(N)$ for all $v \in \mathcal{G}_N$.

$\tilde{v}(g) = \tilde{v}(g|_N) = \tilde{v}_g(N) = v_g(N)$ because $f_i^{MJW}(\tilde{v}, g) = \psi_i(\tilde{v}_g)$ and the efficiency of $\psi_i(\tilde{v}_g)$. However, if it is true, we have $f_1^{MJW}(\tilde{v}, g_1) = f_1^{MJW}(\tilde{v}, g_1) = 1$ and $f_3^{MJW}(\tilde{v}, g_1) = 0$ because

$$\tilde{v}(g_1|_S) = \begin{cases} 2 & \text{if } S = \{1,2\} \text{ or } N, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, there is no value function $\tilde{v}$ which satisfies $\phi_i(g) = f_i^{MJW}(\tilde{v}, g)$ although this game admits a network potential.

A reason behind this observation is that the number of degrees of freedom for a value function is smaller than that of a network characteristic function because the former is a function from $\mathbb{G}^N$ to $\mathbb{R}$, whereas the latter is a function from $\mathbb{G}^N \times 2^N$ to $\mathbb{R}$.

**Shapley value consistency**. For a network game $\phi$, Chakrabarti and Gilles (2007) consider the following TU game: for each $g \in \mathbb{G}^N$, let us define $U_{\phi,g} : 2^N \to \mathbb{R}$ such that

$$U_{\phi,g}(S) = \sum_{i \in S} \phi_i(g|_S)$$

for all $S \in 2^N$.

We say that a network game $\phi$ is *Shapley value consistent* if for every $g \in \mathbb{G}^N$, it holds that $\phi_i(g) = \psi_i(U_{\phi,g})$. They show the following result, which is also implied by Theorem 2.1.

**Corollary 2.2.** (Theorem 3.7 of Chakrabarti and Gilles, 2007). If a network game $\phi$ is Shapley value consistent, then a network game $\phi$ admits a network potential.

Chakrabarti and Gilles (2007) show an example where a network potential exists although $\phi$ is not Shapley value consistent.[5] If $\phi$ is Shapley value consistent, then the family of games $\{U_{\phi,g}\}_{g \in \mathbb{G}^N}$ is a TU game on networks. However, for the existence of a network potential, Theorem 2.1 says that we need TU games on networks, which is not necessary Shapley value consistent.

[5]See Appendix B.6 for the formal argument. In the example, we demonstrate that there is a network characteristic function $\{v_g\}_{g \in \mathbb{G}^N}$ which satisfies the condition of Theorem 2.1, whereas the network game is not Shapley value consistent.

**Potential games**. For each $i \in N$, let $A_i$ be the set of strategies and $u_i : A \to \mathbb{R}$ be the payoff function where $A = A_1 \times \cdots \times A_n$. We denote $(N, A, u)$ a strategic form game. Monderer and Shapley (1996) define the class of potential games.

**Definition 2.3.** A game $(N, A, u)$ is called a *potential game* if there is a function $V : A \to \mathbb{R}$ such that, for each $i \in N, a'_i \in A_i$ and $a \in A$,

$$u_i(a'_i, a_{-i}) - u_i(a) = V(a'_i, a_{-i}) - V(a).$$

Let us consider a collection of TU games $\{v_a\}_{a\in A}$ such that $v_a(S) = v'_a(S)$ if $a_S = a'_S$, which is called a *TU game with action choices*. We denote by $\mathcal{G}_{N,A} = \{\{v_a\}_{a\in A} | v_a(S) = v'_a(S) \text{ if } a_S = a'_S\}$ the set of all TU games with action choices. Ui (2000) shows the following relationship between potential games and the Shapley value, and hence the potential of a TU game.

**Theorem 2.2.** (Theorem 2 of Ui, 2000). For any game $(N, A, u)$, the following statements are equivalent:

(i) $(N, A, u)$ is a potential game.

(ii) There exists a TU game with action choices $\{v_a\}_{a\in A} \in \mathcal{G}_{N,A}$ such that

$$u_i(a) = \psi_i(v_a) \text{ for all } i \in N.$$

Furthermore, a potential function $V$ is given by

$$V(a) = P(v_a).$$

We consider the following network formation game by Myerson (1991) to state another characterization result of the existence of a network potential in terms of potential games by Chakrabarti and Gilles (2007). This game is called a *consent game*. Given a network game $\phi = (\phi_i)_{i\in N}$, let $A_i = \{(l_{ij})_{j\neq i} | l_{ij} \in \{0, 1\}\}$ be the set of actions for player $i$. We denote by $l_i = (l_{ij})_{j\neq i}$ a typical element of $A_i$. Let $\sigma(l) = \{ij \in g^N | l_{ij} \cdot l_{ji} = 1\}$ be an network induced by the action profile $l = (l_i)_{i\in N}$. We call $(N, A, \pi_\phi)$ a consent game corresponding a network game $\phi$ if $\pi_{\phi,i}(l) = \phi_i(\sigma(l))$. Let $A_g = \{l \in A | \sigma(l) = g\}$ be the set of strategy profiles which induce the network $g$. Note that each strategy profile $l$ induces the unique network $\sigma(l)$, but there are many strategy profiles which induce the same network. We define $\hat{l}_g \in A_g$ as the (unique) *non-superfluous* strategy profile such that, for all pair $i, j \in N$, $\hat{l}_{g,ij} = 1$ if and only if $ij \in g$. Chakrabarti and Gilles (2007) show the following characterization result. By using Theorem 2.1 and 2.2, we can provide much simpler proof than the original proof.

**Theorem 2.3.** (Theorem 3.3 of Chakrabarti and Gilles, 2007). A network game $\phi = (\phi_i)_{i \in N}$ admits a network potential if and only if the corresponding consent game $(N, A, \pi_\phi)$ is a potential game.

According to this result, we have to consider whether the corresponding consent game is a potential game or not to check the existence of a network potential. In contrast, Theorem 2.1 can be directly applied to the underlying network game $\phi$.

Combing Theorem 2.1, 2.2 and 2.3, we can say the following unified relationship among potential functions of different types of games: network games, cooperative games, and noncooperative games.

**Corollary 2.3.** A network game $\phi$ admits a network potential $\omega(\cdot)$ if and only if there exists a TU game on networks $\{v_g\}_{g \in \mathbb{G}^N} \in \mathcal{G}_{N, \mathbb{G}^N}$ such that $\omega(g) = P(v_g) = V(\hat{l}_g)$ where $V(\cdot)$ is a potential function of the corresponding consent game $(N, A, \pi_\phi)$.

## 2.4 Examples

In this section, we consider the following examples to demonstrate the applicability of theorem 2.1 to find a network potential. For each network $g \in \mathbb{G}^N$ and for each distinct $i, j \in N$, let

$$g_{ij} = \begin{cases} 1 & \text{if } ij \in g, \\ 0 & \text{otherwise.} \end{cases}$$

**Example 2.1** (Social distance (Iijima and Kamada, 2016).[6] Suppose that each player $i$ has a multidimensional ($m$-dimension) characteristic, which is called type, $x_i = (x_{i1}, \cdots, x_{im}) \in [0,1]^m$. Let $d : X \times X \to \mathbb{R}$ represent a metric, called *social distance*, which measures the relation/similarity among players. They assume that each player's payoff function is as follows:

$$\phi_i(g) = \left( \sum_{j \in N_i(g)} b(d(x_i, x_j)) \right) - c(q_i),$$

[6] We thank Ryota Iijima for suggesting this model.

where $b(\cdot) > 0$ is a weakly decreasing, left-continuous function, and $q_i = |N_i(g)|$. Assume that $c(\cdot)$ is linear i.e., $c(q_i) = cq_i$ for some constant $c > 0$.[7]

Let us define

$$\zeta_S(g|_S) = \begin{cases} (b(d(x_i, x_j)) - c)g_{ij} & \text{if } S = \{i, j\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, an interaction network potential $\{\zeta_S\}_{S \in 2^N}$ satisfies

$$\phi_i(g) = \sum_{S \in 2^N, i \in S} \zeta_S(g|_S)$$

for all $i \in N$.

Thus, by Theorem 2.1, this network game $\phi$ admits a network potential and it is given by

$$\omega(g) = \sum_{i<j} (b(d(x_i, x_j)) - c)g_{ij} = \frac{1}{2} \sum_{i \in N} \phi_i(g).$$

By Proposition 2.1, a pairwise stable network exists. Moreover, for a network $g \notin \text{argmax}_{g' \in \mathbb{G}^N} w(g')$, (i) there is a pair $ij \in g$ such that $b(d(x_i, x_j)) < c$ or (ii) there is a pair $ij \notin g$ such that $b(d(x_i, x_j)) > c$, which implies that $g$ is not pairwise stable. Therefore, $g$ is pairwise stable if and only if $g \in \text{argmax}_{g' \in \mathbb{G}^N} w(g') = \text{argmax}_{g' \in \mathbb{G}^N} \sum_{i \in N} \phi_i(g')$, which implies that a pairwise stable network $g$ is efficient.[8] This argument shows the following.

**Corollary 2.4.** Suppose that a cost function is linear. Then, a pairwise stable network $g$ exists and is efficient.

We should remark that the concept of pairwise stability employed by Iijima and Kamada (2016), which we call IK-pairwise stability, is slightly stronger than the usual one, Definition 2.1. Their definition of the first part is identical to that of Definition 2.1, but the second part is as follows: (ii') for all $ij \notin g$, if $\phi_i(g) \leq \phi_i(g + ij)$ then $\phi_j(g) > \phi_j(g + ij)$. Generically in payoff, these two definitions are equivalent. However, if there is a pair $ij$ such that $b(d(x_i, x_j)) = c$, this concept is strictly stronger requirement for a network to be stable. In such cases, only the IK-pairwise stable network is the maximal element of $\text{argmax}_{g' \in \mathbb{G}^N} w(g')$, which is unique. Let $\hat{d}$ be

[7] Iijima and Kamada (2016) consider the general case, in which $c(\cdot)$ is a strictly increasing function, but they mainly discuss this linear cost case.

[8] A network $g$ is said to be efficient if $g \in \text{argmax}_{g' \in \mathbb{G}^N} \sum_{i \in N} \phi_i(g')$.

the maximal value of $d(x_i, x_j)$ such that $b(d(x_i, x_j)) \geq c$, which exists because of the properties of $b(\cdot)$. Hence, $g = \{ij | d(x_i, x_j) \leq \hat{d}\}$ is the unique IK-pairwise stable network. By Corollary 2.4 and this argument, we obtain the following characterization of IK-pairwise stability by Iijima and Kamada (2016).

**Corollary 2.5.** (Lemma 1 of Iijima and Kamada, 2016). Suppose that a cost function is linear. Then, an IK-pairwise stable network $g$ exists and it is unique and efficient. Furthermore, $g$ is generated by a cutoff rule with a homogeneous cutoff value profile.[9]

**Example 2.2** (Games on networks). In the study of games on networks, we assume that each player has a payoff function, which depends not only on the action profile but also the network. The typical examples are represented by the following quadratic payoff function $u_i : A \times \mathbb{G}^N \to \mathbb{R}$ such that

$$u_i(a, g) = \alpha_i a_i - \frac{1}{2} a_i^2 + \psi a_i \sum_{j \neq i} g_{ij} a_j - \gamma a_i \sum_{j \neq i} a_j,$$

where $a_i \in A_i = \mathbb{R}$ is an action for $i \in N$ and $\alpha_i, \psi, \gamma \in \mathbb{R}$ are parameters.[10]

Note that, given network $g \in \mathbb{G}^N$, the game $\Big(N, A, \big(u_i(\cdot, g)\big)_{i \in N}\Big)$ is a potential game with a potential function $V(a) = \sum_{i \in N} \big(\alpha_i a_i - \frac{1}{2} a_i^2\big) + \sum_{i<j} \big(\psi g_{ij} - \gamma\big) a_i a_j$.

Given action profile $a = (a_i)_{i \in N}$, let us define

$$\zeta_S(g|_S) = \begin{cases} \psi g_{ij} a_i a_j & \text{if } S = \{i, j\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, an interaction network potential $\{\zeta_S\}_{S \in 2^N}$ satisfies

$$\phi_i(g) = \sum_{S \in 2^N, i \in S} \zeta_S(g|_S)$$

for all $i \in N$.

Thus, by Theorem 2.1, this network game $\big(u_i(a, \cdot)\big)_{i \in N}$ admits a network potential and it is given by

$$\omega(g) = \psi \sum_{i<j} g_{ij} a_i a_j.$$

[9] $g$ is generated by a cutoff rule with $(\hat{d}_1, \cdots, \hat{d}_n) \in \mathbb{R}_+$ if $ij \in g \Leftrightarrow d(i, j) \leq \min\{\hat{d}_i, \hat{d}_j\}$. A cutoff rule is homogeneous if $\hat{d}_i = \hat{d}_j$ for any $i, j \in N$.

[10] See chapter 4 of Jackson and Zenou (2014) for detailed surveys.

Therefore, given action profile $a = (a_i)_{i \in N}$, a pairwise stable network exists by Proposition 2.1. The argument here will be useful to consider what network is stable and what action profile is an equilibrium simultaneously like Staudigl (2011) and Hsie et al. (2016), who consider the dynamics where each player can change both a link in a network and an action simultaneously.

## 2.5 Conclusion

In this chapter, we provide a characterization for network potentials in terms of the Shapley value and potentials in cooperative games. To show this, we introduce the new concept: interaction network potentials. Our representation theorem shows that a network potential exists if and only if each player's payoff function can be decomposed into the sum of the symmetric interaction terms, which coincides with the Shapley value of a TU-game on networks.

A network potential function is useful to discuss the plausibility of stable networks. If there are many stable networks, focusing on the stability is not sufficient to obtain the sharp prediction about what network emerge.[11] In the noncooperative games, this problem raises a huge literature of the equilibrium selection. Among them, there are two prominent approaches: one is a dynamic process approach and the other one is an incomplete information approach. In both approaches, it is known that the strategy profile that maximizes the potential function is selected in many cases.[12] We believe that a network potential function can be useful to consider a research program of the "stable network selection".

[11] In general, the set of stable networks is large. For instance, see Jackson and Watts (2002).

[12] There are many papers showing that the maximizers of the potential function survived from the equilibrium selections. In the dynamic process approach, Blume (1993) shows that strategy profiles that maximize the potential function have the highest probabilities in the stationary distribution of the stochastic evolutionary process, which is a version of Kandori et al. (1993) and Young (1993). Hofbauer and Sorger (1999) show that the maximizer of the potential function has stable properties in the perfect foresight dynamics by Matsui and Matsuyama (1995). In the incomplete information approach by Kajii and Morris (1997), Ui (2001) shows that the unique potential maximizer has a robust property.

# Chapter 3 Two-sided Matching under Irrational Choice Behavior

## 3.1 Introduction

Matching theory is one of the most influential theories recently. Designs of an assignment mechanism in labor markets (i.e., a matching between hospitals and medical students), in school choice problems (i.e., a matching between public schools and students), and in marriage markets (i.e., a matching between men and women) are typical applications. In the standard analysis, we assume that all agents are rational in the sense that they have complete and transitive strict preference relations over the other side. Under this rationality assumption, Gale and Shapley (1962) show that we can always obtain a desirable outcome called *stable matching* by using the deferred-acceptane algorithm (DA algorithm). Moreover, the DA algorithm has good properties like strategy-proofness and one-side optimality.[1]

To use the DA algorithm, the assumption of agents' rationality is indispensable. However, there is ample evidence in psychology and behavioral economics literature that the rationality assumption is doubtful. Taking an example in the context of the school choice problem, it is natural that some attributes like the commute time, the location, or the curriculum are considered simultaneously. Then, people sometimes consider the school A is better than the school B and the school B is better than the school C in terms of the location but, in terms of the curriculum, they may consider the school C is better than the school A. As a result, the preference over schools can be intransitive.

It is also natural that market participants in the marriage market have irrational preferences. An acceptable reason is that, like on-line matching services, people cannot know full information about all other participants. Even at a speed dating event, we cannot construct a rational preference if the number of the participants are large enough. Indeed, most on-line matching services do not require a rational

[1] For classical results, see Roth and Sotomayor (1992). For recent surveys, see Roth (2008) and Kojima and Troyan (2011).

preference of each user over the other side but require general favorable characteristics about potential partners like, age, income, educational background, living place, and other social status.[2] In spite of such evidence, we cannot say even whether the stable matchings exist or not without the assumption of agents' rationality.

Given these observations, our purpose of this chapter is to propose a way to treat irrational preferences in the two-sided matching problem. We first show that if we assume that each agent's preference is acyclic, we can guarantee the existence of a stable matching by extending the DA algorithm. Our proposed algorithm is divided into two steps. The first step is based on the completion of preference relations by way of Suzumura (1976). That is, if the preference is acyclic, we can extend it to a rational preference. This seems a natural way to treat irrational preferences as the same flavor of using a tie-breaking rule in the case of weak preferences. However, his proof is not constructive and any explicit construction is not known in general.

To overcome this difficulty, we develop a completion algorithm to construct an extended order. In the algorithm, we focus on a directed graph induced by each agent's preference. If an agent's preference is acyclic, we can construct a directed graph on the set of choice alternatives with respect to the preference relation. Owing to this structure, we can construct a *pseudo-ranking* in the sense that we identify each choice alternative with the unique number and then we can put each alternative into the same ranked set. This ranking constitutes a partition of the set of choice alternatives. Finally, we construct a strict order by picking elements from the partition in an appropriate manner. In the second step, we simply use the DA algorithm in terms of the constructed rational preferences in the first step. We show that the final matching induced from this two-step algorithm is stable in terms of the reported (not necessary rational) preference profile.

In addition to stability, this two-step algorithm is stratgy-proof. Despite of such merits, we can show that the DA algorithm with any completion rule sometimes cannot induce a (wo)man-optimal matching. Since the DA algorithm is a unique rule which satisfies the stability and the strategy-proofness shown by Dubins and

[2]One example is Match.com, which is the most biggest on-line matching service over the world. In the services, the participants publish their own profile to the web site and, then, they search for a favorite partner based on this information. The other example is eHarmony, which is also a widely used on-line matching service and has many users over countries. In the services, the participants answer questions about desirable partners and, then, based on the answers, an algorithm of the service recommends matching partners.

Freedman (1981), Roth (1982), and Alcalde and Barberà (1994), this observation implies the following impossibility result: whenever we use a usual algorithm beyond the rational preference domain with a completion rule, strategy-proofness and one-side optimality are incompatible. Therefore, we face a trade-off between strategic non-manipulability and one-side optimality beyond the rational preference domain.

Our study is closely related to recent works of the mechanism design under irrational preferences/choices. Bade (2008, 2016) considers the agents' irrational choice behavior in the housing problem by Shapley and Scarf (1974). de Clippel (2014) considers the general implementation problem and the housing problem as its sub-case in the complete information environments. He characterizes the implementability of the social choice function similar to Maskin (1999).[3] Except the housing market problem, none of these papers considers the two-sided matching problem. Our study is the first attempt to consider the irrational preferences in the two-sided matching problem.

In the decision theory literature, some authors consider a scenario that, even if an agent's behavior is inconsistent with the maximization based on the rational preference relation, he has a true rational preference and some restrictions, which induces the seemingly irrational behavior (Masatlioglu et al., 2012; Cherepanov et al., 2013). Other authors consider a scenario that an agent has multiple rational preference relations and uses them sequentially, which also induces the seemingly irrational behavior (Kalai et al., 2002; Manzini and Mariotti, 2007). Different from these scenarios, we do not assume that agents have the true rational preferences. As we mentioned above, agents can have incomplete and/or intransitive preferences as their own primitive. This view is consistent with that of Aumann (1962), Bewley (2002), and others.[4]

The rest of this chapter is organized as follows. In section 3.2, we offer the formal model and motivating examples. In section 3.3, we show our main results. Section 3.4 is the conclusion of this chapter. All omitted proofs are relegated to Appendix C.

[3]See also Korpela (2012).

[4]See, Ok (2002) and Nishimura and Ok (2016) for the more comprehensive discussion of this issue.

## 3.2 Preliminary

### 3.2.1 Model

Let $M$ and $W$ be the set of men and women, respectively. We assume that these sets are enumerated arbitrary. Throughout this chapter, for convenience, we only write the definition of each concept in terms of the man side if there is no confusion.

Each agent $i \in M \cup W$ has a binary relation $\succ_i$ on the other side. We assume that $\succ_i$ is antisymmetric (i.e., $x \succ_i y \to y \not\succ_i x$).[5] In other words, we assume that each agent has a strict preference. Let $\mathcal{P}_i$ be the set of all preferences of agent $i$ and $\mathcal{P} = \prod_{i \in M \cup W} \mathcal{P}_i$ be the set of all preference profiles.[6] We do not assume that each agent's preference relation $\succ_i$ is either complete nor transitive.[7] Let $\mathcal{P}^r \subset \mathcal{P}$ be the set of all complete and transitive preference profiles. Let us call $\mathcal{E} = (M, W, (\succ_i)_{i \in M \cup W})$ an economy.

A matching is a function $\phi : M \cup W \to M \cup W$ such that (1) $\phi(m) \in W \cup \{m\}$ for any $m \in M$, (2) $\phi(w) \in M \cup \{w\}$ for any $w \in W$ and (3) $\phi(\phi(i)) = i$ for any $i \in M \cup W$. Let $\mathcal{X}$ be the set of all matchings. A mechanism is a function $g : \mathcal{P} \to \mathcal{X}$. In particular, we denote $g^{DA}$ by the mechanism induced by the man proposing DA algorithm. Formally, the man proposing DA algorithm works as follows:

**Step** 0: put a preference profile $\succ \in \mathcal{P}^r$. All the following arguments are based on $\succ$.

**Step** 1: each man proposes to his most preferred woman. Then, each woman rejects unacceptable proposals and "keeps" her most preferred man from the proposals.

**Step** $k(\geq 2)$: each man rejected at step $k-1$ proposes to his most preferred woman who has not rejected him yet. Then, each woman rejects unacceptable proposals and "keeps" her most preferred man from the new proposals and man whom she has kept.

If there are no additional proposals, the algorithm stops. The final matching $g^{DA}(\succ)$ is determined by the pairs in the final step.

---

[5] We sometimes denote $X$ by a set of alternatives and typical elements are denoted by $x, y, z \in X$.

[6] For any $\succ \in \mathcal{P}$ and $i \in M \cup W$, we write $\succ_{-i}$ for the preference profile other than $i$.

[7] We say that a preference relation $\succ_i$ is *complete* if for each $x, y, x \succ_i y$ or $y \succ_i x$ holds. We also say that a preference relation $\succ_i$ is *transitive* if for each $x, y$ and $z$ such that $x \succ_i y$ and $y \succ_i z$, then $x \succ_i z$.

### 3.2.2 Stability

One of the desirable criteria for matchings is called stability. This concept requires that, under the matching, no agent is matched with someone who is not preferred to being single and no pair wants to deviate from their partners to be a new pair. Gale and Shapley (1962) show that there exists a stable matching in the domain $\mathcal{P}^r$. We say that a matching $\phi$ is *individually rational* (IR) if for any $i \in M \cup W$, (1) $i \not\succ_i \phi(i)$, or (2) $\phi(i) = i$. We should comment the difference from the usual definition. In our model, there exists a case that $m \not\succ_m w$ and $w \not\succ_m m$ for some $w$ because of the incompleteness of $\succ_m$. Therefore, if we write $\phi(i) \succ_i i$ in the first part as the usual way, such $w$ is not individually rational for $m$ even though $m$ does not consider being single is preferred to $w$. Our definition avoids such a situation. Hence, we require that under a individually rational matching, no one is matched with someone who is not preferred to being single, which is consistent with the usual definition. Indeed, note that this is equivalent to the standard (IR) condition if $\succ_m$ is complete. Similarly, we say that a pair $(m, w) \in M \times W$ *blocks* a matching $\phi$ if $w \succ_m \phi(m)$ and $m \succ_w \phi(w)$. We say that a matching $\phi$ is *stable* if it satisfies (IR) and there is no blocking pair. Let us denote $S(\mathcal{E})$ by the set of stable matchings in the economy $\mathcal{E}$. Again, note that this is equivalent to the standard stability condition if we consider the above stability concept in $\mathcal{P}^r$.

Bernheim and Rangel (2009) define the core where each agent has a choice function as primitive based on a generalized revealed preference relation, which is called an *unambiguous preference*. We say that a matching $\phi'$ *dominates* $\phi$ if there exists $U \subset M \cup W$ such that for any $i \in U$, (1) $\phi'(i) \succ_i \phi(i)$ and (2) $\phi'(i) \in U$. The core is the set of matchings which are not dominated by any matchings. Let us denote $C(\mathcal{E})$ by the *core* in the economy $\mathcal{E}$. Our definition of the core coincides with that of Bernheim and Rangel (2009) in this environment. We formally state it in Appendix C.2. The following proposition shows that an equivalence between the set of stable matchings and the core is satisfied in our environment as in $\mathcal{P}^r$.

**Proposition 3.1.** For each economy $\mathcal{E}$, $S(\mathcal{E})$=$C(\mathcal{E})$.

### 3.2.3 Motivating examples

In this subsection, we give some examples, in which agents are not rational and so the DA algorithm does not work.

**Example 3.1**. Let $M = \{m_1\}$ and $W = \{w_1, w_2\}$. Each agent's preference is defined as follows:

$m_1 : w_1 \succ_{m_1} w_2 \succ_{m_1} m_1$, but $m_1 \succ_{m_1} w_1$,

$w_1 : m_1 \succ_{w_1} w_1$,
$w_2 : m_1 \succ_{w_2} w_2$.

Note that $\succ_{m_1}$ is cyclic. Then, by this cyclic preference, there always exists a blocking pair whatever matchings we consider. Therefore, there is no stable matching in this example.

**Example 3.2**. Let $M = \{m_1, m_2, w_3\}$ and $W = \{w_1, w_2, w_3\}$. Then, each agent's preference is defined as follows:

$m_1 : w_2 \succ_{m_1} w_3, w_1 \succ_{m_1} w_3$, and $m_1 \succ_{m_1} w_1$,
$m_2 : w_1 \succ_{m_2} w_2$ and $w_1 \succ_{m_2} w_3$,
$m_3 :$ all agents are incomparable,

$w_1 :$ all agents are incomparable,
$w_2 : m_1 \succ_{w_2} m_2$ and $m_1 \succ_{w_2} m_3$,
$w_3 : m_1 \succ_{w_3} m_2$ and $m_1 \succ_{w_3} m_3$.

In this case, we cannot use the DA algorithm to obtain a stable matching because, for instance, $m_3$ cannot choose the most preferred agent even in the first step of the algorithm. So, we cannot know whether there is a stable matching or not in the economy. Nevertheless, a matching $\{(m_1, w_2), (m_2, w_3), (m_3, w_1)\}$ is stable.

Example 3.1 shows that, without the rationality assumption of agents' preferences, the existence result of a stable matching does not hold in some cases. In contrast, Example 3.2 shows that the DA algorithm does not work in some cases, so that we cannot verify whether a stable matching exists or not by using the DA algorithm even if a stable matching exists. From these observations, the following questions arise. When does a stable matching exist? How can we obtain it even if a stable matching exist? We answer these questions by proposing a new algorithm in the next section.

## 3.3 Main results

### 3.3.1 The DA algorithm with completion rules

As Example 3.1 suggests, the cyclical preferences for some agents may cause the non-existence of a stable matching. Thus, hereafter, we assume that each agent's preference is acyclic. Formally, $\succ_m$ is *acyclic* if for each $w^1, w^2, \cdots, w^k \in W \cup \{m\}$, $w^1 \succ_m w^2 \succ_m \cdots \succ_m w^k$, then $w_k \not\succ_m w_1$. Let $\mathcal{P}^{acyc} \subset \mathcal{P}$ be the set of acyclic preference profiles. We say that a preference relation $\succ_m^* \in \mathcal{P}_m$ is an order extension or a completion of $\succ_m$ if $\succ_m^*$ is (1) complete, transitive and (2) if $w \succ_m w'$, then $w \succ_m^* w'$ for all $w, w' \in W \cap \{m\}$. Suzumura (1976) shows that a binary relation can be extended to a complete and transitive relation if and only if it satisfies *Suzumura consistency*.[8] In our setting, the acyclicity is equivalent to the Suzumura consistency since we assume that each preference is strict. Hence, an order extension is possible.

We denote $f : \mathcal{P}^{acyc} \to \mathcal{P}^r$ by a completion rule. Let $\mathcal{F}$ be the set of all completion rules. Based on this idea, we can say the following result.

**Proposition 3.2.** Suppose that $\succ \in \mathcal{P}^{acyc}$. Then, for any completion rule $f \in \mathcal{F}$, a matching $g^{DA} \circ f(\succ)$ is stable in the economy $\mathcal{E}$.

In general, an explicit construction of a completion rule is not known. Therefore, Proposition 3.2 does not immediately tell us how to obtain a stable matching. However, since our domain is finite, we can construct a rule explicitly. The following procedure is an example of the construction.

For each $m \in M$, first, let us consider a directed graph on $W \cup \{m\}$ with respect to $\succ_m$. Let $V_m = W \cup \{m\}$ be the set of nodes. We define the edge as $w_1 \to w_2$ if $w_1 \succ_m w_2$. Let $E$ be the set of edges. Then, we say that there is a path from $w^1$ to $w^k$ if there is a sequence $w^1, w^2, \cdots, w^k \in W \cup \{m\}$ such that $w^1 \to w^2 \to \cdots \to w^k$. Let $P(V_m)$ be the set of paths on $W \cup \{m\}$. For each $p \in P(V_m)$ whose length is $k$, we define the number $\ell(p, w^l) = l$ for each $w^l (1 \leq l \leq k)$. That is, $\ell(p, w)$ is the number

[8]See Suzumura (1976) for the exact definition and the detailed discussion. Also, to find the discussion of the order extension, see Fishburn (1973).

of $w$ in the path $p \in P(V_m)$. Let $P(V_m, w) \subset P(V_m)$ be the set of paths including $w$. Since $\succ_m$ is acyclic, the graph $(V_m, E)$ is a directed graph with no cycle. Next, we define a number for each member of $V_m$ and, by using these numbers, we construct a *pseudo-ranking* with respect to $\succ_m$. Formally, for each $w \in W \cup \{m\}$, define

$$N_m(w) = \begin{cases} \max_{p \in P(V_m, w)} \ell(p, w) & \text{if } P(V_m, w) \neq \emptyset, \\ 1 & \text{otherwise.} \end{cases}$$

Since $\succ_m$ is acyclic, $N_m(w)$ is uniquely assigned to each $w \in W \cup \{m\}$. By construction, for each $w \in W \cup \{m\}$, if $w' \succ_m w$ then it must be the case that $N_m(w') < N_m(w)$. Finally, we construct a complete order by the following rule: $w \succ_m^* w'$ if and only if (1) $N_m(w) < N_m(w')$ or (2) $N_m(w) = N_m(w')$ and $w$'s index is smaller than that of $w'$. Let us denote $f^*$ by the completion rule obtained from this procedure. We can say the following result.

**Corollary 3.1.** Suppose that $\succ \in \mathcal{P}^{acyc}$. A matching $g^{DA} \circ f^*(\succ)$ is stable in the economy $\mathcal{E}$.

### 3.3.2 Strategic manipulation

We consider strategic manipulability of the matching mechanism. We say that a mechanism $g : \mathcal{P}^{acyc} \to \mathcal{X}$ is *strategy-proof* for $M$ if for any $m \in M$, $\succ_m, \succ'_m \in \mathcal{P}^{acyc}_m$ and $\succ_{-m} \in \mathcal{P}^{acyc}_{-m}$, we have $g(\succ'_m, \succ_{-m}) \not\succ_m g(\succ_m, \succ_{-m})$. Again, if $\succ_m$ is complete, this definition coincides with the standard definition of strategy-proofness. We can say the following positive result.

**Proposition 3.3.** Let $f \in \mathcal{F}$ be a completion rule. Then, a mechanism $g^{DA} \circ f$ : $\mathcal{P}^{acyc} \to \mathcal{X}$ is strategy-proof for $M$.

### 3.3.3 Optimality

In the domain $\mathcal{P}^r$, it is known that the DA algorithm induces a proposer-side optimal matching. We consider whether this property can be obtained by some mechanisms in the domain $\mathcal{P}^{acyc}$ or not. For each economy $\mathcal{E}$, a stable matching $\phi^*$ is $M$-optimal if for any stable matching $\phi \in S(\mathcal{E}) \backslash \{\phi^*\}$, (1) $\phi(m) \not\succ_m \phi^*(m)$ for all $m \in M$ and (2) $\phi^*(m) \succ_m \phi(m)$ for some $m \in M$. Note that a $M$-optimal matching is unique if it exists. In $\mathcal{P}^{acyc}$, it may happen that the $M$-optimal matching does not exist.

Therefore, we say that a mechanism $g : \mathcal{P}^{acyc} \to \mathcal{X}$ is $M$-optimal if it always induces a stable matching and, if it exists, induces a $M$-optimal matching. In contrast to the positive result in the domain $\mathcal{P}^r$, we obtain the following negative result.

**Proposition 3.4.** There is no mechanism $g : \mathcal{P}^r \to \mathcal{X}$ and completion rule $f \in \mathcal{F}$ such that the mechanism $g \circ f : \mathcal{P}^{acyc} \to \mathcal{X}$ is strategy-proof and $M$-optimal.

Proposition 3.4 shows a trade-off between strategy-proofness and one-side optimality when we use a mechanism with completion rules. Roughly speaking, a completion rule adds extra information to the reported preference profile even if the reported profile is the true preference profile. With arbitrary additional information, a usual mechanism $g : \mathcal{P}^r \to \mathcal{X}$ like $g^{DA}$ kills a desirable matching with respect to the true preferences. Hence, we have to consider more general mechanisms if we purse the possibility to obtain both strategy-proofness and one-side optimality.

We can observe a similar phenomenon when each agent's preference is a weak order, namely, we allow indifference to the preference. In such a case, one can obtain an example that there is no $M$-optimal stable matching.[9] In contrast, our problem is different. We show that we cannot obtain the $M$-optimal matching by using usual mechanisms with a completion rule even if it exits (see Appendix C.4). Hence, a problem which stems from each agent's irrational preference is different from a problem which stems from each agent's weak preference.[10]

## 3.4 Conclusion

This chapter considers the two-sided one-to-one matching problem where we do not assume either complete nor transitive preferences. We show that, in the domain $\mathcal{P}^{acyc}$, many positive results can hold like in $\mathcal{P}^r$. The new method used in this chapter is the order extension argument, which is similar flavor of a tie-breaking rule under weak preferences. Our result can be extended to the many-to-one matching problem if we consider responsive preferences for the many-side.

Although such an order extension argument is natural, we also show that one-side

[9] See, for example, Roth and Sotomayor (1992), p.34.

[10] Eliaz and Ok (2006) argue that indecisiveness (incompleteness) and indifference are different behavior. They give a choice theoretic foundation for incomplete preferences, which has different behavioral implication from the case of indifference.

optimality is incompatible with strategy-proofness when we use a usual stable matching mechanism like the DA algorithm with any completion rule. To know whether we can obtain a possibility result for strategy-proofness and one-side optimality or not, we have to consider a different idea of the technique used in this chapter. This is left for the future research.

# Chapter 4 The Weighted Egalitarian Shapley Values

This chapter is based on the same title of the joint work with Takaaki Abe.

## 4.1 Introduction

The most eminent allocation rule for cooperative games with transferable utility (TU-games) may be the Shapley value introduced by Shapley (1953b). After the celebrated study of Shapley (1953b), many other axiomatic foundations for the rule were intensively studied.[1] In particular, Young (1985) shows that the Shapley value is the unique efficient rule that satisfies *strong monotonicity* and *symmetry*. These two properties focus on each player's contributions in a game to determine their rewards. More precisely, (1) strong monotonicity states that each player receives more as his/her contributions increase and (2) symmetry requires that any two players receive the same amount if their contributions are equal. Therefore, the Shapley value can be thought of as an allocation rule completely based on each player's performance.

In actual situations, however, we often use an allocation rule which can assign positive payoff to each player even if he/she cannot contribute for some reason. For example, in the case of wage assignment in a firm, each worker may receive a basic salary in addition to a reward for her contribution. This system may be more secure than a system without a basic salary given the possibility that employees cannot contribute because of, for instance, raising children or hospitalization. Constructing an allocation rule which integrates this kind of social equity, which is often referred to as a solidarity principle, with contribution based rule, is one of the main concerns in recent literature of cooperative game theory (Nowak and Radzik, 1994; Joosten, 1996; Casajus and Huettner, 2013, 2014; van den Brink et al., 2013; Joosten, 2016). Also, in the same example, the wage may be affected by some index independent of one's contributions, such as seniority, educational background, and entitlements. Moreover,

[1] For recent studies, see Casajus (2011, 2014) and Casajus and Yokote (2017a).

in the case of redistribution of income in a society, each household has a heterogeneous background, such as the number of children or handicaps (Casajus 2015, 2016; Casajus and Yokote 2017b; Abe and Nakada, 2017).[2] Modern taxation systems consider such heterogeneities. This observation raises the following question: what allocation rule reconciles performance-based evaluation with a solidarity principle and takes players' heterogeneity into consideration?

To answer this question, this chapter considers rules satisfying weaker monotonicity and symmetry. We show that these axioms (elaborated later) characterize the new class of rules, *weighted egalitarian Shapley values*, where each rule in this class is given as a convex combination of the Shapley value and the weighted division (Béal et al., 2016). Each rule in our class can be interpreted as a redistribution rule, via which a player keeps a part of his income and offers the other part as a tax. After every player's tax is collected, the total tax revenue is redistributed. As the name of our class of rules indicates, weights are featured in our class. To formulate weights for players, we employ the following two approaches.

First, we characterize the weighted egalitarian Shapley values where a weight profile is endogenously determined. For monotonicity, we employ weak monotonicity, i.e., each player receives more if his contributions and the worth of the grand coalition increase.[3] This property does not require each player's evaluation to depend only on his contribution but rather allows it to depend also on the worth of the grand coalition to reflect a solidarity principle. For the symmetry property, we consider the following two new axioms: *weak differential marginality for symmetric players* (WDMSP) and *ratio invariance for null players* (RIN). The former axiom (WDMSP) requires the following property. Suppose that there are two null players, i.e., players whose marginal contributions are zero. If the worth of the grand coalition keeps unchanged but these two players make the same additional contributions, then the two players should receive the same additional payoff. The latter axiom (RIN) is a minimal fairness requirement for null players, which requires that the payoff ratio between two null players does not vary as long as they are null players. We show

[2]See Roth (1979), Kalai and Samet (1987), Chun (1988, 1991), and Nowak and Radzik (1995) for other examples.

[3]Weak monotonicity without weights was introduced by van den Brink et al. (2013) to characterize egalitarian Shapley values. We discuss this topic later.

that a rule satisfies these axioms, efficiency and nullity[4] if and only if it is a weighted egalitarian Shapley value.

Second, we suppose that each player's heterogeneity is parametrized by an exogenous weight profile. In this case, how we should integrate the exogenous weights into the two axioms, monotonicity and symmetry, is a problem. The three axioms efficiency* (E*), weak monotonicity* ($M^{-*}$), and ratio invariance for null players* (RIN*) correspond to the axioms in the first case. For the symmetry property, we consider the following two new axioms: *symmetry** (SYM*) and *fair evaluation for contribution* (FEC*). The former axiom (SYM*) states that any two players receive the same payoff if their contributions and weights are equal. In other words, even though their contributions are equal, we admit that the two players receive different payoffs if their weights are different. The latter axiom (FEC*) states that if a player additionally contributes, then a reward for his additional contributions should be evaluated impartially, regardless of his weight. That is, this axiom requires that we should take each player's contributions and weight into consideration separately. We show that a rule satisfies these axioms if and only if it is a weighted egalitarian Shapley value with an exogenous weight profile.

Our results contribute to the literature regarding axiomatization of variants of the Shapley value to accommodate a solidarity principle and heterogeneity, in particular, the egalitarian Shapley values (Joosten, 1996) and the weighted Shapley values (Shapley, 1953a). We analyze the axiomatic differences among our rules, the egalitarian Shapley values and the weighted Shapley values.

The egalitarian Shapley values are a class of rules that are convex combinations of the Shapley value and the equal division. That is, each player can obtain some amount of payoff equally and extra amount depending on his/her own contributions. In this sense this rule is considered as a compromise between marginalism and egalitarianism. Since the equal division is a special case of the weighted divisions, our rules subsume this class of rules. We show that the difference between this class and ours stems from different symmetry properties by comparing our result with that of Casajus and Huettner (2014); they characterize their rules by efficiency, weak monotonicity, and symmetry.[5]

---

[4]Nullity is a very weak feasibility condition, which requires that every player receives nothing if no coalition induces any value. Nullity is called *triviality* in Chun (1989).

[5]For other characterization of the egalitarian Shapley values, see Casajus and Huettner (2013)

The weighted Shapley value is an allocation rule based on weighted contributions. Although this rule and ours only share efficiency, the difference is understood as a consequence of the requirement of monotonicity and symmetry by comparing our result with that of Nowak and Radzik (1995); they characterize the class of rules by efficiency, strong monotonicity, mutual dependence and positivity, where mutual dependence implies (RIN) but (WDMSP) and positivity implies nullity under efficiency and ($M^-$).[6]

Joosten (2016) introduces the egalitarian weighted Shapley values, that is, a class of rules that are convex combinations of the weighted Shapley value and the equal division. The difference between his class and ours lies in how to address a weight profile of players, that is, heterogeneity. In our rules, each player's contributions are evaluated without weights, while the weights determine players' "basic payoffs," namely, the weighted division. In contrast, the egalitarian weighted Shapley values take into account the weights to evaluate each player's contributions, while the "basic payoffs" are given as the equal division. In this sense, Joosten (2016)'s class and ours can be thought of as two different generalizations of the egalitarian Shapley values: the egalitarian Shapley value takes a middle ground between the weighted egalitarian Shapley value and the egalitarian weighted Shapley values.[7]

The remainder of this chapter is organized as follows. In Section 4.2, we provide basic definitions and notation. In Section 4.3, we offer the main characterization of the weighted egalitarian Shapley values. In Section 4.4, we offer the characterization of the weighted egalitarian Shapley values in the case of exogenous weight profiles. Section 4.5 is the conclusion of this chapter. All proofs are relegated to Appendix D.

## 4.2 Preliminaries

Let $N = \{1, \cdots, n\}$ be the set of players and the function $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$ denote a characteristic function. A *coalition* of players is defined as a subset of the player set, $S \subseteq N$. Let $|S|$ denote the cardinality of coalition $S$. We sometimes use $n$ to denote $|N|$. A cooperative TU game is $(N, v)$. Fixing the player set $N$, we denote by $\mathcal{G}$ the set of all TU games with the player set $N$. An allocation rule is denoted by

and van den Brink et al. (2013).

[6] For other characterizations of the weighted Shapley value, see Kalai and Samet (1987), Chun (1991), Hart and Mas-colell (1989), and Yokote (2014).

[7] In Section 4.5, we discuss the future direction to unify these two classes.

$f : \mathcal{G} \to \mathbb{R}^N$. Player $i$'s marginal contribution to a coalition $S \subseteq N \setminus \{i\}$ is defined as $v(S \cup \{i\}) - v(S) = 0$. For each $v \in \mathcal{G}$, we say that player $i \in N$ is a *null player* in $v$ if $v(S \cup \{i\}) - v(S) = 0$ for all $S \subseteq N \setminus \{i\}$. We also say that two players $i, j \in N$ are symmetric in $v$ if $v(S \cup \{i\}) - v(S) = v(S \cup \{j\}) - v(S)$ for all $S \subseteq N \setminus \{i, j\}$. For any nonempty coalition $T \subseteq N$, the unanimity game $u_T \in \mathcal{G}$ is defined as follows: for any $S \in 2^N$,

$$u_T(S) = \begin{cases} 1 & \text{if } \ T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

The Shapley value, $Sh(v)$, is given as follows: for any player $i \in N$,

$$Sh_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)).$$

The Shapley value assigns an average of marginal contributions to each player. Young (1986) shows that the Shapley value satisfies the following properties:

**Efficiency, E**. For any $v \in \mathcal{G}$, $\sum_{i \in N} f_i(v) = v(N)$.

**Strong Monotonicity, M**. For any $v, v' \in \mathcal{G}$ and $i \in N$, if $v(S \cup \{i\}) - v(S) \geq v'(S \cup \{i\}) - v'(S)$ for all $S \subseteq N \setminus \{i\}$, then $f_i(v) \geq f_i(v')$.

**Symmetry, SYM**. For any $v \in \mathcal{G}$ and $i, j \in N$, if $i, j$ are symmetric in $v$, then we have $f_i(v) = f_j(v)$.[8]

The following theorem shows that the Shapley value is the unique solution satisfying these three properties.

**Theorem 4.1** (Young, 1986)**.** An allocation rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E), (M) and (SYM) if and only if $f_i(v) = Sh_i(v)$.

Since the Shapley value determines each player's payoff only depending on his/her contributions, it ignores both equity/solidarity and heterogeneity among players, which do not depend on contributions. In the following section, we introduce our class of rules that exhibit these features.

[8] This is also known as the *equal treatment property*.

## 4.3 Axiomatization of the weighted egalitarian Shapley values

We define $w = (w_i)_{i \in N} \in \mathbb{R}^N_+$ with $\sum_{i \in N} w_i = 1$ as a weight profile and $\mathcal{W}$ as the set of all possible weight profiles.

First, we consider the following weaker version of monotonicity which was introduced by van den Brink et al. (2013).

**Weak Monotonicity, M$^-$.** For each $v, v' \in \mathcal{G}$ with $v(N) \geq v'(N)$, if $v(S) - v(S \backslash \{i\}) \geq v'(S) - v'(S \backslash \{i\})$ for all $S \subseteq N$ with $i \in S$, then $f_i(v) \geq f_i(v')$.

This property states that a player's payoff weakly increases as his marginal contributions and the total value weakly increase. In contrast with (M), this property does not insist that each player's evaluation totally depend on his contributions but rather allows that it can depend on the total value, which is a flavor of the solidarity principle.

The next axiom is a requirement for the treatment of null players.

**Ratio Invariance for Null Players, RIN.** For any $v, v' \in \mathcal{G}$ and $i, j \in N$ such that $i, j$ are null players in $v, v'$, we have $f_i(v) \cdot f_j(v') = f_i(v') \cdot f_j(v)$.

Ratio invariance for null players requires that as long as some players, say $i, j$, contribute zero in both games $v$ and $v'$, the ratio of their payoffs, $f_i(v)/f_j(v)$, does not vary.

The following axiom is a requirement for null players about additional contribution.

**Weak Differential Marginality for Symmetric Players, WDMSP.** For any $i, j \in N$ and $v, v' \in \mathcal{G}$ such that $i, j$ are null players in $v$, if $i, j$ are symmetric in $v'$ and $v(N) = v'(N)$, then $f_i(v) - f_i(v') = f_j(v) - f_j(v')$.

Note that players $i, j$ are also symmetric in $v$. Suppose that there are two null players in $v$. If the the worth of the grand coalition keeps unchanged but these two players make the same additional contributions (from $v$ to $v'$), then this axiom requires that the two players should receive the same additional payoff.[9] That is, the contribution itself is fairly evaluated as the same as under the original symmetry axiom, but the axiom (WDMSP) does not exclude the possibility that the payoffs of two players can differ because of their heterogeneity, that is, $f_i(v) \neq f_j(v)$ and $f_i(v') \neq f_j(v')$ can be allowed.

The following axiom is a harmless feasibility requirement.

**Nullity, NY**.[10] Let $\mathbf{0}$ be the null game. For any $i \in N$, $f_i(\mathbf{0}) = 0$.

Now, we introduce the following new class of allocation rules, which we call *weighted egalitarian Shapley values*:

$$f_i(v) = \delta \cdot Sh_i(v) + (1 - \delta) \cdot w_i v(N) \text{ where } \delta \in [0, 1] \text{ and } w \in \mathcal{W}.$$

Note that the allocation rule is specified by two parameters $\delta \in [0, 1]$ and $w \in \mathcal{W}$. For $\delta = 1$, the allocation rule coincides with the Shapley value and distributes the surplus $v(N)$ based only on the players' contributions. For $\delta = 0$, our rule coincides with the weighted devision (Béal et al., 2016). As Joosten (1996) describes, $\delta$ is interpreted as the degree of solidarity among players. It is clear that rules in this class satisfy all of the axioms. Now, we are ready to offer our main axiomatization result as follows.

**Theorem 4.2.** Suppose that $n \neq 2$. An allocation rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E), ($M^-$), (RIN), (WDMSP), and (NY) if and only if there exists a $\delta \in [0, 1]$ and a weight profile $w \in \mathcal{W}$ such that $f_i(v) = \delta \cdot Sh_i(v) + (1 - \delta) \cdot w_i v(N)$.

*Proof.* See Appendix D.1. $\square$

For the independence of the axioms, we provide examples in Appendix D.2. Note also that, when $n = 2$, there is an allocation rule which satisfies all the axioms, but it is not included in our class. In this sense, uniqueness of the class of rules does not hold when $n = 2$. We provide a counterexample in Appendix D.3.

[9] Note that (WDMSP) is weaker than (SYM). It is also a weaker version of *differential marginality* defined by Casajus (2010, 2011). Formally, differential marginality is defined as follows: for any $i, j \in N$ and $v, v' \in \mathcal{G}$, if $v(S \cup \{i\}) - v(S \cup \{j\}) = v'(S \cup \{i\}) - v'(S \cup \{j\})$ for all $S \subseteq N \setminus \{i, j\}$, then $f_i(v) - f_j(v) = f_i(v') - f_j(v')$.

[10] Note that nullity is named *triviality* in Chun (1989).

## 4.4 Axiomatization of the weighted egalitarian Shapley values with an exogenous weight

In this section, we assume that a profile $w = (w_i)_{i \in N} \in \mathcal{W}$ is exogenously given. That is, we consider an allocation rule in the extended domain $f : \mathcal{G} \times \mathcal{W} \to \mathbb{R}^N$.

We first introduce the analogs of the axioms in the previous section.

**Efficiency$^*$, E$^*$.** For any $v \in \mathcal{G}$ and $w \in \mathcal{W}$, $\sum_{i \in N} f_i(v, w) = v(N)$.

**Weak Monotonicity$^*$, M$^{-*}$.** For any $v, v' \in \mathcal{G}$, $w \in \mathcal{W}$ and $i \in N$, if $v(N) \geq v'(N)$ and $v(S \cup \{i\}) - v(S) \geq v'(S \cup \{i\}) - v'(S)$ for all $S \subseteq N \setminus \{i\}$, then $f_i(v, w) \geq f_i(v', w)$.

**Ratio Invariance for Null players$^*$, RIN$^*$.** For any $v \in \mathcal{G}$, $w \in \mathcal{W}$ and any null players $i, j \in N$, we have $w_i \cdot f_j(v, w) = w_j \cdot f_i(v, w)$.

Next, we introduce two weaker notions of symmetry.

**Symmetry$^*$, SYM$^*$.** For any $v \in \mathcal{G}$, $w \in \mathcal{W}$ and $i, j \in N$, if $i, j$ are symmetric in $v$ and $w_i = w_j$, then we have $f_i(v, w) = f_j(v, w)$.

Symmetry$^*$ requires that two players whose contributions and weights are the same should receive the same amount. If we fix $w$ to the equal weight $(\frac{1}{n}, ..., \frac{1}{n})$, this is equivalent to usual symmetry. Note that we allow different payoffs for two players, even if their contributions are the same, as long as their weights are different.

In addition to the axioms above, we impose the following requirement to take each player's contributions and weights into consideration separately, which supports our motivation to consider heterogeneity.

**Fair Evaluation for Contribution, FEC$^*$.** For any $v \in \mathcal{G}$, $w, w' \in \mathcal{W}$ and $i \in N$,

$$f_i(v, w) - f_i(v(N)u_{N \setminus \{i\}}, w) = f_i(v, w') - f_i(v(N)u_{N \setminus \{i\}}, w').$$

This axiom describes how the payoff of a player changes due to a shift between the two weight profiles. The axiom requires that the payoff difference because of this shift should be the same as long as productivity of the other players keeps unchanged and this player $i$ is a null player. The following result shows that the rule satisfies these five axioms if and only if it is the weighted egalitarian Shapley values with an exogenously determined weight profile $w \in \mathcal{W}$.

**Theorem 4.3.** Suppose $n \neq 2$. An allocation rule $f : \mathcal{G} \times \mathcal{W} \to \mathbb{R}^N$ satisfies (E*), (M$^{-*}$), (RIN*), (SYM*), and (FEC*) if and only if there exists a $\delta \in [0,1]$ such that $f_i(v,w) = \delta \cdot Sh_i(v) + (1-\delta) \cdot w_i v(N)$.

*Proof.* See Appendix D.4. □

For the independence of the axioms, we provide examples in Appendix D.5.

## 4.5 Conclusion

In this chapter, we propose and axiomatically characterize a new class of allocation rules called weighted egalitarian Shapley values. This allocation rule integrates equity and heterogeneity with the Shapley value.

As briefly argued in Section 4.1, monotonicity and symmetry distinguish our class of rules from the other rules, such as the egalitarian Shapley values and the weighted Shapley values. Below, we elaborate on the differences among these classes of rules by comparing the axioms.

Joosten (1996) introduces the class of the egalitarian Shapley values, which are convex combinations of the Shapley value and the equal division:

$$f_i(v) = \delta \cdot Sh_i(v) + (1-\delta) \cdot \frac{v(N)}{n} \text{ where } \delta \in [0,1].$$

Note that this class of rules is a subset of weighted egalitarian Shapley values (i.e., $w = (\frac{1}{n}, ..., \frac{1}{n})$). To clarify the difference between our allocation rules and the egalitarian Shapley values, we consider the characterization of Casajus and Huettner (2014).

**Theorem 4.4** (Casajus and Huettner, 2014)**.** Suppose $n \neq 2$. An allocation rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E), (M$^-$), (SYM) if and only if there exists a $\delta \in [0,1]$ such that $f_i(v) = \delta \cdot Sh_i(v) + (1-\delta) \cdot \frac{v(N)}{n}$.

By comparing Theorem 4.2 with Theorem 4.4, the difference between these two rules is observed to be the requirement of symmetry. As Table 4.1 shows, the weighted egalitarian Shapley values no longer satisfy symmetry, while all egalitarian Shapley values obey symmetry.[11] This difference should be ascribed to the weight which each weighted egalitarian Shapley value contains. To see what properties the weight makes a solution obey/violate, we briefly introduce another variation of the Shapley value, the weighted Shapley value.

Shapley (1953a) introduces the weighted Shapley value $Sh_i^w(v)$, which is a unique linear solution such that for each unanimity game $u_T$, there is a weight $w \in \mathbb{R}^N_{++}$ with $\sum_{j \in N} w_j = 1$ such that

$$Sh_i^w(u_T) = \begin{cases} \frac{w_i}{\sum_{j \in T} w_j} & \text{if} \quad i \in T, \\ 0 & \text{otherwise.} \end{cases}$$

Similar to the Shapley value, the weighted Shapley value satisfies (M). Therefore, this rule can be considered as a performance-based rule and does not satisfy equity. However, the weighted Shapley values allow us to allocate players' surplus based not only on their contributions but also on their weights, that is, heterogeneity is taken into account.

Nowak and Radzik (1995) consider the following weak symmetry and regularity properties.

**Mutual Dependence, MD**. For any two players $i, j \in N$ and $v, v' \in \mathcal{G}$, if $i, j$ are symmetric in $v$ and $v'$, then $f_i(v)f_j(v') = f_i(v')f_j(v)$.

**Strict Positivity, SP**. For any monotonic $v \in \mathcal{G}$ such that there are no null players, we have $f_i(v) > 0$ for all $i \in N$.[12]

They show that these properties, together with (E) and (M), characterize the weighted Shapley values.

**Theorem 4.5** (Nowak and Radzik, 1995)**.** An allocation rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E), (M), (MD) and (SP) if and only if there exists a weight $w \in \mathbb{R}^N_{++}$ with $\sum_{j \in N} w_j = 1$ such that $f_i(v) = Sh_i^w$.

[11] Abbreviation "w-egSh" means the weighted egalitarian Shapley values, "egSh" means the egalitarian Shapley values, and "w-Sh" is the weighted Shapley values discussed later.

[12] A game $v$ is monotonic if $v(T) \geq v(S)$ for any $S \subseteq T$.

Note that (MD) implies (RIN) but not (WDMSP). Moreover, under (E) and ($M^-$), (SP) implies (NY).[13] Therefore, considering Theorem 4.2 and Theorem 4.5, monotonicity and symmetry are the differences between these two rules.

| | E | M | $M^-$ | SYM | WDMSP | RIN | MD | NY | SP |
|---|---|---|---|---|---|---|---|---|---|
| w-egSh | + | | + | | + | + | | + | |
| egSh | + | | + | + | | | | | |
| w-Sh | + | + | | | | | + | | + |

Table 4.1: Axioms and Rules

Given that our rule has an axiomatic foundation similar to that of the weighted Shapley values, we may weaken the properties and obtain another class of meaningful rules. For example, there exist $\delta \in [0,1]$, $w \in \mathcal{W}$ and $z \in \mathbb{R}^N_{++}$ with $\sum_{j \in N} z_j = 1$ such that

$$f_i(v) = \delta \cdot Sh^z_i(v) + (1-\delta) \cdot w_i v(N).$$

When $w = (1/n, \cdots, 1/n)$, as briefly mentioned in Section 4.1, the rules are called the egalitarian weighted Shapley values introduced by Joosten (2016), which is another generalization of the egalitarian Shapley values. Hence, the class above contains both the weighted egalitarian Shapley values and the egalitarian weighted Shapley values, whose intersection is the egalitarian Shapley values. We conjecture that this general class can be characterized by weak monotonicity and a variant of symmetry.

[13] If $f_i(\mathbf{0}) = \delta < 0$ for all $i \in N$, it contradicts to (E). Hence, supposing that $f_i(\mathbf{0}) = \delta > 0$ for some $i$, we consider $v \in \mathcal{G}_N$ such that $v(N) = \varepsilon \in (0,\delta)$ and $v(S) = 0$ for $S \neq N$. By ($M^-$), $f_i(v) \geq f_i(\mathbf{0}) = \delta$. Also, by (SP), $f_i(v) > 0$ for all $i \in N$. This contradicts (E) because $v(N) = \varepsilon \in (0,\delta)$.

# Chapter 5 Robust Voting under Uncertainty

This chapter is based on the same title of the joint work with Shmuel Nitzan and Takashi Ui.

## 5.1 Introduction

Consider a choice of a voting rule on a succession of two alternatives (such as "yes" or "no") by a group of individuals who are uncertain about their future preferences. Each individual presumes that the gain from the passage of a favorable issue equals the loss from the passage of an unfavorable issue. Imagine that someone proposes a voting rule such that the expected loss of every individual is greater than the expected gain. Then, the group will not agree to adopt it. In fact, such a voting rule is problematic because the probability that the outcome agrees with an individual's preference, i.e., responsiveness, is less than one-half for all individuals. This means that a group decision reflects minority preferences on average and that the decision can be eventually not only unfair ex post facto but also more likely incorrect. Moreover, by adopting another voting rule whose collective decision always disagrees with that of the original rule, the group can make responsiveness of every individual greater than one-half; that is, this rule is better than the original rule for all individuals in terms of responsiveness.

To evaluate the expected net gain, individuals must know the true probability distribution of their preferences. However, in reality, they face Knightian uncertainty and have little confidence regarding the true probabilities.[1] This makes it difficult for them to figure out whether the expected net gain is positive or negative, which raises the following questions. Does there exist a voting rule such that the expected net gain of every individual is never negative whatever the underlying probability distribution is? If the answer is yes, what is it?

[1] Knight (1921) distinguishes risky situations, where a decision maker knows the probabilities of all events, and uncertain situations, where a decision maker does not know them.

This chapter proposes two normative criteria for voting rules under Knightian uncertainty and provides answers to the above questions. Our criteria require that a voting rule should avoid the following worst-case scenarios. The first worst-case scenario is that the true responsiveness of every individual is less than or equal to one-half, or equivalently, the expected net gain of every individual is nonpositive. By replacing "less than or equal to" with "strictly less than" in this scenario, we obtain a slightly more severe scenario. The second worst-case scenario is that the true responsiveness of every individual is strictly less than one-half, or equivalently, the expected net gain of every individual is strictly negative. A voting rule is said to be *robust*[2] if, for any probability distribution of preferences, it avoids the first worst-case scenario. A voting rule is said to be *weakly robust* if, for any probability distribution of preferences, it avoids the second worst-case scenario. Because the first scenario is less severe than the second one, robustness is a stronger requirement than weak robustness: under robust rules, responsiveness of at least one individual must be strictly greater than one-half, whereas, under weakly robust rules, responsiveness of every individual can be less than or equal to one-half, as long as responsiveness of at least one individual is equal to one-half, in which case a collective decision is at best neutral to each individual's choice on average.

In the main result, we show that a voting rule is robust if and only if it is a weighted majority rule (WMR) without ties. We also show that a voting rule is weakly robust if and only if it is a WMR allowing ties with an arbitrary tie-breaking rule. The proofs of both results are based upon the theorem of alternatives due to von Neumann and Morgenstern (1944), which is also known as a corollary of Farkas' lemma. Because Farkas' lemma is mathematically equivalent to the fundamental theorem of asset pricing (cf. Dybvig and Ross 2003, 2008), the proofs can be understood in terms of the following imaginary asset for each individual $i$: one unit of asset $i$ yields $+1$ if individual $i$'s preference agrees with the collective decision and $-1$ otherwise. Using the fundamental theorem of asset pricing, we can show that there exists a portfolio with nonnegative weights in all assets (i.e. no short selling) yielding a strictly positive payoff in each state if and only if, for any arbitrage-free price vector, the price of at least one asset is strictly positive. The former condition is true if and only if the

[2] We borrow the term "robustness" from robust statistics, statistics with good performance for data drawn from a wide range of probability distributions Huber (1981).

voting rule is a WMR and the latter condition is true if and only if responsiveness of at least one individual is greater than one-half, thus implying the equivalence of a robust rule and a WMR.

We apply the above result to anonymous rules, which are considered to be fair, and obtain the following characterization of robust anonymous rules. A simple majority rule (SMR) is a unique robust anonymous rule when the number of individuals is odd, whereas no anonymous rule is robust when the number of individuals is even. In the latter case, however, a SMR with an anonymous tie-breaking rule is a weakly robust anonymous rule. This implies that we face a trade-off between robustness and anonymity when the number of individuals is even: we must be content with a nonanonymous rule if we require robustness and we must be content with a weakly robust rule if we require anonymity.

To illustrate the difference between robustness and weak robustness as well as the trade-off between robustness and anonymity, assume that the number of individuals is even and consider SMRs. A SMR with some tie-breaking rule is robust if and only if it is represented as a WMR allowing no ties. For example, a SMR with a casting (tie-breaking) vote is a robust nonanonymous rule. On the other hand, a SMR with any tie-breaking rule is weakly robust. In particular, a SMR with the status quo tie-breaking rule (i.e., the status quo is followed whenever there is a tie) is a weakly robust anonymous rule. Because no anonymous rule is robust when the number of individuals is even, the choice between the two SMRs depends upon which criterion to prioritize, and notice that, in the real world, both rules are widely used. A SMR with a casting vote is used by legislatures such as the United States Senate, the Australian House of Representatives, and the National Diet of Japan. A SMR with the status quo tie-breaking rule is used by legislatures such as the New Zealand House of Representatives, the British House of Commons, and the Australian Senate.

Our characterization of a WMR complements the well-known characterization of a WMR due to Rae (1969), Taylor (1969), and Fleurbaey (2008). Their result, which we call the Rae-Taylor-Fleurbaey (RTF) theorem, states that a voting rule is a WMR if and only if it maximizes the corresponding weighted sum of responsiveness over all individuals, where a probability distribution of preferences is assumed to be known. The normative implication of the RTF theorem is efficiency of WMRs;[3] that is, a

[3]See also Schmitz and Tröger (2012) and Azrieli and Kim (2014).

voting rule is efficient in terms of responsiveness if and only if it is a WMR. Note that an efficient rule achieves the optimal outcomes, whereas a robust rule avoids the worst outcomes. Thus, robustness together with efficiency gives a dual characterization of a WMR in terms of responsiveness.

We emphasize that efficiency and robustness are distinct sufficient conditions for WMRs. A voting rule is efficient if, under a fixed probability distribution of preferences, any random voting rule is never better than this rule for all individuals. In contrast, a voting rule is robust if, under any probability distribution of preferences, the voting rule whose collective decision always disagrees with this rule is never better than this rule for all individuals. Thus, to check efficiency, we need to consider all random voting rules, i.e., all probability distributions of voting rules, ensuring that the optimal outcome is achieved. On the other hand, to check robustness, we need to consider all probability distributions of preferences, ensuring that the worst outcome is avoided. Because robustness and efficiency are different requirements, the sufficiency of robustness for WMRs is not implied by the RTF theorem

This chapter not only contributes to the literature on the axiomatic foundations of a SMR or a WMR (May, 1952; Fishburn, 1973), but also joins a recently growing literature on economic design with worst-case objectives. Most studies in the latter literature, however, have focused on mechanism design. For example, Chung and Ely (2007) consider a revenue maximization problem in a private value auction where the auctioneer does not know agents' belief structures exactly[4] and show that the optimal auction rule is a dominant-strategy mechanism when the auctioneer evaluates rules by their worst-case performance. On the other hand, Carroll (2015) considers a moral hazard problem where the principal does not know the agent's set of possible actions exactly and shows that the optimal contract is linear when the principal evaluates contracts by their worst-case performance.[5] In contrast to these papers, we consider a choice of voting rules with the worst-case objective to characterize WMRs, where the constitution-maker does not know the probability distribution of preferences, thus demonstrating that this approach is also useful in the study of voting and social choice.

[4]This is a standard assumption in robust mechanism design (Bergemann and Morris, 2005).

[5]Other recent examples of economic design with worst-case objectives include Bergemann and Schlag (2011), Yamashita (2015), Bergemann et al. (2016), and Carroll (2017) among others. In decision theory, Gilboa and Schmeidler (1989) is a seminal paper.

The rest of this chapter is organized as follows. Section 5.2 summarizes properties of WMRs which will be used in the subsequent sections. Section 5.3 introduces the concepts of robustness and weak robustness and Section 5.4 characterizes robust rules and weakly robust rules. Section 5.5 compares our result and the RTF theorem. We conclude the chapter in Section 5.6.

## 5.2 Weighted majority rules

Consider a group of individuals $N = \{1, \cdots, n\}$ that faces a choice between two alternatives (such as "yes" or "no"). The choice of individual $i \in N$ is represented by a decision variable $x_i \in \{-1, 1\}$. The choices of the group members are summarized by a decision profile $x = (x_i)_{i \in N}$. Let $\mathcal{X} = \{-1, 1\}^N$ denote the set of all possible profiles.

A voting rule is a mapping $\phi : \mathcal{X} \to \{-1, 1\}$. Let $\Phi$ denote the set of all voting rules. A voting rule $\phi \in \Phi$ is a weighted majority rule (WMR)[6] if there exists a nonzero weight vector[7] $w = (w_i)_{i \in N} \in \mathbb{R}^N$ satisfying

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{i \in N} w_i x_i > 0, \\ -1 & \text{if } \sum_{i \in N} w_i x_i < 0. \end{cases}$$

A simple majority rule (SMR) is a special case with positive equal weights, i.e., $w_i = w_j > 0$ for all $i, j \in N$. When there is a tie, i.e. $\sum_{i \in N} w_i x_i = 0$, a tie-breaking rule is used to determine a voting rule. For example, a SMR requires a tie-breaking rule if $n$ is even.

A voting rule $\phi$ is anonymous if it is symmetric in its $n$ variables; that is, $\phi(x) = \phi(x^\pi)$ for each $x \in \mathcal{X}$ and each permutation $\pi : N \to N$, where $x^\pi = (x_{\pi(i)})_{i \in N}$. A SMR is anonymous if $n$ is odd or if $n$ is even and its tie-breaking rule is anonymous, i.e., symmetric in its $n$ variables. A WMR with nonnegative weights is anonymous if and only if it is an anonymous SMR.

The following characterization of WMRs, which is immediate from the definition, plays an important role in the subsequent analysis.

[6] For an axiomatic foundation of weighted majority rules, see Fishburn (1973), Fishburn and Gehrlein (1977), Nitzan and Paroush (1981), Eeiny and Lehrer (1989), and Taylor and Zwicker (1992) among others.

[7] We allow negative weights, which appears in Proposition 5.3 and Appendix E.1.

**Lemma 5.1.** A voting rule $\phi \in \Phi$ is a WMR with a weight vector $w \in \mathbb{R}^N$ if and only if

$$\phi(x) \sum_{i \in N} w_i x_i \geq 0 \text{ for all } x \in \mathcal{X}.$$

A voting rule $\phi \in \Phi$ is a WMR with a weight vector $w \in \mathbb{R}^N$ allowing no ties if and only if

$$\phi(x) \sum_{i \in N} w_i x_i > 0 \text{ for all } x \in \mathcal{X}.$$

This lemma states that $\phi$ is a WMR allowing no ties (allowing ties) if and only if the corresponding weighted sum of $\phi(x)x_i$ over $i \in N$ is positive (nonnegative) for all $x \in \mathcal{X}$ because the left-hand side of the inequality is $\sum_{i \in N} w_i(\phi(x)x_i)$. Note that $\phi(x)x_i$ equals $+1$ if $i$'s choice agrees with the collective decision and $-1$ otherwise. Thus, we can regard $\phi(x)x_i$ as individual $i$'s payoff by assuming that the gain from the passage of a favorable issue and the loss from the passage of an unfavorable issue are equal and normalized to one.

Note that a weight vector $w$ representing a WMR is a solution to a system of linear inequalities in Lemma 5.1. This observation leads us to the next lemma, which shows that the set of WMRs with nonnegative weights (i.e., $w_i \geq 0$ for all $i \in N$) coincides with that with positive weights (i.e., $w_i > 0$ for all $i \in N$) if there are no ties.

**Lemma 5.2.** A WMR with nonnegative weights allowing no ties is represented as a WMR with positive weights allowing no ties.

*Proof.* Let $\phi$ be a WMR with nonnegative weights allowing no ties. Consider the set of all weight vectors representing $\phi$, which is $W \equiv \{w \in \mathbb{R}^N : \sum_{i \in N} w_i(\phi(x)x_i) > 0 \text{ for all } x \in \mathcal{X}\}$ by Lemma 5.1. Note that $W$ is an open convex polyhedron containing a nonnegative vector. This implies that $W$ also contains a positive vector. $\square$

## 5.3 Voting under uncertainty

Assume that $x \in \mathcal{X}$ is randomly drawn according to a probability distribution $p \in \Delta(\mathcal{X}) \equiv \{p \in \mathbb{R}_+^{\mathcal{X}} : \sum_{x \in \mathcal{X}} p(x) = 1\}$. Let

$$p(\phi(x) = x_i) \equiv p(\{x \in \mathcal{X} : \phi(x) = x_i\}) = \sum_{x \in \mathcal{X} : \phi(x) = x_i} p(x)$$

be the probability that $i$'s choice agrees with the collective decision, which is referred to as responsiveness or the Rae index (Rae1969). It is calculated as

$$p(\phi(x) = x_i) = (E_p[\phi(x)x_i] + 1)/2, \tag{5.1}$$

where $E_p[\phi(x)x_i] \equiv \sum_{x \in \mathcal{X}} p(x)\phi(x)x_i$ is the expected value of individual $i$'s payoff $\phi(x)x_i$, because

$$E_p[\phi(x)x_i] = \sum_{x:\phi(x)=x_i} p(x) - \sum_{x:\phi(x)\neq x_i} p(x) = 2p(\phi(x) = x_i) - 1. \tag{5.2}$$

Suppose that, under a voting rule $\phi \in \Phi$ and a probability distribution $p \in \Delta(\mathcal{X})$, the responsiveness of every individual is less than or equal to one-half:

$$p(\phi(x) = x_i) \leq 1/2 \text{ for all } i \in N. \tag{5.3}$$

Let $\phi' \in \Phi$ be the voting rule whose collective decision always disagrees with that of $\phi$, i.e., $\phi'(x) = -\phi(x)$ for all $x \in \mathcal{X}$. We call $\phi'$ the inverse rule of $\phi$. Then, under $\phi'$ and $p$, the responsiveness of every individual is greater than or equal to one-half:

$$p(\phi'(x) = x_i) = p(\phi(x) = -x_i) = 1 - p(\phi(x) = x_i) \geq 1/2 \text{ for all } i \in N. \tag{5.4}$$

That is, the responsiveness under $\phi$ is less than or equal to that under $\phi'$ for all individuals, which is a good enough reason for the group of individuals not to adopt $\phi$.

Given the above, imagine that the individuals agree not to adopt a voting rule if the responsiveness of every individual is less than or equal to one-half. However, if they have no information about the true probability distribution of their preferences facing Knightian uncertainty, they cannot evaluate the responsiveness. Under these circumstances, the worst-case scenario is that the true responsiveness of every individual is less than or equal to one-half. We say that a voting rule is robust if, for any probability distribution of preferences, it avoids this worst-case scenario.

**Definition 5.1.** A voting rule $\phi \in \Phi$ is *robust* if, for each $p \in \Delta(\mathcal{X})$, responsiveness of at least one individual is strictly greater than one-half:

$$\max_{i \in N} p(\phi(x) = x_i) > 1/2 \text{ for all } p \in \Delta(\mathcal{X}). \tag{5.5}$$

For example, a WMR with nonnegative weights is robust if there are no ties. In fact, by Lemma 5.1, $\sum_{i\in N} w_i E_p[\phi(x)x_i] > 0$ for all $p \in \Delta(\mathcal{X})$, so there exists $i \in N$ such that $w_i E_p[\phi(x)x_i] > 0$, i.e., $p(\phi(x) = x_i) > 1/2$.

By replacing "less than or equal to" with "strictly less than" in the above worst-case scenario, we obtain a slightly more severe worst-case scenario in which the responsiveness of every individual is strictly less than one-half. We say that a voting rule is weakly robust if, for any probability distribution of preferences, it avoids this worst-case scenario.

**Definition 5.2.** A voting rule $\phi \in \Phi$ is *weakly robust* if, for each $p \in \Delta(\mathcal{X})$, responsiveness of at least one individual is greater than or equal to one-half:

$$\max_{i\in N} p(\phi(x) = x_i) \geq 1/2 \text{ for all } p \in \Delta(\mathcal{X}). \tag{5.6}$$

For example, a WMR with nonnegative weights is weakly robust even if there are ties. In fact, by Lemma 5.1, $\sum_{i\in N} w_i E_p[\phi(x)x_i] \geq 0$ for all $p \in \Delta(\mathcal{X})$, so there exists $i \in N$ such that $w_i > 0$ and $E_p[\phi(x)x_i] \geq 0$, i.e., $p(\phi(x) = x_i) \geq 1/2$.

If $\phi \in \Phi$ is weakly robust but not robust, there exists $p \in \Delta(\mathcal{X})$ such that

$$\max_{i\in N} p(\phi(x) = x_i) = 1/2.$$

If such $p$ is the true probability distribution, a collective decision is at best neutral to each individual's choice on average, which is never the case with robust rules.

By the equivalence of (5.3) and (5.4), we can rewrite the definitions of robustness and weak robustness as follows, which we will use when robustness and efficiency are compared in Section 5.5.

**Lemma 5.3.** A voting rule $\phi \in \Phi$ is robust if and only if there exists no $p \in \mathcal{X}$ such that

$$p(\phi(x) = x_i) \leq p(\phi'(x) = x_i) \text{ for all } i \in N,$$

and $\phi$ is weakly robust if and only if there exists no $p \in \mathcal{X}$ such that

$$p(\phi(x) = x_i) < p(\phi'(x) = x_i) \text{ for all } i \in N,$$

where $\phi' \in \Phi$ is the inverse rule of $\phi$.

## 5.4 Main results

In this section, we present our main result characterizing robust and weakly robust rules. The result is stated and discussed in Section 5.4.1. The proofs are given in Section 5.4.2.

### 5.4.1 Characterizations

First, we characterize robust rules. A WMR with nonnegative weights allowing no ties is robust as discussed in Section 5.3. Our first main result establishes that every robust rule must be such a WMR.

**Proposition 5.1.** A voting rule is robust if and only if it is a WMR with nonnegative weights such that there are no ties.

Next, we characterize weakly robust rules. A WMR with nonnegative weights is weakly robust even if there are ties as discussed in Section 5.3. Our second main result establishes that every weakly robust rule must be such a WMR.

**Proposition 5.2.** A voting rule is weakly robust if and only if it is a WMR with nonnegative weights.

As a corollary of Proposition 5.1, we characterize robust anonymous rules. If $n$ is odd, a SMR is the unique rule that is both robust and anonymous. In Appendix E.1, we give another characterization of a SMR with odd $n$ using a stronger version of robustness. However, if $n$ is even, no anonymous rule is robust. That is, there is a trade-off between robustness and anonymity.

**Corollary 5.1.** Suppose that $n$ is odd. Then, a voting rule is robust and anonymous if and only if it is a SMR. Suppose that $n$ is even. Then, no voting rule is both robust and anonymous.

*Proof.* A voting rule is robust and anonymous if and only if it is an anonymous WMR with nonnegative weights allowing no ties, which is a SMR with odd $n$. □

Corollary 5.1 implies that an anonymous rule is not robust if $n$ is even or if it is not a SMR. In particular, a supermajority rule is not robust because it is anonymous. To illustrate it by a numerical example, consider a two-thirds rule with a very large number of individuals. Suppose that $x_i = 1$ with probability $p \in (1/2, 2/3)$ independently and identically for each $i \in N$. By the law of large numbers, the group decision is $-1$ with probability close to one, so responsiveness of each individual is close to $1 - p < 1/2$, which implies that this rule is not robust.[8]

As a corollary of Proposition 5.2, we characterize weakly robust anonymous rules. Although no anonymous rule is robust when $n$ is even, there exists a weakly robust anonymous rule regardless of $n$, which is an anonymous SMR.

**Corollary 5.2.** A voting rule is weakly robust and anonymous if and only if it is an anonymous SMR.

*Proof.* A voting rule is weakly robust and anonymous if and only if it is an anonymous WMR with nonnegative weights, which is an anonymous SMR. □

To illustrate the difference between robustness and weak robustness as well as the trade-off between robustness and anonymity, suppose that $n$ is even and consider SMRs. By Proposition 5.1, a SMR with some tie-breaking rule is robust if and only if it is represented as a WMR allowing no ties and, by Corollary 5.1, such a SMR is nonanonymous. For example, a SMR with a casting (tie-breaking) vote is a robust nonanonymous rule. To see this, we consider two cases. First, assume that the presiding officer with a casting vote is a member of the group of $n$ individuals. This rule is equivalent to a WMR such that the presiding officer's weight is slightly greater than the others' weights. Next, assume that the presiding officer is not a member of the group of $n$ individuals and that he or she votes only when there is a tie. This rule is equivalent to a WMR with $n + 1$ individuals including the presiding officer such that the presiding officer's weight is very close to zero. Each of these WMRs does not have ties and is robust

On the other hand, a SMR with any tie-breaking rule is weakly robust by Proposition 5.2. In particular, a SMR with the status quo tie-breaking rule (i.e., the status quo is followed whenever there is a tie) is a weakly robust anonymous rule, but it is not robust by Corollary 5.1.

[8] Even if $n$ is not so large, we can find a probability distribution $p \in \Delta(\mathcal{X})$ such that the responsiveness of every individual is less than or equal to one-half. See an example in Appendix E.6.

### 5.4.2 Proofs

This section provides the proofs of Propositions 5.1 and 5.2. In the proofs, we use the following inequality symbols. For vectors $\xi$ and $\eta$, we write $\xi \geq \eta$ if $\xi_i \geq \eta_i$ for each $i$, $\xi > \eta$ if $\xi_i \geq \eta_i$ for each $i$ and $\xi \neq \eta$, and $\xi \gg \eta$ if $\xi_i > \eta_i$ for each $i$.

We enumerate elements in $\mathcal{X}$ as $\{x^j\}_{j \in M}$, where $M \equiv \{1, \ldots, m\}$ is an index set with $m = 2^n$. Consider an $n \times m$ matrix

$$L = [l_{ij}]_{n \times m} = \left[\phi(x^j) x_i^j\right]_{n \times m}.$$

Note that $l_{ij}$ equals $+1$ if $i$'s choice agrees with the collective decision and $-1$ otherwise. Using this matrix, we can restate the conditions in Proposition 5.1 as follows.

(a) By Lemma 5.1, a voting rule $\phi$ is a WMR with nonnegative weights allowing no ties if and only if there exists $w = (w_i)_{i \in N} \geq 0$ such that

$$\sum_{i \in N} w_i l_{ij} = \sum_{i \in N} w_i \big(\phi(x^j) x_i^j\big) > 0$$

for each $j \in M$, or equivalently, $w^\top L \gg 0$.

(b) By definition, a voting rule is *not* robust if and only if there exists $p = (p_j)_{j \in M} > 0$ such that

$$\sum_{j \in N} l_{ij} p_j = \sum_{j : \phi(x^j) = x_i^j} p_j - \sum_{j : \phi(x^j) \neq x_i^j} p_j \leq 0$$

for each $i \in N$, or equivalently, $Lp \leq 0$.

Proposition 5.1 states that exactly one of (a) and (b) holds. The following theorem of alternatives due to von Neumann and Morgenstern (1944)[9] guarantees that this is true. The same result also appears in Gale(1960, Theorem 2.10) as a corollary of Farkas' lemma.

**Lemma 5.4.** Let $A$ be an $n \times m$ matrix. Exactly one of the following alternatives holds.

- There exists $\xi \in \mathbb{R}^n$ satisfying

$$\xi^\top A \gg 0, \; \xi \geq 0.$$

[9] von Neumann and Morgenstern (1944) use this result to prove the minimax theorem.

- There exists $\eta \in \mathbb{R}^m$ satisfying

$$A\eta \leq 0, \ \ \eta > 0.$$

*Proof of Proposition 5.1.* Plug $L$, $w$, and $p$ into $A$, $\xi$, and $\eta$ in Lemma 5.4, respectively. Then, Lemma 5.4 implies that exactly one of (a) and (b) holds. □

We can interpret Lemma 5.4 as a corollary of the fundamental theorem of asset pricing,[10] which is equivalent to Farkas' lemma. Thus, we can explain why Proposition 5.1 is true in terms of arbitrage-free pricing in an imaginary asset market as discussed in the introduction, which is elaborated in Appendix E.2.

We can prove Proposition 5.2 similarly by restating the conditions in the proposition as follows.

(a′) By Lemma 5.1, a voting rule $\phi$ is a WMR with nonnegative weights if and only if there exists $w = (w_i)_{i \in N} > 0$ such that

$$\sum_{i \in N} w_i l_{ij} = \sum_{i \in N} w_i \big(\phi(x^j) x_i^j\big) \geq 0$$

for each $j \in M$, or equivalently, $w^\top L \geq 0$.

(b′) By definition, a voting rule is *not* weakly robust if and only if there exists $p = (p_j)_{j \in M} > 0$ such that

$$\sum_{j \in N} l_{ij} p_j = \sum_{j : \phi(x^j) = x_i^j} p_j - \sum_{j : \phi(x^j) \neq x_i^j} p_j < 0$$

for each $i \in N$, or equivalently, $Lp \ll 0$.

Proposition 5.2 states that exactly one of (a′) and (b′) holds. To prove it, we use Lemma 5.4 again, but in another way.

*Proof of Proposition 5.2.* Plug $-L^\top$, $p$, and $w$ into $A$, $\xi$, and $\eta$ in Lemma 5.4, respectively, where we replace $(n, m)$ with $(m, n)$. Then, Lemma 5.4 implies that exactly one of (a′) and (b′) holds. □

[10] For details on the fundamental theorem of asset pricing, see Dybvig and Ross, (2003, 2008) and references therein.

## 5.5 Robustness vs. efficiency

Rae (1969) and Taylor (1969) were the first to use responsiveness to characterize voting rules, followed by Straffin (1977) and Fleurbaey (2008). In this section, we discuss their results in comparison to our result.

Note that, by Lemma 5.1, $\phi \in \Phi$ is a WMR with a weight vector $w \in \mathbb{R}^N$ if and only if $\phi(x)\sum_{i\in N} w_i x_i = |\phi(x)\sum_{i\in N} w_i x_i|$ for all $x \in \mathcal{X}$, which is equivalent to the following inequality: for all $\phi' \in \Phi$ and $x \in \mathcal{X}$,

$$\phi(x)\sum_{i\in N} w_i x_i = \left|\phi(x)\sum_{i\in N} w_i x_i\right| = \left|\phi'(x)\sum_{i\in N} w_i x_i\right| \geq \phi'(x)\sum_{i\in N} w_i x_i. \tag{5.7}$$

This is true if and only if, for all $p \in \Delta(\mathcal{X})$,

$$\sum_{i\in N} w_i E_p[\phi(x)x_i] = \max_{\phi'\in\Phi} \sum_{i\in N} w_i E_p[\phi'(x)x_i], \tag{5.8}$$

or equivalently,

$$\sum_{i\in N} w_i p(\phi(x) = x_i) = \max_{\phi'\in\Phi} \sum_{i\in N} w_i p(\phi'(x) = x_i). \tag{5.9}$$

That is, a necessary and sufficient condition for a voting rule to be a WMR is that it maximizes the corresponding weighted sum of responsiveness over all voting rules for each $p \in \Delta(\mathcal{X})$. This result is summarized in the following proposition due to Fleurbaey (2008),[11] where the sufficient condition is weaker.[12]

**Proposition 5.3.** If $\phi$ is a WMR with a weight vector $w$, then *(5.9)* holds for each $p \in \Delta(\mathcal{X})$. For fixed $p \in \Delta(\mathcal{X})^\circ \equiv \{p \in \Delta(\mathcal{X}) : p(x) > 0 \textit{ for each } x \in \mathcal{X}\}$, where every $x$ is possible, if *(5.9)* holds, then $\phi$ is a WMR with a weight vector $w$.

We call the above result the Rae-Taylor-Fleurbaey (RTF) theorem because it generalizes the Rae-Taylor theorem[13] which focuses on a SMR. Note that a WMR in the RTF theorem can have negative weights. In Appendix E.1, we characterize a WMR with possibly negative weights by introducing a further weaker version of robustness.

[11] See also Brighouse and Fleurbaey (2010), who discuss the implication of this result for democracy.

[12] To see why a weaker condition suffices, suppose that $\phi$ is not a WMR. Then, (5.7) does not hold for some $\phi' \in \Phi$ and $x \in \mathcal{X}$, which contradicts (5.8) and (5.9) for each $p \in \Delta(\mathcal{X})^\circ$.

[13] See Rae (1969), Taylor (1969), Straffin (1977), and references in Fleurbaey (2008).

The normative implication of the RTF theorem is efficiency of WMRs.[14] To give the formal definitions of efficiency in this context, we consider the set of random voting rules. For $\phi, \phi' \in \Phi$ and $\lambda \in [0,1]$, the convex combination $\lambda\phi + (1-\lambda)\phi' : \mathcal{X} \to [-1,1]$ is given by $(\lambda\phi+(1-\lambda)\phi')(x) = \lambda\phi(x)+(1-\lambda)\phi'(x)$ for each $x \in \mathcal{X}$. We regard $\phi'' : \mathcal{X} \to [-1,1]$ in the convex hull of $\Phi$ as the following random voting rule: the collective decision is $+1$ with probability $(1+\phi''(x))/2$ and $-1$ with probability $(1-\phi''(x))/2$. Let $\Delta(\Phi)$ denote the set of all random voting rules. For each $\phi \in \Delta(\Phi)$, the responsiveness of individual $i$ is calculated as $(E_p[\phi(x)x_i]+1)/2$.[15]

A voting rule $\phi \in \Delta(\Phi)$ is said to be as good as $\phi' \in \Delta(\Phi)$ under $p \in \Delta(\mathcal{X})$ if $E_p[\phi'(x)x_i] \geq E_p[\phi(x)x_i]$ for all $i \in N$. A voting rule $\phi$ is said to be better than $\phi'$ under $p$ if $E_p[\phi'(x)x_i] \geq E_p[\phi(x)x_i]$ for all $i \in N$ with strict inequality holding for at least one individual. A voting rule $\phi$ is said to be strictly better than $\phi'$ under $p$ if $E_p[\phi'(x)x_i] > E_p[\phi(x)x_i]$ for all $i \in N$. Then, we can define three types of efficiency.

**Definition 5.3.** Fix $p \in \Delta(\mathcal{X})^\circ$. A voting rule $\phi \in \Phi$ is strictly efficient if any other random voting rule is not as good as $\phi$ under $p$. A voting rule $\phi \in \Phi$ is efficient if any random voting rule is not better than $\phi$ under $p$. A voting rule $\phi \in \Phi$ is weakly efficient if any random voting rule is not strictly better than $\phi$ under $p$.

It should be noted that, even if no deterministic voting rule is better than $\phi$, there may exist a random voting rule that is better than $\phi$, which is demonstrated in Appendix E.4.

Using the RTF theorem, we can obtain the following normative characterizations of WMRs.[16]

**Proposition 5.4.** Fix $p \in \Delta(\mathcal{X})^\circ$. A voting rule is strictly efficient if and only if it is a WMR with positive weights such that there are no ties. A voting rule is efficient if and only if it is a WMR with positive weights. A voting rule is weakly efficient if and only if it is a WMR with nonnegative weights.

---

[14] This issue is not formally discussed in Fleurbaey (2008). Instead, Fleurbaey (2008) considers the optimality of a WMR by assuming that $w_i$ is proportional to $i$'s utility, where the weighted sum of responsiveness is the total sum of expected utilities.

[15] When $x \in \mathcal{X}$ is given, the conditional probability that $i$'s decision agrees with the collective decision is $(1+\phi(x))/2$ if $x_i = 1$ and $(1-\phi(x))/2$ if $x_i = -1$. Thus, the conditional probability is equal to $(\phi(x)x_i + 1)/2$.

[16] Another normative implication of Proposition 5.3 is optimality of WMRs in terms of Paretian social preferences, which is immediate from Harsanyi's utilitarianism theorem (Harsanyi, 1955). See Appendix E.3.

*Proof.* To the best of the authors' knowledge, the proofs as well as the above formal statements have never appeared in the literature. We give the proofs in Appendix E.5. □

Propositions 5.1 and 5.4 characterize a WMR using responsiveness. In particular, the class of robust rules coincides with that of strictly efficient rules[17] and the class of weakly robust rules coincides with that of weakly robust rules. On the other hand, efficient rules achieve the optimal outcomes, whereas robust rules avoid the worst outcomes. Thus, robustness together with efficiency gives a dual characterization of WMRs in terms of responsiveness.

Let us elaborate on the difference between efficiency and robustness. Note that if a voting rule $\phi$ is a WMR with positive weights, then, under every $p \in \Delta(\mathcal{X})$, no random voting rule is strictly better than $\phi$, as implied by (5.8). The converse is also true, but a weaker condition suffices. In fact, weak efficiency and weak robustness are distinct sufficient conditions for WMRs with positive weights. Weak efficiency requires that, under a fixed $p \in \Delta(\mathcal{X})^\circ$, any random voting rule is not strictly better than $\phi$. Thus, we need to consider all random voting rules (i.e., all probability distributions of voting rules) to check weak efficiency, where it is assumed that individuals can choose any random voting rule. On the other hand, according to Lemma 5.3, weak robustness requires that, under every $p \in \Delta(\mathcal{X})$, the inverse rule of $\phi$ is not strictly better than $\phi$. Thus, we need to consider all probability distributions of decision profiles to check weak robustness, where it is assumed that individuals do not know the true probability distribution. The following table summarizes the difference between the two sufficient conditions (see also Appendix E.4 for a numerical example).

| | What rule is strictly better than $\phi$? | Under what distribution? |
|---|---|---|
| WMRs | all $\phi' \in \Delta(\Phi)$ | all $p \in \Delta(\mathcal{X})$ |
| weakly efficient rules | all $\phi' \in \Delta(\Phi)$ | fixed $p \in \Delta(\mathcal{X})^\circ$ |
| weakly robust rules | the inverse rule $\phi' \in \Phi$ | all $p \in \Delta(\mathcal{X})$ |

Finally, we emphasize that our characterization of WMRs is not a restatement of the RTF theorem. As discussed in Section 5.3, it is obvious that a WMR is robust. Thus, Proposition 5.4 implies that a strictly efficient rule is robust. However, neither

[17]A robust rule is a WMR with nonnegative weights allowing no ties, which can be represented as a WMR with positive weights allowing no ties by Lemma 5.2.

Proposition 5.4 nor the RTF theorem says anything about whether a robust rule is strictly efficient. Our contribution is to identify the set of all robust rules, which is not implied by the RTF theorem.

## 5.6 Conclusion

The justification of WMRs and, in particular, SMRs based on efficiency arguments or axiomatic characterizations has yielded some of the celebrated contributions to the social choice and voting literature. The two paramount examples rationalizing a SMR within a dichotomous setting are Condorcet's jury theorem and May's theorem,[18] where the rationalization of a voting rule is based on asymptotic (i.e., infinite-individual) probabilistic criteria or deterministic criteria. An alternative approach based on non-asymptotic (i.e., finite-individual) probabilistic criteria was pioneered by Rae (1969), who suggested the aggregate expected net gain or the aggregate responsiveness as a meaningful criterion for evaluating the performance of a voting rule in the constitutional stage, namely, where the veil of ignorance prevails.

This chapter contributes to the latter literature by joining the recently growing literature on economic design with worst-case objectives discussed in the introduction. That is, we introduce normative criteria for voting rules under Knightian uncertainty about individuals' preferences, robustness and weak robustness. Robustness requires that a voting rule should avoid the worst-case scenario in which the true responsiveness of every individual is less than or equal to one-half, and weak robustness requires that a voting rule should avoid the worst-case scenario in which the true responsiveness of every individual is less than one-half. We establish that a voting rule is robust if and only if it is a WMR without any ties and that a voting rule is weakly robust if and only if it is a WMR with any tie-breaking rule when there are ties. We also find that we face a trade-off between robustness and anonymity when the number of individuals is even: we must be content with a nonanonymous rule if we require robustness and we must be content with a weakly robust rule if we require anonymity.

Our result and the RTF theorem (Rae,1969; Taylor, 1969; Fleurbaey, 2008) have in common that both examine WMRs using responsiveness. However, the RTF theorem characterizes WMRs as efficient or weakly efficient rules achieving the optimal outcomes, whereas our result characterizes WMRs as robust or weakly robust rules

[18] See May (1952), Fishburn (1973), and Dasgupta and Maskin (2008).

avoiding the worst outcomes. Hence, our result complements the renowned RTF theorem by providing a dual characterization of WMRs and, in particular, of SMRs in terms of responsiveness under Knightian uncertainty.

# Appendix A Appendix to Chapter 1

## A.1 Proof of Proposition 1.1

To prove Proposition 1.1, we use some lemmas. We first show that the switching strategies constitute an equilibrium and each cutoff point is uniquely determined. After that, we show that this strategy is the unique strategy survived by the iterated deletion of interim dominated strategies.

**Lemma A.1.** The equilibrium cut off points $\hat{\theta}_B$ and $\hat{\theta}_C$ are the solutions of the following simultaneous equations:

$$
\begin{aligned}
(\frac{\alpha_B}{\gamma_B}+\frac{\alpha_C}{\gamma_C})\hat{\theta}_B \quad &+ \quad \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B-\frac{\alpha_B}{\gamma_B}\eta-\frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big)-1 \\
&- \quad (\frac{\alpha_B}{\gamma_B}+\frac{\alpha_C}{\gamma_C})\eta-\frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})+\frac{\beta_C}{\gamma_C}\Phi(\sqrt{2}\frac{\eta}{\sigma})-\frac{\beta_C}{\gamma_C}=0. \quad \text{(A.1)} \\
(\frac{\alpha_B}{\gamma_B}+\frac{\alpha_C}{\gamma_C})\hat{\theta}_C \quad &+ \quad \sqrt{2}\sigma\frac{\alpha_B}{\gamma_B}\Phi^{-1}\Big(\frac{\alpha_C}{\gamma_C}\hat{\theta}_C+\frac{\alpha_C}{\gamma_C}\eta-\frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma})\Big)-1 \\
&+ \quad (\frac{\alpha_B}{\gamma_B}+\frac{\alpha_C}{\gamma_C})\eta-\frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})+\frac{\beta_C}{\gamma_C}\Phi(\sqrt{2}\frac{\eta}{\sigma})-\frac{\beta_C}{\gamma_C}=0. \quad \text{(A.2)}
\end{aligned}
$$

*Proof.* By rearranging the equations (1.1) and (1.2), we have

$$
\Phi(\frac{\theta_C-\theta_B}{\sqrt{2}\sigma}+\sqrt{2}\frac{\eta}{\sigma})=\frac{\alpha_B}{\gamma_B}\theta_B-\frac{\alpha_B}{\gamma_B}\eta-\frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}), \tag{A.3}
$$

$$
\Phi(\frac{\theta_B-\theta_C}{\sqrt{2}\sigma}-\sqrt{2}\frac{\eta}{\sigma})=\frac{\alpha_C}{\gamma_C}\theta_C+\frac{\alpha_C}{\gamma_C}\eta-\frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma}). \tag{A.4}
$$

Let $\Phi^{-1}(\cdot):[0,1]\to\mathbb{R}$ be the inverse function of $\Phi(\cdot)$. By using this function, we obtain the following expressions,

$$
\theta_C=\theta_B-2\eta+\sqrt{2}\sigma\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\theta_B-\frac{\alpha_B}{\gamma_B}\eta-\frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big), \tag{A.5}
$$

$$
\theta_B=\theta_C+2\eta+\sqrt{2}\sigma\Phi^{-1}\Big(\frac{\alpha_C}{\gamma_C}\theta_C+\frac{\alpha_C}{\gamma_C}\eta-\frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma})\Big). \tag{A.6}
$$

Define $A = \frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})$. By substituting (A.5) into (A.6), we have

$$\begin{aligned}\hat{\theta}_B &= \hat{\theta}_B - 2\eta + \sqrt{2}\sigma\Phi^{-1}(A) + 2\eta \\ &+ \sqrt{2}\sigma\Phi^{-1}\Big(\frac{\alpha_C}{\gamma_C}\hat{\theta}_B - 2\frac{\alpha_C}{\gamma_C}\eta + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}(A) + \frac{\alpha_C}{\gamma_C}\eta - \frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma})\Big).\end{aligned}$$

By rearranging it, we have

$$\Phi^{-1}(A) + \Phi^{-1}\Big(\frac{\alpha_C}{\gamma_C}\hat{\theta}_B - 2\frac{\alpha_C}{\gamma_C}\eta + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}(A) + \frac{\alpha_C}{\gamma_C}\eta - \frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma})\Big) = 0.$$

Note that $-\Phi^{-1}(x) = \Phi^{-1}(y) \Leftrightarrow 1 - x = y$. Therefore, we have

$$1 - A = \frac{\alpha_C}{\gamma_C}\hat{\theta}_B - 2\frac{\alpha_C}{\gamma_C}\eta + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}(A) + \frac{\alpha_C}{\gamma_C}\eta - \frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma}).$$

Moreover, this is equivalent to

$$\begin{aligned}1 &- \frac{\alpha_B}{\gamma_B}\hat{\theta}_B + \frac{\alpha_B}{\gamma_B}\eta + \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}) \\ &= \frac{\alpha_C}{\gamma_C}\theta_B - 2\frac{\alpha_C}{\gamma_C}\eta + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}(A) + \frac{\alpha_C}{\gamma_C}\eta - \frac{\beta_C}{\gamma_C}\Phi(-\sqrt{2}\frac{\eta}{\sigma}).\end{aligned}$$

Finally, this gives us the required expression

$$\begin{aligned}(\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B})\theta_B &+ \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \\ &- 1 - (\frac{\alpha_B}{\gamma_B} + \frac{\alpha_C}{\gamma_C})\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}) + \frac{\beta_C}{\gamma_C}\Phi(\sqrt{2}\frac{\eta}{\sigma}) - \frac{\beta_C}{\gamma_C} = 0,\end{aligned}$$

where, in the last part, we use the fact that $\Phi(x) + \Phi(-x) = 1$ for all $x \in \mathbb{R}$.

By the similar calculation for $\hat{\theta}_C$, we also have (A.2). □

**Lemma A.2.** The cutoff points $\hat{\theta}_B$ and $\hat{\theta}_C$ exist.

*Proof.* For the equation (A.1), the left hand side is continuous with respect to $\hat{\theta}_B$. Also, note that $0 \leq \frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}) \leq 1 \Leftrightarrow \theta_B^*(\sigma,\eta) \leq \hat{\theta}_B \leq \theta_B^{**}(\sigma,\eta)$. By differentiating the left hand side of the equation (A.1) with respect to $\hat{\theta}_B$, we have

$$\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B} + \sqrt{2}\sigma\frac{\alpha_B\alpha_C}{\gamma_B\gamma_C}\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) > 0.$$

Therefore, the left hand side is strictly increasing with respect to $\hat{\theta}_B$. Also, note that

$$\begin{aligned}(\frac{\alpha_C}{\gamma_C} &+ \frac{\alpha_B}{\gamma_B})\hat{\theta}_B + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \\ &- 1 + \frac{t}{\gamma_C} - (\frac{\alpha_B}{\gamma_B} + \frac{\alpha_C}{\gamma_C})\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}) + \frac{\beta_C}{\gamma_C}\Phi(\sqrt{2}\frac{\eta}{\sigma}) - \frac{\beta_C}{\gamma_C} \to -\infty,\end{aligned}$$

$$\text{as } \hat{\theta}_{\mathrm{B}} \to \theta^{*}_{\mathrm{B}}(\sigma,\eta),$$

$$\begin{aligned}(\frac{\alpha_C}{\gamma_C} &+ \frac{\alpha_B}{\gamma_B})\hat{\theta}_B + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \\ &- 1 - (\frac{\alpha_B}{\gamma_B} + \frac{\alpha_C}{\gamma_C})\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma}) + \frac{\beta_C}{\gamma_C}\Phi(\sqrt{2}\frac{\eta}{\sigma}) - \frac{\beta_C}{\gamma_C} \to \infty,\end{aligned}$$

$$\text{as } \hat{\theta}_{\mathrm{B}} \to \theta^{**}_{\mathrm{B}}(\sigma,\eta).$$

By the above calculations, there exists the unique value $\hat{\theta}_B$ such that $\hat{\theta}_B \in [\theta^*_B(\sigma,\eta), \theta^{**}_B(\sigma,\eta)]$. The same argument for (A.2) show that there exists the unique value $\hat{\theta}_C$ such that $\hat{\theta}_C \in [\theta^*_C(\sigma,\eta), \theta^{**}_C(\sigma,\eta)]$. □

To see the uniqueness of $\hat{\theta}_B$ and $\hat{\theta}_C$, we consider the following iterated procedure. For each signal $x_i$ and $n \in \mathbb{Z}_+$, define the following strategies,

$$\text{Market B}: \quad s_i^B(x_i) = \begin{cases} W & \text{if } x_i \le b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), \\ NW & \text{if } x_i > b_B^{n-1}(\overline{\theta}_B, \overline{\theta}_C), \end{cases}$$

$$\text{Market C}: \quad s_i^C(x_i) = \begin{cases} A & \text{if } x_i \le b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C), \\ NA & \text{if } x_i > b_C^{n-1}(\overline{\theta}_B, \overline{\theta}_C), \end{cases}$$

where $\underline{\theta}_B = \underline{\theta}_C = -\infty, \overline{\theta}_B = \overline{\theta}_C = \infty$ and

$$b_B^n(\theta_B, \theta_C) = \begin{cases} \underline{\theta}_B & \text{if } n = 0, \\ b_B(b_B^{n-1}(\theta_B, \theta_C), b_C^{n-1}(\theta_B, \theta_C)) & \text{if } n \ge 1, \end{cases}$$

$$b_C^n(\theta_B, \theta_C) = \begin{cases} \underline{\theta}_C & \text{if } n = 0, \\ b_C(b_B^{n-1}(\theta_B, \theta_C), b_C^{n-1}(\theta_B, \theta_C)) & \text{if } n \ge 1. \end{cases}$$

Let $s$ be the strategy profile such that player $i$ takes $s_i^B(\cdot)$ if $i$ is in the market B and takes $s_i^C(\cdot)$ if $i$ is in the market C.

**Lemma A.3.** A strategy profile $s$ is a unique equilibrium.

*Proof.* We only show the iteration starting from $(\underline{\theta}_B, \underline{\theta}_C)$. The case of $(\overline{\theta}_B, \overline{\theta}_C)$ can be shown symmetrically. We prove it by induction. It is clearly satisfied for $n = 1$ since they are dominance actions.

Suppose that each player follows the strategy until $n = k \geq 1$. Then, note that

$$Prob(s_j^B(x_j) = W|x_i) \geq Prob(s_j(b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C)) = W|x_i),$$

and

$$Prob(s_j^C(x_j) = A|x_i) \geq Prob(s_j(b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C)) = A|x_i).$$

Therefore, given opponents' strategies, the (maximin) expected payoff from taking NW is less than or equal to $V_B(x_i, (b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C)))$. By the definition of $b_B$, we can say that

$$V_B(x_i, (b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C))) \leq 0$$

if $x_i \leq b_B(b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C))$, which implies that following $s^B$ is the best response at $n = k + 1$. Similarly, the (maximin) expected payoff from taking A is greater than or equal to $V_C(x_i, (b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C)))$. By the definition of $b_C$, we can say that

$$V_C(x_i, (b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C))) \geq 0$$

if $x_i \leq b_B(b_B^{n-1}(\underline{\theta}_B, \underline{\theta}_C), b_C^{n-1}(\underline{\theta}_B, \underline{\theta}_C))$, which implies that following $s^C$ is the best response at $n = k + 1$. □

**Lemma A.4.**

$$\lim_{n\to\infty} b_B^n(\underline{\theta}_B, \underline{\theta}_C) = \lim_{n\to\infty} b_B^n(\overline{\theta}_B, \overline{\theta}_C) = \hat{\theta}_B$$

and

$$\lim_{n\to\infty} b_C^n(\underline{\theta}_B, \underline{\theta}_C) = \lim_{n\to\infty} b_C^n(\overline{\theta}_B, \overline{\theta}_C) = \hat{\theta}_C.$$

*Proof.* We also only show the iteration starting from $(\underline{\theta}_B, \underline{\theta}_C)$. The case of $(\overline{\theta}_B, \overline{\theta}_C)$ can be shown symmetrically. By the properties of $V_B$ and $V_C$, we have $b_B(\theta_B, \theta_C) < \hat{\theta}_B, b_C(\theta_B, \theta_C) < \hat{\theta}_C$ if $(\theta_B, \theta_C) < (\hat{\theta}_B, \hat{\theta}_C)$. Also, $b_B^1(\underline{\theta}_B, \underline{\theta}_C) = b_B(\underline{\theta}_B, \underline{\theta}_C) > \underline{\theta}_B$ and $b_C^1(\underline{\theta}_B, \underline{\theta}_C) = b_C(\underline{\theta}_B, \underline{\theta}_C) > \underline{\theta}_C$. Therefore, $b_B^2(\underline{\theta}_B, \underline{\theta}_C) \geq b_B^1(\underline{\theta}_B, \underline{\theta}_C)$ and $b_C^2(\underline{\theta}_B, \underline{\theta}_C) \geq b_C^1(\underline{\theta}_B, \underline{\theta}_C)$. Thus, the both sequences $\{b_B^n(\underline{\theta}_B, \underline{\theta}_C)\}_{n=0}^{\infty}$ and $\{b_C^n(\underline{\theta}_B, \underline{\theta}_C)\}_{n=0}^{\infty}$ are increasing and bounded by $\hat{\theta}_B$ and $\hat{\theta}_C$, which imply $\lim_{n\to\infty} b_B^n(\underline{\theta}_B, \underline{\theta}_C) = \hat{\theta}_B$, and $\lim_{n\to\infty} b_C^n(\underline{\theta}_B, \underline{\theta}_C) = \hat{\theta}_C$. □

Finally, Proposition 1.1 follows from Lemma A.2, A.3 and A.4. □

## A.2 Proof of Proposition 1.2

Let $\eta = 0$ in (A.1). Then, by differentiating both sides of the equation with respect to $\sigma$, we have

$$\begin{aligned}(\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B})\frac{\partial\tilde{\theta}_B}{\partial\sigma} &+ \sqrt{2}\frac{\alpha_C}{\gamma_C}\Phi^{-1}(\frac{\alpha_B}{\gamma_B}\tilde{\theta}_B - \frac{1}{2}\frac{\beta_B}{\gamma_B}) \\ &+ \sqrt{2}\sigma\frac{\alpha_B\alpha_C}{\gamma_B\gamma_C}\frac{\partial\tilde{\theta}_B}{\partial\sigma}\Phi^{-1\prime}(\frac{\alpha_B}{\gamma_B}\tilde{\theta}_B - \frac{1}{2}\frac{\beta_B}{\gamma_B}) = 0.\end{aligned}$$

This is equivalent to

$$\begin{aligned}\{\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B} + \sqrt{2}\sigma\frac{\alpha_B\alpha_C}{\gamma_B\gamma_C}\Phi^{-1\prime}(\frac{\alpha_B}{\gamma_B}\tilde{\theta}_B - \frac{1}{2}\frac{\beta_B}{\gamma_B})\}\frac{\partial\tilde{\theta}_B}{\partial\sigma} \\ = -\sqrt{2}\frac{\alpha_C}{\gamma_C}\Phi^{-1}(\frac{\alpha_B}{\gamma_B}\tilde{\theta}_B - \frac{1}{2}\frac{\beta_B}{\gamma_B}).\end{aligned}$$

Since $\Phi^{-1\prime} > 0$ and $\Phi < 0$ if and only if $0 \leq x < \frac{1}{2}$, we obtain that

$$\begin{aligned}\frac{1}{2}\frac{\beta_B}{\alpha_B} \leq \tilde{\theta}_B < \frac{1}{2}\frac{\gamma_B}{\alpha_B} + \frac{1}{2}\frac{\beta_B}{\alpha_B} &\Rightarrow \frac{\partial\tilde{\theta}_B}{\partial\sigma} > 0, \\ \frac{1}{2}\frac{\gamma_B}{\alpha_B} + \frac{1}{2}\frac{\beta_B}{\alpha_B} \leq \tilde{\theta}_B \leq \frac{\gamma_B}{\alpha_B} + \frac{1}{2}\frac{\beta_B}{\alpha_B} &\Rightarrow \frac{\partial\tilde{\theta}_B}{\partial\sigma} \leq 0.\end{aligned}$$

Note that $\frac{1}{2}\frac{\beta_B}{\alpha_B} = \theta_B^*$, $\frac{1}{2}\frac{\gamma_B}{\alpha_B} + \frac{1}{2}\frac{\beta_B}{\alpha_B} = \frac{1}{2}(\theta_B^* + \theta_B^{**})$, and $\frac{\gamma_B}{\alpha_B} + \frac{1}{2}\frac{\beta_B}{\alpha_B} = \theta_B^{**}$. Therefore, we obtain the result.

We can apply the same argument for $\hat{\theta}_C$, which completes the proof. □

## A.3 Proof of Proposition 1.3

By differentiating both sides of equation (A.1) with respect to $\sigma$, we have

$$\begin{aligned}&\sqrt{2}\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) + \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \\ &\times\Big(\frac{\alpha_B}{\gamma_B}\frac{\partial\hat{\theta}_B}{\partial\sigma} + \sqrt{2}\frac{\beta_B}{\gamma_B}\frac{\eta}{\sigma^2}\phi(\sqrt{2}\frac{\eta}{\sigma})\Big) + (\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B})\frac{\partial\hat{\theta}_B}{\partial\sigma} + \frac{\sqrt{2}}{\sigma}\frac{\beta_B}{\gamma_B}\phi(\sqrt{2}\frac{\eta}{\sigma}) - \frac{\sqrt{2}}{\sigma}\frac{\beta_C}{\gamma_C}\phi(\sqrt{2}\frac{\eta}{\sigma}) = 0.\end{aligned}$$

This is equivalent to

$$a \times \frac{\partial\hat{\theta}_B}{\partial\sigma} = -b - c + \Big(\frac{\beta_C}{\gamma_C} - \frac{\beta_B}{\gamma_B}\Big)d.$$

where

$$
\begin{aligned}
a &= \frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B} + \sqrt{2}\sigma\frac{\alpha_B\alpha_C}{\gamma_B\gamma_C}\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) > 0,\\
b &= \sqrt{2}\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big),\\
c &= \frac{\alpha_C\beta_B}{\gamma_B\gamma_C}\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big)\phi(\sqrt{2}\frac{\eta}{\sigma})\frac{\eta}{\sigma} > 0,\\
d &= \sqrt{2}\frac{\eta}{\sigma^2}\phi(\sqrt{2}\frac{\eta}{\sigma}) > 0.
\end{aligned}
$$

Define $X_\sigma = \frac{\gamma_B\gamma_C(b+c)}{d}$. Then, we can see that $\frac{\partial\hat{\theta}_B}{\partial\sigma} < 0$ if and only if $\beta_C\gamma_B - \beta_B\gamma_C < X_\sigma$. The same argument can be applied for $\frac{\partial\hat{\theta}_C}{\partial\sigma}$. □

## A.4 Proof of Proposition 1.4

By differentiating the both side of the equation (A.1) with respect to $\eta$, we have

$$
\begin{aligned}
&\sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Big(\frac{\alpha_B}{\gamma_B}\frac{\partial\hat{\theta}_B}{\partial\eta} - \frac{\alpha_B}{\gamma_B} - \frac{\beta_B}{\gamma_B}\frac{\sqrt{2}}{\sigma}\phi(\sqrt{2}\frac{\eta}{\sigma})\Big)\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big)\\
&\quad + (\frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B})\frac{\partial\hat{\theta}_B}{\partial\eta} - \frac{\alpha_B}{\gamma_B} - \frac{\alpha_C}{\gamma_C} - \frac{\sqrt{2}}{\sigma}\frac{\beta_B}{\gamma_B}\phi(\sqrt{2}\frac{\eta}{\sigma}) + \frac{\sqrt{2}}{\sigma}\frac{\beta_C}{\gamma_C}\phi(\sqrt{2}\frac{\eta}{\sigma}) = 0.
\end{aligned}
$$

This is equivalent to

$$e \times \frac{\partial\hat{\theta}_B}{\partial\eta} = f + g + \Big(\frac{\beta_B}{\gamma_B} - \frac{\beta_C}{\gamma_C}\Big)h$$

where

$$
\begin{aligned}
e &= \frac{\alpha_C}{\gamma_C} + \frac{\alpha_B}{\gamma_B} + \sqrt{2}\sigma\frac{\alpha_B\alpha_C}{\gamma_B\gamma_C}\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) > 0,\\
f &= \sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Big(\frac{\alpha_B}{\gamma_B} + \frac{\beta_B}{\gamma_B}\frac{\sqrt{2}}{\sigma}\phi(\sqrt{2}\frac{\eta}{\sigma})\Big)\Phi^{-1\prime}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) > 0,\\
g &= \frac{\alpha_B}{\gamma_B} + \frac{\alpha_C}{\gamma_C} > 0,\\
h &= \frac{\sqrt{2}}{\sigma}\phi(\sqrt{2}\frac{\eta}{\sigma}) > 0.
\end{aligned}
$$

Define $X_\eta = \frac{\gamma_B\gamma_C(f+g)}{h}$. Then, we can show that $\frac{\partial\hat{\theta}_B}{\partial\eta} < 0$ if and only if $X_\eta < \beta_C\gamma_B - \beta_B\gamma_C$. The same argument can be applied for $\frac{\partial\hat{\theta}_C}{\partial\eta}$. □

## A.5 Proof of Proposition 1.5

(i) We only show the case of $\theta_B^*(\xi) < \theta_C^*(\xi) < \theta_B^{**}(\xi) < \theta_C^{**}(\xi)$. By a similar argument, we can show the other cases. Assume that $\theta_B^*(\xi) < \theta_C^*(\xi) < \theta_B^{**}(\xi) < \theta_C^{**}(\xi)$. We first show that $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0}$. Suppose, on the contrary, that $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} - \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} > 0$. As $\sigma \to 0$ and $\eta \to 0$, we have

$$\frac{\hat{\theta}_C - \hat{\theta}_B}{\sqrt{2}\sigma} + \sqrt{2}\frac{\eta}{\sigma} = \frac{\hat{\theta}_C - \hat{\theta}_B + 2\eta}{\sqrt{2}\sigma} \to -\infty,$$

$$\frac{\hat{\theta}_B - \hat{\theta}_C}{\sqrt{2}\sigma} - \sqrt{2}\frac{\eta}{\sigma} = \frac{\hat{\theta}_B - \hat{\theta}_C - 2\eta}{\sqrt{2}\sigma} \to \infty.$$

Therefore, by (A.3) and (A.4),

$$\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \frac{\beta_B}{\alpha_B}\xi = \theta_B^*(\xi),$$
$$\hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \frac{\gamma_C}{\alpha_C} + \frac{\beta_C}{\alpha_C} - \frac{\beta_C}{\alpha_C}\xi = \theta_C^{**}(\xi).$$

However, by our hypothesis, $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} - \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \theta_B^*(\xi) - \theta_C^{**}(\xi) > 0$, which contradicts our assumption about $\theta_B^*(\xi) < \theta_C^*(\xi) < \theta_B^{**}(\xi) < \theta_C^{**}(\xi)$. Hence, we have $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} - \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} \leq 0$. By the symmetric argument, we have $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} - \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} \geq 0$, which means that $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0}$.

We next show that $\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}(\xi)$ where $\hat{\theta}(\xi) = (\gamma_B\gamma_C + r\gamma_C - t\gamma_B + \beta_C\gamma_B + (\beta_B\gamma_C - \beta_C\gamma_B)\xi)/(\alpha_B\gamma_C + \alpha_C\gamma_B)$. It suffices to show that, in equation (A.1), $\sqrt{2}\sigma\frac{\alpha_C}{\gamma_C}\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \to 0$ as $\sigma \to 0$ and $\eta \to 0$. Since $\hat{\theta}_C$ and $\hat{\theta}_C$ satisfy

$$\hat{\theta}_C = \hat{\theta}_B - 2\eta + \sqrt{2}\sigma\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big),$$

and

$$\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} = \hat{\theta}(\xi),$$

we obtain

$$\sqrt{2}\sigma\Phi^{-1}\Big(\frac{\alpha_B}{\gamma_B}\hat{\theta}_B - \frac{\alpha_B}{\gamma_B}\eta - \frac{\beta_B}{\gamma_B}\Phi(\sqrt{2}\frac{\eta}{\sigma})\Big) \to 0.$$

Therefore, we have

$$(\frac{\alpha_C}{\gamma_C}+\frac{\alpha_B}{\gamma_B})\hat{\theta}(\xi)-1-\frac{\beta_B}{\gamma_B}\xi+\frac{\beta_C}{\gamma_C}\xi-\frac{\beta_C}{\gamma_C}=0.$$

as $\sigma\to 0$ and $\eta\to 0$. Finally, this is equivalent to

$$\begin{aligned}\hat{\theta}(\xi) &= \frac{\gamma_B\gamma_C+r\gamma_C-t\gamma_B+\beta_C\gamma_B+(\beta_B\gamma_C-\beta_C\gamma_B)\xi}{\alpha_B\gamma_C+\alpha_C\gamma_B}\\ &= \frac{\alpha_B\gamma_C}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_B^*+\frac{\alpha_C\gamma_B}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_C^{**}\\ &= \frac{\alpha_B\gamma_C}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_B^{**}+\frac{\alpha_C\gamma_B}{\alpha_B\gamma_C+\alpha_C\gamma_B}\theta_C^*\end{aligned}$$

as we desired.

(ii) Suppose that $\theta_B^*(\xi)<\theta_B^{**}(\xi)<\theta_C^*(\xi)<\theta_C^{**}(\xi)$. This implies that $\hat{\theta}_B<\hat{\theta}_C$. As $\sigma\to 0$ and $\eta\to 0$, we have

$$\begin{aligned}\frac{\hat{\theta}_C-\hat{\theta}_B}{\sqrt{2}\sigma}+\sqrt{2}\frac{\eta}{\sigma} &= \frac{\hat{\theta}_C-\hat{\theta}_B+2\eta}{\sqrt{2}\sigma}\to\infty,\\ \frac{\hat{\theta}_B-\hat{\theta}_C}{\sqrt{2}\sigma}-\sqrt{2}\frac{\eta}{\sigma} &= \frac{\hat{\theta}_B-\hat{\theta}_C-2\eta}{\sqrt{2}\sigma}\to-\infty.\end{aligned}$$

By (A.3) and (A.4), we have

$$\begin{aligned}\hat{\theta}_B(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} &= \frac{\gamma_B}{\alpha_B}+\frac{\beta_B}{\alpha_B}\xi=\theta_B^{**}(\xi),\\ \hat{\theta}_C(\sigma,\eta)|_{\sigma\to 0,\eta\to 0} &= \frac{\beta_C}{\alpha_C}-\frac{\beta_C}{\alpha_C}\xi=\theta_C^*(\xi).\end{aligned}$$

The same argument is applied for case (iii). □

## A.6 Proof of Proposition 1.6

(i) By the direct calculation, we have

$$\frac{d\hat{\theta}(\xi)}{d\xi}=\frac{\beta_B\gamma_C-\beta_C\gamma_B}{\alpha_B\gamma_C+\alpha_C\gamma_B},$$

which implies that

$$\beta_B\gamma_C\geq\beta_C\gamma_B\Leftrightarrow\frac{d\hat{\theta}(\xi)}{d\xi}\geq 0.$$

(ii) By the direct calculation, we have,

$$\frac{d\theta_B^{**}}{d\xi} = \frac{\beta_B}{\alpha_B} > 0,$$
$$\frac{d\theta_C^{*}}{d\xi} = -\frac{\beta_C}{\alpha_C} < 0.$$

(iii) By the direct calculation, we have,

$$\frac{d\theta_B^{*}}{d\xi} = \frac{\beta_B}{\alpha_B} > 0,$$
$$\frac{d\theta_C^{**}}{d\xi} = -\frac{\beta_C}{\alpha_C} < 0.$$

□

# Appendix B Appendix to Chapter 2

## B.1 Proof of Lemma 2.1

($\Leftarrow$) Suppose $g|_S = g'|_S$ implies $v_g^S = v_{g'}^S$. Then, for any $S \in 2^N$,

$$v_g(S) = \sum_{T \subset 2^N} v_g^T u_T(S) = \sum_{T \subset S} v_g^T.$$

Note that $g|_S = g'|_S$ implies $g|_T = g'|_T$ for any $T \subset S$. Thus, $v_g(S) = v_{g'}(S)$, which means that $\{v_g\}_{g \in \mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$.

($\Rightarrow$) Suppose that $\{v_g\}_{g \in \mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$. Note that

$$v_g^S = \sum_{T \subset S} (-1)^{|S \setminus T|} v_g(T),$$

and, for all $T \subset S$, $v_g(T) = v_{g'}(T)$ if $g|_S = g'|_S$. Thus, $g|_S = g'|_S$ implies $v_g^S = v_{g'}^S$. $\square$

## B.2 Proof of Theorem 2.1

We first show the following result which provides one of the necessary and sufficient conditions for the existence of network potentials. This result is similar to that of Slade (1994) and Facchini et al. (1997) in noncooperative games. The following lemma will be used in the proof of Theorem 2.1.

**Lemma B.1.** A network game $\phi = (\phi_i)_{i \in N}$ admits a network potential if and only if there exist functions $\omega : \mathbb{G}^N \to \mathbb{R}$ and $\lambda_i : \mathbb{G}^{N \setminus \{i\}} \to \mathbb{R}$ such that, for any $g \in \mathbb{G}^N$ and $i \in N$,

$$\phi_i(g) = \omega(g) + \lambda_i(g|_{N \setminus \{i\}})$$

where $\omega$ is a potential function.

*Proof.* ($\Leftarrow$) By the direct calculation, for any $g \in \mathbb{G}^N, i \in N$ and for any $j \in N_i(g)$, we obtain

$$\begin{aligned}\phi_i(g) - \phi_i(g - ij) =& (\omega(g) - \lambda_i(g|_{N\setminus\{i\}})) - (\omega(g - ij) - \lambda_i((g - ij)|_{N\setminus\{i\}}))\\ =& \omega(g) - \omega(g - ij)\end{aligned}$$

because $g|_{N\setminus\{i\}} = (g - ij)|_{N\setminus\{i\}}$.

($\Rightarrow$) Define $\lambda_i(g) = \phi_i(g + ij) - \omega(g + ij)$ for any $g \in \mathbb{G}^{N\setminus\{i\}}$ and $j \in N\setminus\{i\}$. By definition of the network potential, this value is well-defined, which completes the proof. □

Now, we are in a position to prove the theorem. Proof is dividend into three steps. First, we show that (i) $\Leftarrow$ (iii). Second, we show that (iii) $\Leftarrow$ (i). Finally, we show that (i) $\Leftrightarrow$ (ii).

(i) $\Leftarrow$ (iii). Let $\{\zeta_S\}_{S\in 2^N}$ be an interaction potential satisfying the conditions. Define $\omega(g) = \sum_{S\in 2^N} \zeta_S(g|_S)$. Then

$$\begin{aligned}\omega(g) - \omega(g - ij) =& \sum_{S\in 2^N} \zeta_S(g|_S) - \sum_{S\in 2^N} \zeta_S((g - ij)|_S)\\ =& \sum_{S\in 2^N, i\in S} \zeta_S(g|_S) + \sum_{S\in 2^N, i\notin S} \zeta_S(g|_S)\\ &- \sum_{S\in 2^N, i\in S} \zeta_S((g - ij)|_S) - \sum_{S\in 2^N, i\notin S} \zeta_S((g - ij)|_S)\\ =& \sum_{S\in 2^N, i\in S} \zeta_S(g|_S) - \sum_{S\in 2^N, i\in S} \zeta_S((g - ij)|_S)\\ =& \phi_i(g) - \phi_i(g - ij),\end{aligned}$$

where the third equality follows from the observation that $g|_S = (g - ij)|_S$ if $i \notin S$. Thus, $\phi$ admits a network potential $\omega(\cdot)$.

(i) $\Rightarrow$ (iii). Let $\omega(\cdot)$ be a network potential. By Lemma B.1, let $\lambda_i(g|_{N\setminus\{i\}}) = \phi_i(g) - \omega(g)$. For $S \in 2^N$, let us define

$$\zeta_S(g|_S) = \begin{cases} \omega(g) + \sum_{i\in N} \lambda_i(g|_{N\setminus\{i\}}) & \text{if } S = N,\\ -\lambda_i(g|_{N\setminus\{i\}}) & \text{if } S = N\setminus\{i\} \text{ for some } i,\\ 0 & \text{if } |S| \leq |N| - 2.\end{cases}$$

Then each $i \in N, S \in 2^N$, and $g|_S \in \mathbb{G}^S$,

$$\begin{aligned}\sum_{S\in 2^N, i\in S} \zeta_S(g|_S) =& \sum_{j\in N\setminus\{i\}} \zeta_{N\setminus\{j\}}(g|_{N\setminus\{j\}}) + \zeta_N(g)\\ =& - \sum_{j\in N\setminus\{i\}} \lambda_j(g|_{N\setminus\{j\}}) + \omega(g) + \sum_{j\in N} \lambda_j(g|_{N\setminus\{j\}})\\ =&\omega(g) + \lambda_i(g|_{N\setminus\{i\}}) = \phi_i(g).\end{aligned}$$

Also, for each $S \in 2^N$ and $g|_S \in \mathbb{G}^S$,

$$\begin{aligned}\sum_{S\in 2^N} \zeta_S(g|_S) =& \sum_{j\in N} \zeta_N(g|_{N\setminus\{j\}}) + \zeta_N(g)\\ =& - \sum_{j\in N} \lambda_j(g|_{N\setminus\{j\}}) + \omega(g) + \sum_{j\in N} \lambda_j(g|_{N\setminus\{j\}})\\ =&\omega(g).\end{aligned}$$

(i) $\Leftrightarrow$ (ii). By Lemma 2.1, there is a one-to-one correspondence between $\{v_g\}_{g\in\mathbb{G}^N}$ and $\{\zeta_S\}_{S\in 2^N}$ such that

$$\zeta_S(g|_S) = v_g^S/|S|.$$

Then, by the equivalence of (i) and (iii) and the fact that

$$\sum_{S\in 2^N, i\in S} \zeta_S(g|_S) = \sum_{S\in 2^N, i\in S} v_g^S/|S| = \psi(v_g),$$

we establish the result. $\square$

## B.3 Proof of Corollary 2.1

Given a value function $\tilde{v} \in \tilde{\mathcal{G}}_N$ and network $g \in \mathbb{G}^N$, let us define a TU-game such that $\tilde{v}_g(S) = \tilde{v}(g|_S)$ for each $S \in 2^N$. By construction, for each $g, g' \in \mathbb{G}^N$, if $g|_S = g'|_S$, then $\tilde{v}_g(S) = \tilde{v}_{g'}(S)$. This implies that $\{\tilde{v}_g\}_{g\in\mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$. Moreover, we can see that, for each $i \in N$,

$$
\begin{aligned}
\phi_i(g) = f_i^{MJW}(\tilde{v}, g) &= \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(\tilde{v}(g|_{S\cup\{i\}}) - \tilde{v}(g|_S)) \\
&= \sum_{S \subset N, i \in S} \frac{(|S|-1)!(|N|-|S|)!}{|N|!}(\tilde{v}(g|_S) - \tilde{v}(g|_{S\setminus\{i\}})) \\
&= \sum_{S \subset N, i \in S} \frac{(|S|-1)!(|N|-|S|)!}{|N|!}(v_g(S) - v_g(S\setminus\{i\})) \\
&= \psi_i(\tilde{v}_g).
\end{aligned}
$$

Therefore, by Theorem 2.1, $\phi$ admits a potential $\omega$. Its potential is given by

$$
\omega(g) = \sum_{S \subset N} \frac{(|S|-1)!(|N|-|S|)!}{|N|!}(\tilde{v}(g|_S)).
$$

□

## B.4 Proof of Corollary 2.2

By construction, $U_{\phi,g}(S) = U_{\phi,g'}(S)$ if $g|_S = g'|_S$ for each $S \in 2^N$. Thus, $(U_{\phi,g})_{g\in\mathbb{G}^N} \in \mathcal{G}_{N,\mathbb{G}^N}$. By Theorem 2.1, $\phi$ admits a network potential $\omega$. □

## B.5 Proof of Theorem 2.3

($\Rightarrow$). By Theorem 2.1, let $\{v_g\}_{g\in\mathbb{G}^N}$ be a TU game on networks corresponding to the network game $\phi = (\phi_i)_{i\in N}$. For any $l \in A$, let $v_l = v_{\sigma(l)}$. Note that for any $l \in A, l_S = l'_S$ implies $\sigma(l)|_S = \sigma(l')|_S$ because $g|_S$ is not affected by the actions of $N\setminus S$ for all $g \subset g^N$ and $S \in 2^N$. Then, this implies that $v_l(S) = v_{\sigma(l)}(S) = v_{\sigma(l')}(S) = v_{l'}(S)$. Hence, $\{v_l\}_{l\in A}$ is a TU game with action choices and which satisfies $\pi_i(l) = \phi_i(\sigma(l)) = \psi_i(v_{\sigma(l)}) = \psi_i(v_l)$. By Theorem 2.2, we obtain the result.

($\Leftarrow$). By Theorem 2.2, let $\{v_l\}_{l\in A}$ be a TU game with action choices corresponding the potential game $\Gamma_\phi$. Let us define the function $h : \mathbb{G}^N \to A$ such that $h(g) = \hat{l}_g$ for all $g \in \mathbb{G}^N$. This function is well-defined because $\hat{l}_g$ is unique for all $g \in \mathbb{G}^N$. Next, define $v_g = v_{h(g)}$ for all $g \in \mathbb{G}^N$. By definition of $\hat{l}_g$, $g|_S = g'|_S$ implies $\hat{l}_g|_S = \hat{l}_{g'}|_S$. Then, this implies that $v_g(S) = v_{g'}(S)$ by construction. Hence, $\{v_g\}_{g\in\mathbb{G}^N}$ is a TU game on networks and which satisfies $\phi_i(g) = \phi_i(g(\hat{l}_g)) = \pi_i(\hat{l}_g) = \psi_i(v_{\hat{l}_g}) = \psi_i(v_g)$. By Theorem 2.1, we obtain the result. □

## B.6 Additional discussion

We discuss the relation with the Shapley value consistency and the network characteristic function. As a counter-example for that the Shapley value consistency is not a necessary condition for the existence of a network potential, Chakrabarti and Gilles (2007) give the following example. Each player's payoff function and corresponding network potential $\omega$ are given in Table B.1.

| Network | $\phi_1(g)$ | $\phi_2(g)$ | $\phi_3(g)$ | $\omega(g)$ |
|---|---|---|---|---|
| $g_0 = \emptyset$ | 0 | 0 | 0 | 0 |
| $g_1 = \{12\}$ | 1 | 1 | 2 | 1 |
| $g_2 = \{13\}$ | 1 | 0 | 1 | 1 |
| $g_3 = \{23\}$ | 0 | 1 | 1 | 1 |
| $g_4 = \{12, 13\}$ | 3 | 2 | 4 | 3 |
| $g_5 = \{12, 23\}$ | 2 | 3 | 4 | 3 |
| $g_6 = \{13, 23\}$ | 2 | 2 | 3 | 3 |
| $g_7 = g^N$ | 5 | 5 | 7 | 6 |

Table B.1: Example 3.8 of Chakrabarti and Gilles (2007).

For $g_1$, we can construct $U_{\phi,g_1}$ as

$$U_{\phi,g_1}(S) = \begin{cases} 2 & \text{if } S = \{1,2\}, \\ 4 & \text{if } S = N, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\psi(U_{\phi,g_1}) = (\frac{5}{3}, \frac{5}{3}, \frac{2}{3}) \neq (1, 1, 2) = \phi(g_1)$, which implies that the network game $\phi$ is not Shapley-consistent.

Next, consider the following network characteristic function $\{v_g\}_{g \in \mathbb{G}^N}$ given in Table B.2.

| $v_g \backslash S$ | $\{1\}$ | $\{2\}$ | $\{3\}$ | $\{1,2\}$ | $\{1,3\}$ | $\{2,3\}$ | $N$ |
|---|---|---|---|---|---|---|---|
| $v_{g_0}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $v_{g_1}$ | 0 | 0 | 0 | $-2$ | 0 | 0 | 4 |
| $v_{g_2}$ | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| $v_{g_3}$ | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| $v_{g_4}$ | 0 | 0 | 0 | $-2$ | 2 | 0 | 9 |
| $v_{g_5}$ | 0 | 0 | 0 | $-2$ | 0 | 2 | 9 |
| $v_{g_6}$ | 0 | 0 | 0 | 0 | 2 | 2 | 7 |
| $v_{g_7}$ | 0 | 0 | 0 | $-2$ | 2 | 2 | 17 |

Table B.2: An example of $\{v_g\}_{g \in \mathbb{G}^N}$ such that $\phi_i(g) = \psi_i(v_g)$.

We can show that, for any $g \in \mathbb{G}^N, \phi(g) = \psi(v_g)$. Therefore, this network game $\phi$ admits a network potential $\omega$ by Theorem 2.1 although it is not Shapley value consistent.

# Appendix C Appendix to Chapter 3

## C.1 Proof of Proposition 3.1

Take any $\phi \notin C(\mathcal{E})$. We want to show that $\phi \notin S(\mathcal{E})$. By definition, $\phi \notin C(\mathcal{E})$ means that $\phi$ is dominated by some matching $\phi'$ via coalition $U$. If $\phi$ satisfies (IR), otherwise we are done, $\phi'(w) \in M$ for all $w \in U$. Indeed, if $\phi(w) = w$, we can say that $\phi'(w) \neq w$ by definition. If $\phi(w) \neq w$, since $\phi$ satisfies (IR) and $\phi'$ dominates $\phi$, $\phi'(w) \neq w$. Let $w$ be in $U$ and $m = \phi'(w)$. Then, by definition,

$$w = \phi'(m) \succ_m \phi(m) \text{ and } m = \phi'(w) \succ_w \phi(w),$$

which means that $(m, w)$ blocks $\phi$. Hence, $\phi \notin S(\mathcal{E})$.

We want to show the other direction. Take any $\phi \notin S(\mathcal{E})$. If $\phi$ does not satisfy (IR) for an agent $i \in M \cup W$, (s)he can block $\phi$ by a single coalition. If $\phi$ is blocked by some pairs, then this pair blocks $\phi$. Thus, $\phi \notin C(\mathcal{E})$. $\square$

## C.2 Proof of Proposition 3.2

Let $\succ^* = f(\succ)$ be the preference profile obtained from the completion of $\succ$ and $\phi^{DA} = g^{DA} \circ f(\succ) = g^{DA}(\succ^*)$ denoted by the obtained matching. Since $\phi^{DA}$ is stable with respect to $\succ^*$, $\phi^{DA}$ satisfies (IR) and there is no blocking pair with respect to $\succ^*$. We want to show that $\phi^{DA}$ satisfies (IR) and there is no blocking pair with respect to $\succ$.

By (IR) with respect to $\succ^*$, either $\phi^{DA}(i) \succ_i^* i$ or $\phi^{DA}(m) = i$ for all $i \in M \cup W$. In either case, by definition of $\succ^*$, we have $i \not\succ_i \phi^{DA}(i)$, which implies that $\phi^{DA}$ satisfies (IR) with respect to $\succ$.

Suppose that there is a pair $(m, w) \in M \times W$ blocks $\phi^{DA}$ with respect to $\succ$. Then, this is equivalent to that $w \succ_m \phi^{DA}(m)$ and $m \succ_w \phi^{DA}(w)$. However, since $\succ^*$ is obtained from the completion of $\succ$, we must have $w \succ_m^* \phi^{DA}(m)$ and $m \succ_w^* \phi^{DA}(w)$, which implies that a pair $(m, w) \in M \times W$ blocks $\phi^{DA}$ with respect to $\succ^*$, a contradiction. Therefore, there is no blocking pair with respect to $\succ$. $\square$

## C.3 Proof of Proposition 3.3

Take any $\succ \in \mathcal{P}$ as a true preference profile. Let $\succ^* = f(\succ)$ be the preference profile obtained from the completion of $\succ$. Consider the hypothetical situation that $\succ^*$ is a true preference profile. Since $g^{DA}$ is strategy-proof, for any $m \in M$, $\succ'_m \in \mathcal{P}_m$ and $\succ_{-m} \in \mathcal{P}_{-m}$, we have

$$\begin{aligned} g^{DA} \circ f(\succ_m, \succ_{-m}) &= g^{DA}\big(\succ^*_m, f_{-m}(\succ)\big) \\ &\succsim^*_m g^{DA}\big(f_m(\succ'_m, \succ_{-m}), f_{-m}(\succ'_m, \succ_{-m})\big) \\ &= g^{DA} \circ f(\succ'_m, \succ_{-m}). \end{aligned}$$

Then, this implies that for any $m \in M$, $\succ'_m \in \mathcal{P}_m$ and $\succ_{-m} \in \mathcal{P}_{-m}$, we have

$$\begin{aligned} g^{DA} \circ f(\succ'_m, \succ_{-m}) &= g^{DA}\big(f_m(\succ'_m, \succ_{-m}), f_{-m}(\succ'_m, \succ_{-m})\big) \\ &\not\succ_m g^{DA}\big(\succ^*_m, f_{-m}(\succ)\big) \\ &= g^{DA} \circ f(\succ_m, \succ_{-m}). \end{aligned}$$

Therefore, $g^{DA} \circ f$ is strategy-proof. □

## C.4 Proof of Proposition 3.4

Since $\mathcal{P}^r \subset \mathcal{P}^{acyc}$ and $f(\succ) = \succ$ for all $\succ \in \mathcal{P}^r$, $g \circ f = g$ on $\mathcal{P}^r$. Then, by the fact that $g^{DA}$ is the unique mechanism which induces a stable matching and satisfies strategy-proof for $M$ (Dubins and Freedman, 1981; Roth, 1982, and Alcalde and Barberà, 1994), we must have $g = g^{DA}$. However, we show that there exists an economy $\mathcal{E}$ in which $g^{DA} \circ f$ does not induce the $M$-optimal matching for any $f \in \mathcal{F}$. To see this, consider the following example. Let $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$. We assume that each agent is acceptable for every $i \in M \cup W$. Other specification of preference profile is given as follows. Note that there is no other relation (i.e., incomparable) than the specification:

$m_1 : w_1 \succ_{m_1} w_2$ and $w_1 \succ_{m_1} w_3$,
$m_2 : w_1 \succ_{m_2} w_3$ and $w_3 \succ_{m_2} w_2$,
$m_3 : w_3 \succ_{m_3} w_1$ and $w_3 \succ_{m_3} w_2$,

$w_1 : m_3 \succ_{w_1} m_2$ and $m_2 \succ_{w_1} m_1$,

$w_2 : m_1 \succ_{w_2} m_3$ and $m_2 \succ_{w_2} m_3$,
$w_3 : m_1 \succ_{w_3} m_2$ and $m_3 \succ_{w_3} m_2$.

In this economy, there are four stable matchings:

$$\phi^1 = \{(m_1, w_1), (m_2, w_2), (m_3, w_3)\},\ \phi^2 = \{(m_1, w_2), (m_2, w_1), (m_3, w_3)\},$$

$$\phi^3 = \{(m_1, w_3), (m_2, w_1), (m_3, w_2)\},\ \phi^4 = \{(m_1, w_3), (m_2, w_2), (m_3, w_1)\}.$$

Note that $\phi^1$ is the $M$-optimal matching for $(\succ_m)_{m\in M}$. However, for any completion rule $f \in \mathcal{F}$, $\phi^1$ is not stable with respect to $f(\succ)$ because $w_1 f_{m_2}(\succ) w_2$ and $m_2 f_{w_1}(\succ) m_1$, which means that a pair $(m_2, w_1)$ is a blocking pair in $f(\succ)$.[1] Therefore, $g^{DA} \circ f$ does not induce $\phi^1$ as an outcome. □

## C.5 Relation with Bernheim and Rangel (2009)

We introduce a formal definition of the unambiguous preference and the core based on the relation by Bernheim and Rangel (2009).

For each man $m \in M$, let $C_m : 2^{W\cup\{m\}} \to W \cup \{m\}$ be a choice function for $m$.[2] Similarly, for each woman $w \in W$, let $C_w : 2^{M\cup\{w\}} \to M \cup \{w\}$ be a choice function for $m$. Let us call $\mathcal{E} = (M, W, (C_i)_{i\in M\cup W})$ an economy. A choice function $C$ is rational if there exists a strict, complete and transitive preference $\succ$ such that $C(X) = \text{argmax}_{\succ} X$ where $X$ is an arbitrarily choice set. It is known that a choice function is rational if and only if it satisfies so called independence of irrelevant alternatives (hereafter, we call it IIA).[3]

[1] An example of $f(\succ)$ is as follows:
$m_1 : w_1 \succ_{m_1} w_2 \succ_{m_1} w_3 \succ_{m_1} m_1$,
$m_2 : w_1 \succ_{m_2} w_3 \succ_{m_2} w_2 \succ_{m_2} m_2$,
$m_3 : w_3 \succ_{m_3} w_1 \succ_{m_3} w_2 \succ_{m_3} m_3$,

$w_1 : m_3 \succ_{w_1} m_2 \succ_{w_1} m_1 \succ_{w_1} w_1$,
$w_2 : m_2 \succ_{w_2} m_1 \succ_{w_2} m_3 \succ_{w_2} w_2$,
$w_3 : m_3 \succ_{w_3} m_1 \succ_{w_3} m_2 \succ_{w_3} w_3$.

As we claimed, we can see that $(m_2, w_1)$ is a blocking pair for $\phi^1$.

[2] $C_i(\emptyset) = \emptyset$ for all $i \in M \cup W$ for convenience.

[3] Let $\mathcal{X}$ be the whole set of choice alternatives. We say that a choice function $C : 2^{\mathcal{X}} \to \mathcal{X}$ satisfies *IIA* if for any $X, Y$ such that $Y \subset X$, if $C(X) \in C(Y)$, then $C(X) = C(Y)$. Also, this condition is

For any $w, w' \in W$, we say that $m$ *unambiguously prefers* $w$ to $w'$, denoted by $\succ_m^{UA}$, if $C_m(S) \neq w'$ for all $S \supset \{w, w'\}$. If $m$'s choice behavior shows this phenomenon for a pair $w$ and $w'$, we can deduce $m$ cannot prefer $w'$ to $w$ in all the possible situation where both of them can be chosen. The following result shows that this relation is acyclic.

**Lemma C.1.** $\succ_m^{UA}$ is *acyclic.*

*Proof.* Suppose that $w^1 \succ_m^{UA} w^2 \succ_m^{UA} \cdots \succ_m^{UA} w^k$ for some $w^1, w^2, \cdots, w^k \in W \cup \{m\}$. Let consider the set $S = \{w^1, w^2, \cdots, w^k\}$. By definition of $\succ_m^{UA}$, we must have $C(S) = w_1$. Thus, $w_k \not\succ_m^{UA} w_1$. □

We say that a matching $\phi'$ *dominates* $\phi$ in the sense of Bernheim and Rangel (2009) if there exists $U \subset M \cup W$ such that for any $i \in U$, (1) $\phi'(i) \succ_i^{UA} \phi(i)$ and (2) $\phi'(i) \in U$. Then, core defined by Bernheim and Rangel (2009) is the set of matchings which are not dominated by any matchings in above sense. That is, our definition of stability is consistent with the core defined by Bernheim and Rangel (2009).

If agent's choice function satisfies WARP, then $\succ^{UA}$ rationalizes his choice function, and hence, we can use $\succ_m^{UA}$ as his actual rational preference. Thus, one can think that our assumption of acyclic preferences is a reduced form of Bernheim and Rangel (2009).

**Proposition C.1.** For each $m \in M$, $\succ_m^{UA}$ is complete, transitive and $C_m(X) = \text{argmax}_{\succ_m^{UA}} X$ for any $X \subset W \cup \{m\}$ if and only if the choice function $C_m : 2^{W \cup \{m\}} \to W \cup \{m\}$ satisfies IIA (or, equivalently, WARP).

*Proof.* Suppose that $\succ_m^{UA}$ is complete and transitive. Take any $X, Y \subset W \cup \{m\}$ such that $Y \subset X$ and suppose that $C_m(X) \in Y$. We want to show that $C_m(Y) = C_m(X)$. Suppose not, then by completeness of $\succ_m^{UA}$, either $C_m(X) \succ_m^{UA} C_m(Y)$ or $C_m(Y) \succ_m^{UA} C_m(X)$ holds. If the former case, by definition of $\succ_m^{UA}, C_m(Y) \neq C_m(Y)$, which is a contradiction. If the latter case, since $C_m(Y) \in X$ and, again by definition of $\succ_m^{UA}, C_m(X) \neq C_m(X)$, which is also a contradiction. Therefore, we must conclude that $C_m(Y) = C_m(X)$.

equivalent to the next one, called *WARP* : for any $X, Y$ such that $Y \subset X$ and for any $x \in Y$, if $x \neq C(Y)$, then $x \neq C(X)$.

Next, we show the other direction. Suppose that the choice function $C_m : 2^{W\cup\{m\}} \to W \cup \{m\}$ satisfies IIA (and hence, it satisfies WARP). Take any $w, w' \in W \cup \{m\}$ and suppose that, without loss of generality, $C_m(\{w, w'\}) = w$. Then by WARP, for any $X \subset W \cup \{m\}$ such that $\{w, w'\} \subset X, w' \neq C_m(X)$. This means that $w \succ_m^{UA} w'$. Since this is satisfied for any pair $w, w' \in W \cup \{m\}$, $\succ_m^{UA}$ is complete, i.e., for any $w, w' \in W \cup \{m\}, C_m(\{w, w'\}) = w \Leftrightarrow w \succ_m^{UA} w'$. Then, suppose that $w \succ_m^{UA} w'$ and $w' \succ_m^{UA} w''$. Then, $C_m(\{w, w', w''\}) = w$. By WARP, for any $X \subset W \cup \{m\}$ such that $\{w, w''\} \subset X, w'' \neq C_m(X)$. Thus, $w \succ_m^{UA} w''$, which implies that $\succ_m^{UA}$ is transitive. Finally, by definition of $\succ^{UA}$, $x = C_m(X)$ implies that $y \not\succ_m^{UA} x$ for all $y \in X\backslash\{x\}$ and $X \subset W \cup \{m\}$. Since $\succ_m^{UA}$ is complete, $x \not\succ_m^{UA} y$, which implies $C_m(X) = \text{argmax}_{\succ_m^{UA}} X$ for any $X \subset W \cup \{m\}$. $\square$

# Appendix D Appendix to Chapter 4

## D.1 Proof of Theorem 4.2

Let $\mathcal{G}^c \subset \mathcal{G}$ denote the set of games such that $v(N) = c$. Also, for any player $i \in N$ and $c \in \mathbb{R}$, let $\mathcal{G}^{c,i}$ denote the set of games such that $v(N) = c$ and $i$ is null player. Note that $\mathcal{G}^{c,i} \subset \mathcal{G}^c$ for all $i \in N$ and $c \in \mathbb{R}$. Let $\Delta_i(v) = (v(S \cup \{i\}) - v(S))_{S \subseteq N \setminus \{i\}} \in \mathbb{R}^{2^{(N-1)}}$ be a vector of marginal contributions of $i$ in $v$. Therefore player $i \in N$ is a null player in $v$ if $\Delta_i(v) = \mathbf{0}$. Let $\Lambda^i$ be the set of all vectors of marginal contribution of $i$: $\Lambda^i = \{\Delta_i(v) | v \in \mathcal{G}\}$.

For each $x \in \mathbb{R}^N$, let $m_x \in \mathcal{G}$ be an additive game, $m_x(S) = \sum_{i \in S} x_i$ for all $S \subseteq N$. Let $\mathcal{G}^{add}$ be the set of additive games. Since there is a one-to-one correspondence between $x \in \mathbb{R}^N$ and an additive game $m_x$, we can identify $\mathcal{G}^{add}$ with $\mathbb{R}^N$. Abe and Nakada (2017) provide the following result, which can be useful later.

**Theorem D.1** (Abe and Nakada, 2017)**.** Let $n \neq 2$. $f : \mathcal{G}^{add} \to \mathbb{R}^n$ satisfies (E), (M$^-$), (NY), and (RIN) if and only if there exists a $\delta \in [0, 1]$ and a $w \in \mathcal{W}$ such that $f_i(x) = \delta \cdot x_i + (1 - \delta) \cdot w_i \cdot \sum_{l \in N} x_l$ for all $x \in \mathbb{R}^n$ and $i \in N$.

Now, we offer the proof of Theorem 4.2. It is clear that the rule satisfies all the axioms. We suppose that a rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E), (M$^-$), (RIN), (WDMSP), and (NY).

**Claim 1**: For each $i \in N$, there exist functions $\phi_i : \Lambda^i \times \mathbb{R} \to \mathbb{R}$ and $\alpha_i : \mathbb{R} \to \mathbb{R}$ such that $f_i(v) = \phi_i(\Delta_i(v), v(N)) + \alpha_i(v(N))$.

We first take any $c \in \mathbb{R}$. For any $i \in N$ and $v \in \mathcal{G}^c$, we have the following equation: for any $\bar{v} \in \mathcal{G}^c$ such that $\Delta_i(v) = \Delta_i(\bar{v})$,

$$f_i(v) \overset{(\mathrm{M}^-)}{=} f_i(\bar{v}) \ =: \alpha_i(c, \Delta_i(v)). \tag{D.1}$$

Specifically, we denote

$$\alpha_i(c) = \alpha_i(c, \mathbf{0}). \tag{D.2}$$

Moreover, for any $i \in N$ and $v, v' \in \mathcal{G}^c$, we have

$$\begin{aligned} f_i(v) - f_i(v') &\overset{(D.1)}{=} \alpha_i(c, \Delta_i(v)) - \alpha_i(c, \Delta_i(v')) \\ &=: \phi_i(\Delta_i(v), \Delta_i(v'), c). \end{aligned} \tag{D.3}$$

Hence, for any $i \in N$ and $v \in \mathcal{G}^c$, we obtain the following equation: for any $v' \in \mathcal{G}^{c,i}$,

$$\phi_i(\Delta_i(v), \Delta_i(v'), c) \overset{(D.3)}{=} f_i(v) - f_i(v') \overset{(D.1)}{=} f_i(v) - \alpha_i(c). \tag{D.4}$$

Note that $f_i(v) - \alpha_i(c)$ is independent from $v' \in \mathcal{G}^{c,i}$. For any $i \in N$ and $v \in \mathcal{G}^c$ let

$$\phi_i(\Delta_i(v), c) := f_i(v) - \alpha_i(c). \tag{D.5}$$

Hence, for any $i \in N$ and $v \in \mathcal{G}$, we obtain

$$f_i(v) \overset{(D.5)}{=} \phi_i(\Delta_i(v), v(N)) + \alpha_i(v(N)). \tag{D.6}$$

This completes Claim 1.

**Claim 2**: For any $i \in N$, and $c \in \mathbb{R}$, the function $\phi_i(\cdot, c) : \Lambda^i \to \mathbb{R}$ satisfies (M) within $\mathcal{G}^c$: for any $v, v' \in \mathcal{G}^c$, if $v(S \cup \{i\}) - v(S) \geq v'(S \cup \{i\}) - v'(S)$ for any $S \subset N \setminus \{i\}$, then $\phi(\Delta_i(v), c) \geq \phi(\Delta_i(v'), c)$. Moreover, $\phi_i(\mathbf{0}, c) = 0$.

Let $c = v(N) = v'(N)$. We have

$$\begin{aligned} \phi(\Delta_i(v), c) - \phi(\Delta_i(v'), c) &= \phi(\Delta_i(v), c) + \alpha_i(c) - (\phi(\Delta_i(v'), v'(N)) + \alpha_i(c)) \\ &\overset{\text{C1}}{=} f_i(v) - f_i(v') \\ &\overset{(\text{M}^-)}{\geq} 0. \end{aligned}$$

Moreover, for any $c \in \mathbb{R}$,

$$\phi_i(\mathbf{0}, c) \overset{(D.5),(D.4)}{=} \phi_i(\mathbf{0}, \mathbf{0}, c) \overset{(D.3),(D.1)}{=} \alpha_i(c, \mathbf{0}) - \alpha_i(c, \mathbf{0}) = 0. \tag{D.7}$$

**Claim 3**: The function $\phi$ is symmetric: for any $v \in \mathcal{G}$ and $i, j \in N$, if $i, j$ is symmetric in $v$, then $\phi_i(\Delta_i(v), v(N)) = \phi_j(\Delta_j(v), v(N))$.

For any $i, j \in N$ and $v \in \mathcal{G}$ such that $v(S \cup \{i\}) - v(S) = v(S \cup \{j\}) - v(S)$ for all $S \subseteq N \setminus \{i, j\}$, let $v' = v(N)u_{N\setminus\{i,j\}}$. Then, we have

$$\phi_i(\Delta_i(v), v(N)) \overset{(D.6)}{=} f_i(v) - f_i(v') \overset{\text{(WDMSP)}}{=} f_j(v) - f_j(v') \overset{(D.6)}{=} \phi_j(\Delta_j(v), v(N)). \quad \text{(D.8)}$$

This completes Claim 3.

**Claim 4**: The function $\phi$ satisfies $\delta$-efficiency ($\delta$-E): there is a $\delta \in [0, 1]$ such that $\sum_{i\in N} \phi_i(\Delta_i(v), v(N)) = \delta v(N)$ for any $v \in \mathcal{G}$.

Let $\tilde{f} : \mathcal{G}^{add} \to \mathbb{R}$ be the restriction of $f$ on $\mathcal{G}^{add}$. Then, by Theorem D.1, for each $m_x \in \mathcal{G}^{add}$, we have

$$\begin{aligned} \tilde{f}_i(m_x) &= \phi_i(\Delta_i(m_x), \sum_{l\in N} x_l) + \alpha_i(\sum_{l\in N} x_l) \\ &= \delta \cdot x_i + (1 - \delta) \cdot w_i \cdot \sum_{l\in N} x_l. \end{aligned} \quad \text{(D.9)}$$

for some $\delta \in [0, 1]$ and $w \in \mathcal{W}$. In particular, for $x_i = 0$, (D.7) implies $\alpha_i(\sum_{l\in N} x_l) = (1 - \delta) \cdot w_i \cdot \sum_{l\in N} x_l$. Hence, from Claim 1, it follows that for each $v \in \mathcal{G}$,

$$f_i(v) = \phi_i(\Delta_i(v), v(N)) + (1 - \delta) \cdot w_i \cdot v(N). \quad \text{(D.10)}$$

Since $f$ satisfies (E), $\sum_{i\in N} f_i(v) = \sum_{i\in N} \phi_i(\Delta_i(v), v(N)) + (1-\delta)v(N) = v(N)$, which implies that $\phi$ satisfies the following property:

$$(\delta\text{-E}) : \sum_{i\in N} \phi_i(\Delta_i(v), v(N)) = \delta v(N).$$

This completes Claim 4.

**Claim 5**: There is a $\delta \in [0, 1]$ and a $w \in \mathcal{W}$ such that $f_i(x) = \delta \cdot Sh_i(v) + (1 - \delta) \cdot w_i \cdot v(N)$ for any $v \in \mathcal{G}$.

Fixing $c \in \mathbb{R}$, we write $\phi^c_i(v) = \phi_i(\Delta_i(v), c)$ for each $v \in \mathcal{G}^c$. The function $\phi^c(v) : \mathcal{G}^c \to \mathbb{R}^N$ satisfies ($\delta$-E), (M) and (SYM) within $\mathcal{G}^c$ by Claim 2, 3 and 4. Hence, by the same argument of Theorem 4.1, we have

$$\phi^c_i(v) = \delta Sh_i(v).$$

Since $c \in \mathbb{R}$ is arbitrarily chosen, for any $v \in \mathcal{G}$, we have

$$\phi_i(\Delta_i(v), v(N)) = \phi_i^{v(N)}(v) = \delta Sh_i(v) \tag{D.11}$$

Finally, by (D.10) and (D.11), we obtain $f_i(x) = \delta \cdot Sh_i(v) + (1-\delta) \cdot w_i \cdot v(N)$, which completes the proof. □

## D.2 Independence of axioms for Theorem 4.2

The independence of the axioms is shown in the examples listed below.

**Example D.2.1**. Consider the following function: for any $i \in N$ and $v \in \mathcal{G}$,

$$f_i^{\mathrm{E}}(v) = 0.$$

This function satisfies all axioms except (E).

**Example D.2.2**. Consider the following function: for any $i \in N$ and $v \in \mathcal{G}$,

$$f_i^{\mathrm{M}^-}(v) = 2Sh_i(v) - \frac{v(N)}{n}.$$

This function satisfies all axioms except ($\mathrm{M}^-$).

**Example D.2.3**. Consider the following function: for any $i \in N$ and $v \in \mathcal{G}$,

$$f_i^{\mathrm{RIN}}(v) = \delta Sh_i + (1-\delta)\frac{i + v(N)^2}{\bar{N} + n(v(N))^2} v(N),$$

where $\bar{N} = \sum_{i \in N} i = \frac{n(n-1)}{2}$ and $i$ is the natural number representing player $i$. This rule satisfies (E), ($\mathrm{M}^-$), (WMDSP) and (NY) but not (RIN). To check ($\mathrm{M}^-$), let $h_i(a) = \frac{i+a^2}{\bar{N}+a^2}a = \frac{ia+a^3}{\bar{N}+a^2}$. Then, we have $\frac{dh_i(a)}{da} = \frac{na^4+(3\bar{N}-ni)a^2+i\bar{N}}{(\bar{N}+a^2)^2} > 0$ for all $a \in \mathbb{R}$ because $na^4 + i\bar{N} > 0$ for all $i$ and $3\bar{N} - ni \geq \frac{n(n-3)}{2} \geq 0$.

**Example D.2.4**. Fix a permutation of all players $\sigma$. Consider the following function: for any $i \in N$ and $v \in \mathcal{G}$,

$$f_i^{\mathrm{WMDSP}}(v) = v(P_i^\sigma \cup \{i\}) - v(P_i^\sigma)$$

where $P_i^\sigma$ is the set of predecessors of $i$ in $\sigma$. This function satisfies all the axioms except (WMDSP).

**Example D.2.5**. Consider the following function: for any $i \in N$ and $v \in \mathcal{G}$,

$$f_i^{\mathrm{NY}}(v) = \begin{cases} Sh_1 + 10 & \text{if} \quad i = 1, \\ Sh_i - \frac{10}{n-1} & \text{if} \quad i \neq 1. \end{cases}$$

This rule satisfies all the axioms except (NY).

## D.3 A counterexample to Theorem 4.2 and 4.3 for $n = 2$

Theorem 4.2 and 4.3 fail for $n = 2$. Consider the following allocation rule $f^\heartsuit$ on $N = \{1, 2\}$:

$$(f_1^\heartsuit(v,w), f_2^\heartsuit(v,w)) = \begin{cases} (Sh_1(v), Sh_2(v)), & Sh_1(v) \geq 0 \text{ and } Sh_2(v) \geq 0, \\ (0, v(N)), & Sh_1(v) < 0 \text{ and } Sh_2(v) > 0 \wedge v(N) \geq 0, \\ (v(N), 0), & Sh_1(v) < 0 \text{ and } Sh_2(v) > 0 \wedge v(N) < 0, \\ (Sh_1(v), Sh_2(v)), & Sh_1(v) \leq 0 \text{ and } Sh_2(v) \leq 0, \\ (0, v(N)), & Sh_1(v) > 0 \text{ and } Sh_2(v) < 0 \wedge v(N) \leq 0, \\ (v(N), 0), & Sh_1(v) > 0 \text{ and } Sh_2(v) < 0 \wedge v(N) > 0, \end{cases}$$

for any $v \in \mathcal{G}$ and $w \in \mathcal{W}$.

Note that this function does not depend on $w$. It is clear that $f^\heartsuit$ satisfies (E$^*$) and (M$^{-*}$). It satisfies (SYM$^*$) because if the players 1 and 2 are symmetric in the sense of marginal contribution and have the same weight, they receive $(Sh_1(v), Sh_2(v))$. It satisfies (RIN$^*$) because if the players 1 and 2 are null players, the game $v$ is the null game: $v(12) = v(1) = v(2) = 0$. Since $f^\heartsuit$ does not depend on $w$, it clearly satisfies (FEC$^*$). By the same argument, the uniqueness for Theorem 4.2 fails for $n = 2$.

## D.4 Proof of Theorem 4.3

We first show that (E$^*$), (M$^{-*}$), (SYM$^*$) characterizes the egalitarian Shapley value when we fix $w = (\frac{1}{n}, ..., \frac{1}{n})$, which can be useful in later.

**Lemma D.1.** Suppose that $\mathcal{W} = \{(\frac{1}{n}, ..., \frac{1}{n})\}$ and $n \neq 2$. Then, an allocation rule $f : \mathcal{G} \times \mathcal{W} \to \mathbb{R}^N$ satisfies (E$^*$), (M$^{-*}$), (SYM$^*$) if and only if it is an egalitarian-Shapley value.

*Proof.* This follows from the axioms and arguments in Casajus and Huettner (2014) if $w = (\frac{1}{n}, ..., \frac{1}{n})$. $\square$

Now, we offer the proof of Theorem 4.3. It is clear that the rule satisfies all the axioms. We suppose that a rule $f : \mathcal{G} \to \mathbb{R}^N$ satisfies (E$^*$), (M$^{-*}$), (RIN$^*$), (SYM$^*$), and (FEC$^*$). Claim 1 can be thought of as an analog of that of Theorem 4.2. The differences lie in Claims 2-4. In this proof, we first specify the form of the Shapley value, while we first specify the weighted division in Theorem 4.2.

**Claim 1**: For each $i \in N$, there exists functions $\phi_i(v) : \Lambda^i \times \mathbb{R} \to \mathbb{R}$ and $\alpha_i : \mathcal{W} \times \mathbb{R} \to \mathbb{R}$ such that $f_i(v, w) = \phi_i(\Delta_i(v), v(N)) + \alpha_i(w, v(N))$.

We first take any $c \in \mathbb{R}$. For any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$, we have the following equation: for any $\bar{v} \in \mathcal{G}^c$ such that $\Delta_i(v) = \Delta_i(\bar{v})$,

$$f_i(v, w) \overset{(\mathrm{M}^{-*})}{=} f_i(\bar{v}, w) =: \alpha_i(w, c, \Delta_i(v)). \tag{D.12}$$

Specifically, we denote

$$\alpha_i(w, c) = \alpha_i(w, c, \mathbf{0}). \tag{D.13}$$

By (FEC$^*$), for any $c \in \mathbb{R}$ and $i \in N$, there is a function $\phi_i^c : \mathcal{G}^c \to \mathbb{R}$ such that

$$\begin{aligned} \phi_i^c(v) &= f_i(v, w) - f_i(cu_{N\setminus\{i\}}, w) \\ &\overset{(D.13)}{=} f_i(v, w) - \alpha_i(w, c). \end{aligned}$$

By (D.12), we know that $\phi_i^c(v) = \phi_i^c(\bar{v})$ if $v(N) = \bar{v}(N) = c$ and $\Delta_i(v) = \Delta_i(\bar{v})$. Hence, we can define $\phi_i(\Delta_i(v), c) : \Lambda^i \times \mathbb{R} \to \mathbb{R}$ as $\phi_i(\Delta_i(v), c) =: \phi_i^c(v)$. Therefore, for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$, we obtain $f_i(v, w) = \phi_i(\Delta_i(v), v(N)) + \alpha_i(w, v(N))$. This completes Claim 1.

**Claim 2**: For any $v \in \mathcal{G}$, there exists a $\delta \in [0, 1]$ and $d_i^{v(N)} \in \mathbb{R}$ such that $\phi_i(\Delta_i(v), v(N)) = \delta Sh_i(v) + d_i^{v(N)}$.

Let $w^* = (1/n, \cdots, 1/n) \in \mathcal{W}$, i.e., the equal weight. For any $c \in \mathbb{R}$ and any $v \in \mathcal{G}^c$, by Claim 1, we have

$$f_i(v, w^*) = \phi_i(\Delta_i(v), v(N)) + \alpha_i(w^*, v(N)), \tag{D.14}$$

and, by Lemma D.1, there exists $\delta \in [0,1]$ such that

$$f_i(v, w^*) = \delta Sh_i(v) + (1-\delta)\frac{1}{n}c. \tag{D.15}$$

Note that $\delta$ does not depend on $c \in \mathbb{R}$. For any $v' \in \mathcal{G}^{c,i}$, we have

$$\begin{aligned} \phi_i(\Delta_i(v'), c) + \alpha_i(w^*, c) \overset{(D.14)}{=} f_i(v', w^*) &\overset{(D.15)}{=} \delta Sh_i(v') + (1-\delta)\frac{1}{n}c \\ &= (1-\delta)\frac{1}{n}c. \end{aligned} \tag{D.16}$$

Note that player $i$ is a null player in game $v' \in \mathcal{G}^{c,i}$. Hence, for any $v', v'' \in \mathcal{G}^{c,i}$, we have $\phi_i(\Delta_i(v'), c) + \alpha_i(w^*, c) \overset{(D.16)}{=} (1-\delta)\frac{1}{n}c \overset{(D.16)}{=} \phi_i(\Delta_i(v''), c) + \alpha_i(w^*, c)$ and, so, denote $d_i^c := \phi_i(\Delta_i(v'), c) = \phi_i(\Delta_i(v''), c)$. We obtain

$$\alpha_i(w^*, c) \overset{(D.16), d_i^c}{=} (1-\delta)\frac{1}{n}c - d_i^c. \tag{D.17}$$

Therefore, for every $v \in \mathcal{G}^c$, we must have

$$\phi_i(\Delta_i(v), c) \overset{(D.14)(D.15)(D.17)}{=} \delta Sh_i(v) + d_i^c. \tag{D.18}$$

Since $c \in \mathbb{R}$ is arbitrary chosen, we obtain $\phi_i(\Delta_i(v'), v(N)) = \delta Sh_i(v) + d_i^{v(N)}$ for all $v \in \mathcal{G}$.

**Claim 3**: $\alpha_i(w, v(N)) = (1-\delta) \cdot w_i v(N) - d_i^{v(N)}$ for each $w \in \mathcal{W}$.

Consider any $w \in \mathcal{W}$ and player $k^* \in N$ such that $k^* \in \text{argmin}_{i \in N, w_i > 0} w_i$. Note that $k^*$ is well-defined because $\sum_{i \in N} w_i = 1$ and $w_i \geq 0$ for any $i \in N$. By Claim 2, for any player $i \neq k^*$ and any $c \in \mathbb{R}$,

$$f_k(cu_{\{i\}}, w) = \begin{cases} \delta c + d_k^c + \alpha_k(w, c) & \text{if} \quad k = i, \\ d_k^c + \psi_k^c(w) & \text{otherwise.} \end{cases}$$

Hence, we have

$$\sum_{k \in N} (\alpha_k(w, c) + d_k^c) \overset{(\mathrm{E}^*)}{=} (1-\delta)c. \tag{D.19}$$

Moreover, for any $i \neq k^*$, $j$ ($j \neq i,\ j \neq k^*$) and, by considering a game $cu_{\{j\}}$, we have

$$\alpha_i(w, c) + d_i^c \overset{(\mathrm{RIN}^*)}{=} \frac{w_i}{w_{k^*}}(\alpha_{k*}(w, c) + d_{k^*}^c), \tag{D.20}$$

because $i$ and $k^*$ are null players in $cu_{\{j\}}$. Therefore, for any $i \in N$, we have

$$\begin{aligned}
\alpha_i(w,c) + d_i^c - (1-\delta)w_i c &\overset{(D.20)}{=} w_i \cdot \left[\frac{1}{w_{k^*}}(\alpha_{k*}(w,c) + d_{k^*}^c) - (1-\delta)c\right] \\
&\overset{(D.19)}{=} w_i \cdot \left[\frac{1}{w_{k^*}}(\alpha_{k*}(w,c) + d_{k^*}^c) - \sum_{k\in N}(\alpha_k(w,c) + d_k^c)\right] \\
&\overset{(D.20)}{=} \frac{w_i}{w_{k^*}} \cdot \left[(\alpha_{k*}(w,c) + d_{k^*}^c) - \sum_{k\in N} w_k(\alpha_{k*}(w,c) + d_{k^*}^c)\right] \\
&\overset{\sum_k w_k = 1}{=} \frac{w_i}{w_{k^*}} \cdot \left[(\alpha_{k*}(w,c) + d_{k^*}^c) - (\alpha_{k*}(w,c) + d_{k^*}^c)\right] \\
&= 0.
\end{aligned}$$

Since $c \in \mathbb{R}$ is arbitrary chosen, we obtain $\alpha_i(w, v(N)) = (1-\delta)\cdot w_i v(N) - d_i^{v(N)}$ for all $v \in \mathcal{G}$.

**Claim 4**: For any $v \in \mathcal{G}$ and $w \in \mathcal{W}$, there exists a $\delta \in [0,1]$ such that $f_i(v,w) = \delta \cdot Sh_i(v) + (1-\delta)\cdot w_i v(N)$.

For any $v \in \mathcal{G}$ and $w \in \mathcal{W}$, we have

$$\begin{aligned}
f_i(v,w) &\overset{\text{C1}}{=} \phi_i(\Delta_i(v), v(N)) + \alpha_i(w, v(N)) \\
&\overset{\text{C2}}{=} \delta Sh_i(v) + d_i^{v(N)} + \alpha_i(w, v(N)) \\
&\overset{\text{C3}}{=} \delta Sh_i(v) + d_i^{v(N)} + (1-\delta)\cdot w_i v(N) - d_i^{v(N)} \\
&= \delta Sh_i(v) + (1-\delta)\cdot w_i v(N).
\end{aligned}$$

This completes the proof. □

## D.5 Independence of axioms for Theorem 4.3

The independence of the axioms is shown in the examples listed below.

**Example D.5.1**. Consider the following function: for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$,

$$f_i^{\mathrm{E}^*}(v,w) = 0.$$

Then, the function satisfies all axioms except (E*).

**Example D.5.2**. Consider the following function: for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$,

$$f_i^{\mathrm{M}^{-*}}(v, w) = 2Sh_i(v) - w_i v(N).$$

Then, the function satisfies all axioms except ($\mathrm{M}^{-*}$).

**Example D.5.3**. Consider the following function: for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$,

$$f_i^{\mathrm{RIN}^*}(v, w) = \delta \cdot \frac{v(N)}{|N|} + (1 - \delta) \cdot w_i v(N).$$

The function satisfies all axioms except ($\mathrm{RIN}^*$).

**Example D.5.4**. Consider the following function: for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$,

$$f_i^{\mathrm{SYM}^*}(v, w) = \delta \cdot Sh_i^z(v) + (1 - \delta) \cdot w_i v(N),$$

where $Sh_i^z(v)$ is the weighted Shapley value for a given weight $z \in R_{++}^N$. Since ($\mathrm{SYM}^*$) is defined over $\mathcal{G}$ and $\mathcal{W}$, the function satisfies all axioms except ($\mathrm{SYM}^*$).

**Example D.5.5**. Consider the following function: for any $i \in N$, $v \in \mathcal{G}$ and $w \in \mathcal{W}$,

$$f_i^{\mathrm{FEC}^*}(v, w) = w_{\min} \cdot Sh_i(v) + (1 - w_{\min}) w_i v(N),$$

where $w_{\min} = \min_{j \in N} w_j$. This function satisfies all axioms except ($\mathrm{FEC}^*$).

# Appendix E Appendix to Chapter 5

## E.1 Other robustness concepts

In this appendix, we consider two other robustness concepts.

### Strong robustness and a SMR

Even if a voting rule is robust and responsiveness of at least one individual is strictly greater than one-half, the arithmetic mean of responsiveness of all individuals can be less than one-half, which implies that a collective decision reflects minority preferences on average. To avoid this scenario, a voting rule must satisfy the following stronger requirement.

**Definition E.1.** A voting rule $\phi \in \Phi$ is *strongly robust* if, for each $p \in \Delta(\mathcal{X})$, the arithmetic mean of responsiveness is strictly greater than one-half:

$$\sum_{i \in N} p(\phi(x) = x_i)/n > 1/2 \text{ for all } p \in \Delta(\mathcal{X}). \tag{E.1}$$

Clearly, a strongly robust rule is robust. In the next proposition, we show that a voting rule is strongly robust if and only if it is robust and anonymous; that is, it is a SMR with odd $n$.

**Proposition E.1.** Suppose that $n$ is odd. Then, a voting rule is strongly robust if and only if it is a SMR. Suppose that $n$ is even. Then, no voting rule is strongly robust.

*Proof.* Note that, by (5.1), (E.1) is equivalent to

$$\sum_{i \in N} E_p[\phi(x)x_i] > 0 \text{ for all } p \in \Delta(\mathcal{X}). \tag{E.2}$$

Suppose that $n$ is odd and that $\phi$ is a SMR. Then, $\phi$ satisfies (E.2) by Lemma 5.1, so a SMR is strongly robust.

Suppose that $\phi$ is not a SMR. Then, there exist $y \in \mathcal{X}$ and $S \subsetneq N$ such that $|S| < n/2$ and $\phi(y) = y_i$ if and only if $i \in S$. Let $p \in \Delta(\mathcal{X})$ be such that $p(y) = 1$.

Then, $\sum_{i \in N} E_p[\phi(x)x_i] = |S| - |N \setminus S| < 0$, violating (E.2). Thus, a voting rule is not strongly robust unless it is a SMR.

Suppose that $n$ is even. Let $y \in \mathcal{X}$ be such that $y_i = 1$ for $i \leq n/2$ and $y_i = -1$ for $i \geq n/2 + 1$. For $p \in \Delta(\mathcal{X})$ with $p(y) = 1$ and any $\phi \in \Phi$, it holds that $\sum_{i \in N} E_p[\phi(x)x_i] = 0$, violating (E.2). Thus, no voting rule is strongly robust when $n$ is even. □

**Semi-robustness and a WMR**

We consider the following weaker version of robustness to characterize a WMR with possibly negative weights.

**Definition E.2.** A voting rule $\phi \in \Phi$ is *semi-robust* if, for each $p \in \Delta(\mathcal{X})$, responsiveness of at least one individual is not equal to one-half.

To understand the implication of semi-robustness, imagine that some individuals are more likely to have correct choices and other individuals are more likely to have wrong choices. However, if the responsiveness of every individual is equal to one-half, then it is difficult to extract information from individuals in order to arrive at a correct group decision. A semi-robust rule does not face this problem for any probability distribution.

The following proposition establishes the equivalence of a semi-robust rule and a WMR with possibly negative weights allowing no ties.

**Proposition E.2.** A voting rule is semi-robust if and only if it is a WMR such that there are no ties.

To prove Proposition E.2, we use the following theorem of alternatives, which is referred to as Gordan's theorem. This result also appears in Gale (1960, Theorems 2.9) as a corollary of Farkas' lemma.

**Lemma E.1.** Let $A$ be an $n \times m$ matrix. Exactly one of the following alternatives holds.

- There exists $\xi \in \mathbb{R}^n$ satisfying
$$\xi^\top A \gg 0.$$
- There exists $\eta \in \mathbb{R}^m$ satisfying
$$A\eta = 0, \;\; \eta > 0.$$

*Proof of Proposition E.2.* We can restate the conditions in Proposition E.2 as follows.

(a″) By Lemma 5.1, a voting rule $\phi$ is a WMR allowing no ties if and only if there exists $w = (w_i)_{i\in N} \neq 0$ such that

$$\sum_{i\in N} w_i l_{ij} = \sum_{i\in N} w_i\big(\phi(x^j)x_i^j\big) > 0$$

for each $j \in M$, or equivalently, $w^\top L \gg 0$.

(b″) By definition, a voting rule is *not* semi-robust if and only if there exists $p = (p_j)_{j\in M} > 0$ such that

$$\sum_{j\in N} l_{ij} p_j = \sum_{j:\phi(x^j)=x_i^j} p_j - \sum_{j:\phi(x^j)\neq x_i^j} p_j = 0$$

for each $i \in N$, or equivalently, $Lp = 0$.

Proposition E.2 states that exactly one of (a″) and (b″) holds, which is true by Lemma E.1. In fact, by plugging $L$, $w$, and $p$ into $A$, $\xi$, and $\eta$ in Lemma E.1, respectively, we can conclude that exactly one of (a″) and (b″) holds. □

## E.2 An imaginary asset market

We can explain why Proposition 5.1 is true in terms of arbitrage-free pricing in an imaginary asset market because we can interpret Lemma 5.4 as a corollary of the fundamental theorem of asset pricing.

Let $M$ and $N$ be the set of states and the set of assets, respectively. One unit of asset $i \in N$ yields a payoff $l_{ij}$ when state $j \in M$ is realized. Recall that $l_{ij}$ equals $+1$ if $i$'s choice agrees with the collective decision and $-1$ otherwise. The matrix $L$ is referred to as the payoff matrix. We denote by $q = (q_i)_{i\in N}$ the vector of prices of the $n$ assets.

A portfolio defined by a vector $w = (w_i)_{i\in N}$ consists of $w_i$ units of asset $i$ for each $i \in N$. It yields a payoff $\sum_{i\in N} w_i l_{ij}$ when state $j \in M$ is realized, which is summarized in $w^\top L = (\sum_{i\in N} w_i l_{ij})_{j\in M}$. The price of the portfolio is $w^\top q = \sum_{i\in N} q_i w_i$.

We say that a price vector $q$ is arbitrage-free if $w^\top L \geq 0$ implies $w^\top q \geq 0$; that is, the price of any portfolio yielding a nonnegative payoff in each state is nonnegative. We say that a price vector $q$ is determined by a nonnegative linear pricing rule if there exists a nonnegative vector $p = (p_j)_{j \in M} > 0$, which is referred to as a state price, such that $q = Lp$. The fundamental theorem of asset pricing establishes the equivalence of an arbitrage-free price and the existence of a nonnegative linear pricing rule, which is immediate from Farkas' lemma.

**Claim E.1.** A price vector $q$ is arbitrage-free if and only if it is determined by a nonnegative linear pricing rule. That is, the set of all arbitrage-free price vectors is $\{q : q = Lp,\ p > 0\}$.

The fundamental theorem of asset pricing has the following corollary, which is immediate from Lemma 5.4 (a corollary of Farkas' lemma).

**Claim E.2.** There exists a portfolio with nonnegative weights in all assets (i.e. no short selling) yielding a strictly positive payoff in each state if and only if, for any arbitrage-free price vector, the price of at least one asset is strictly positive.

The former condition is restated as $w^\top L \gg 0$ for some $w \geq 0$ and the latter condition is restated as $Lp \not\leq 0$ for all $p > 0$. Therefore, Claim E.2 implies the equivalence of a robust rule and a WMR with nonnegative weights allowing no ties.

This paper does not discuss how to find optimal weights of WMRs, whereas it is a central topic in modern portfolio theory to determine optimal weights. Thus, modern portfolio theory could be useful to find optimal weights of WMRs.

## E.3 Robustness vs. optimality

In this appendix, we demonstrate that the RTF theorem has another normative implication, optimality in terms of Paretian social preferences.

**Definition E.3.** Fix $p \in \Delta(\mathcal{X})$. A voting rule $\phi \in \Phi$ is optimal with respect to a Paretian von Neumann-Morgenstern (vNM) welfare function if there exists a linear welfare function $v : \mathrm{co}(\Phi) \to \mathbb{R}$ such that (i) $E_p[\phi'(x)x_i] \geq E_p[\phi''(x)x_i]$ for each $i \in N$ implies $v(\phi') \geq v(\phi'')$ for $\phi', \phi'' \in \mathrm{co}(\Phi)$, and (ii) $v(\phi) \geq v(\phi')$ for all $\phi' \in \mathrm{co}(\Phi)$.

Harsanyi's utilitarianism theorem (Harsanyi, 1955) states that if a linear welfare function $v : \mathrm{co}(\Phi) \to \mathbb{R}$ satisfies the condition (i), then there exists a nonnegative vector $w > 0$ such that $v(\phi) = \sum_{i \in N} w_i E[\phi(x)x_i]$ for each $\phi \in \mathrm{co}(\Phi)$ (cf. Domotor,1979; Weymark, 1993; Mandler, 2005). Thus, Proposition 5.3 can be understood as the following normative characterization of WMRs.

**Corollary E.1.** Fix $p \in \Delta(\mathcal{X})^{\circ}$. A voting rule is optimal with respect to a Paretian vNM welfare function if and only if it is a WMR with nonnegative weights.

This corollary and Propositions 5.1 and 5.2 reveal the following relationship between robustness, weak robustness, and optimality.

**Corollary E.2.** A voting rule is optimal with respect to a Paretian vNM welfare function if and only if it is weakly robust. A robust rule is optimal, but an optimal rule is not necessarily robust.

## E.4 Robustness vs. efficiency: a numerical example

In this appendix, we demonstrate the difference between the requirement of robustness and that of efficiency using a numerical example.

Assume that $n = 5$ and let $\phi$ be a SMR with veto power of individual 1, which can be represented as follows:

$$\phi(x) = \begin{cases} -x_1 & \text{if } x \in \mathcal{A}, \\ x_1 & \text{otherwise,} \end{cases}$$

where $\mathcal{A} \equiv \{x \in \mathcal{X} : x_1 = 1,\ \#\{i : x_i = 1\} \leq 2\}$. Fix $p \in \Delta(\mathcal{X})$ with

$$p(x) = \begin{cases} \alpha & \text{if } x \in \mathcal{A}, \\ \beta & \text{otherwise,} \end{cases}$$

where $0 < \alpha < 1/140$ and $\beta = (1 - 5\alpha)/27$. Then, no deterministic voting rule is better than $\phi$ under $p$. To see this, let $\phi' \in \Phi$ be as good as $\phi$ under $p$. It is enough to show that $\phi' = \phi$. We first prove that $\phi'(x) = \phi(x)$ for all $x \notin \mathcal{A}$. This is because otherwise there exists $x \notin \mathcal{A}$ such that $\phi'(x)x_1 = -1$ and thus the expected payoff of individual 1 under $\phi'$ is strictly less that that under $\phi$:

$$\begin{aligned} E_p[\phi(x)x_1] - E_p[\phi'(x)x_1] &= \sum_{x \notin \mathcal{A}} p(x)(\phi(x)x_1 - \phi'(x)x_1) + \sum_{x \in \mathcal{A}} p(x)(\phi(x)x_1 - \phi'(x)x_1) \\ &= \sum_{x \notin \mathcal{A}} \beta(1 - \phi'(x)x_1) + \sum_{x \in \mathcal{A}} \alpha(-1 - \phi'(x)x_1) \\ &\geq 2\beta - 10\alpha = 2(1 - 140\alpha)/27 > 0. \end{aligned}$$

We next prove that $\phi'(x) = \phi(x)$ for all $x \notin \mathcal{A}$. By the above property of $\phi'$, we obtain

$$\sum_{i \in N} E_p[\phi(x)x_i] - \sum_{i \in N} E_p[\phi'(x)x_i] = \alpha \sum_{x \in \mathcal{A}} \left( \sum_{i \in N} \phi(x)x_i - \sum_{i \in N} \phi'(x)x_i \right) \geq 0,$$

where the last inequality follows from $\sum_{i \in N} \phi(x)x_i = |\sum_{i \in N} x_i| \geq \sum_{i \in N} \phi'(x)x_i$ for all $x \in \mathcal{A}$. We also have $\sum_{i \in N} E_p[\phi(x)x_i] \leq \sum_{i \in N} E_p[\phi'(x)x_i]$ because $\phi'$ is as good as $\phi$ under $p$. Thus, $\sum_{i \in N} E_p[\phi(x)x_i] = \sum_{i \in N} E_p[\phi'(x)x_i]$, which implies that $\sum_{i \in N} \phi(x)x_i = \sum_{i \in N} \phi'(x)x_i$ for all $x \in \mathcal{A}$, i.e., $\phi'(x) = \phi(x)$ for all $x \in \mathcal{A}$.

Let us discuss whether $\phi$ is efficient. Although no deterministic voting rule is better than $\phi$ under $p$, the following random voting rule $\phi'$ is strictly better than $\phi$ under $p$:

$$\phi'(x) = \begin{cases} -1 + \delta & \text{if } x_1 = -1 \text{ and } x_i = 1 \text{ for all } i \neq 1, \\ -1 + \gamma & \text{if } x_1 = 1 \text{ and } \#\{i : x_i = 1\} = 2, \\ \phi(x) & \text{otherwise}, \end{cases}$$

where $\delta, \gamma \in (0, 1)$ and $\beta\delta/(4\alpha) < \gamma < \beta\delta/(2\alpha)$. In fact,

$$E_p[\phi'(x)x_1] - E_p[\phi(x)x_1] = -\beta\delta + 4\alpha\gamma > 0,$$

$$E_p[\phi'(x)x_i] - E_p[\phi(x)x_i] = \beta\delta - 2\alpha\gamma > 0 \text{ for each } i \neq 1.$$

Thus, $\phi$ does not satisfy the requirement of weakly efficiency as well as that of efficiency.

Let us discuss whether $\phi$ is robust. Although the inverse rule of $\phi$ is not better than $\phi$ under $p$, it is strictly better than $\phi$ under $p' \in \Delta(\mathcal{X})$ given by

$$p'(x) = \begin{cases} 1/16 & \text{if } x_1 = 1 \text{ and } x_i = -1 \text{ for all } i \neq 1, \\ 2/16 & \text{if } x_1 = 1 \text{ and } \#\{i : x_i = 1\} = 2, \\ 7/16 & \text{if } x_1 = -1 \text{ and } x_i = 1 \text{ for all } i \neq 1, \\ 0 & \text{otherwise}. \end{cases}$$

In fact, $E_{p'}[\phi(x)x_i] = 7/16 - 9/16 = -1/8 < 0$ for all $i$. Thus, if $\phi'$ is the inverse rule of $\phi$ then $E_{p'}[\phi'(x)x_i] = -E_{p'}[\phi(x)x_i] > E_{p'}[\phi(x)x_i]$ for all $i$, which implies that $\phi$ does not satisfy the requirement of weak robustness as well as that of robustness.

## E.5 Proof of Proposition 5.4

We first prove the first statement, which asserts that exactly one of the following holds.

(a‴) Note that, by Lemma 5.1, a voting rule $\phi$ is a WMR with nonnegative weights allowing no ties if and only if there exists $w = (w_i)_{i \in N} \geq 0$ such that, for all $\phi' \in \Phi$ and $x \in \mathcal{X}$, (5.7) is true with strict inequality holding for at least one decision profile $x$. This is true if and only if there exists $w = (w_i)_{i \in N} \geq 0$ such that

$$\sum_{i \in N} w_i E_p[\phi(x) x_i] > \sum_{i \in N} w_i E_p[\phi'(x) x_i] \text{ for all } \phi' \in \Phi \setminus \{\phi\},$$

or equivalently,

$$\sum_{i \in N} w_i (E_p[\phi(x) x_i] - E_p[\phi'(x) x_i]) > 0 \text{ for all } \phi' \in \Phi \setminus \{\phi\}.$$

(b‴) There exists a random voting rule that is as good as $\phi$ if and only if there exists $\rho \in \mathbb{R}_+^{\Phi \setminus \{\phi\}}$ such that

$$E_p[\phi(x) x_i] \leq \sum_{\phi' \in \Phi \setminus \{\phi\}} E_p[\phi'(x) x_i] \rho(\phi') / \sum_{\phi' \in \Phi \setminus \{\phi\}} \rho(\phi') \text{ for all } i \in N,$$

or equivalently,

$$\sum_{\phi' \in \Phi \setminus \{\phi\}} (E_p[\phi(x) x_i] - E_p[\phi'(x) x_i]) \rho(\phi') \leq 0 \text{ for all } i \in N.$$

By Lemma 5.4, exactly one of (a‴) and (b‴) holds. Although the same theorem of alternatives characterizes both robust rules and strictly efficient rules, its use is different from each other. In fact, as discussed in Section 5.5, robustness and strict efficiency are different conditions.

The second and third statements are implied by the well-known theorem of Wald (1950) on admissible decision functions or that of Pearce (1984) on undominated strategies. Proposition 5.3 states that a voting rule is a WMR with a weight vector $w$ if and only if (5.8) holds. Mathematically, (5.8) is equivalent to expected utility maximization, where $\Phi$ is the set of actions, $N$ is the set of states, and $w_i / \sum_j w_j$ is a probability of state $i \in N$. Therefore, we can apply the theorem of Wald (1950) on admissible decision functions or that of Pearce (1984) on undominated strategies.

In particular, Theorems 5.2.1 and 5.2.5 in Blackwell and Girshick (1954) are useful. Theorem 5.2.1 implies that a voting rule is weakly efficient if and only if there exists a weight vector $w > 0$ such that (5.8) holds. Theorem 5.2.5 implies that a voting rule is efficient if and only if there exists a weight vector $w \gg 0$ such that (5.8) holds. Therefore, this corollary holds by Proposition 5.3. □

For its completeness, we review the results of Blackwell and Girshick (1954) and show the formal proofs of second and third statements.

### E.5.1 $S$ games

Consider the following situation where a decision maker makes a statistical decision against nature. This situation is captured by a zero-sum game, which is called a $S$ game. $I_n = \{1, \cdots, n\}$ be a set of actions for player 1. Let $S \subset \mathbb{R}^n$ be an action set of player 2 and its element is denoted by $s = (s_1, \cdots s_n)$. We assume that $S$ is bounded. Player 1, who is called *nature*, chooses an action $i \in I_n$ independently and player 2, who is called *statistician*, chooses an action $s = (s_1, \cdots s_n) \in S$. Finally, payoff for player 1 is determined by $M(i, s) = s_i$ where $s_i$ is the $i$-th coordinate of $s$. The game $G = (I_n, S, M)$ is called a $S$ game.

### E.5.2 Optimal strategies

Since $S \subset \mathbb{R}^n$ and it is bounded, the set of mixed strategies $S^*$ of player 2 is a convex subset of $\mathbb{R}^n$. Similarly, any mixed strategy $\zeta$ of player 1 is described by

$$\zeta = (\zeta(1), \cdots, \zeta(n)), \quad \zeta(i) \geq 0, \quad \sum_i^n \zeta(i) = 1.$$

Thus, the set of mixed strategies $\Xi$ for player 1 is also a convex subset of $\mathbb{R}^n$. For each $\zeta \in \Xi$ and $s \in S^*$, payoff is naturally extended to $M(\zeta, s) = \zeta \cdot s = \sum_{i=1}^n \zeta(i) s_i$. Let $(\Xi, S^*, M)$ be the mixed extension of the $S$ game.

**Definition E.4.**

(1) A strategy $a \in S^*$ is *admissible* if there is no $s \in S^*$ such that $s_i \leq a_i$ for all $i$ and inequality for some $i$. Let $\mathcal{A}$ be the set of all admissible strategies.

(2) A strategy $b \in S^*$ is *Bayes point* if there is $\zeta \in \Xi$ such that

$$\zeta \cdot b = \min_{s \in S^*}(\zeta \cdot s)$$

Let $\mathcal{B}$ be the set of all Bayes points and the point $b$ is said to be Bayes against $\zeta$.

Let $\Xi_+$ be the set of $\zeta \in \Xi$ such that $\zeta(i) > 0$ for all $i$. Then, the set of all $s \in S^*$ which are Bayes against some $\zeta \in \Xi_+$ is denoted by $\mathcal{D}$. In particular, for a finite set $A \subset \Xi_+$, we denote by $\mathcal{D}(A)$ the set of Bayes against some $\zeta \in A$. Let $\bar{\mathcal{D}}$ be the closure of $\mathcal{D}$. For $s \in S^*$, define $T_s = \{t \in \mathbb{R}^n | t_i < s_i \text{ for all } i\}$.

For $S$ games, the following results are well-known.

**Theorem E.1** (Theorem 5.2.1 in Blackwell and Girshick, 1954)**.** Let $(\Xi, S^*, M)$ be the mixed extension of the $S$ game. A strategy $s^* \in S^*$ is a Bayes point if and only if $T_{s^*} \cap S^* = \emptyset$.

**Theorem E.2** (Theorem 5.2.5 in Blackwell and Girshick, 1954)**.** Let $(\Xi, S^*, M)$ be the mixed extension of $S$ game, where $S^*$ is spanned by a finite number of points. Then, there are finite number of strategies $\zeta^1, \cdots, \zeta^k \in \Xi_+$ such that $\mathcal{D} = \mathcal{A} = \bar{\mathcal{D}} = \mathcal{D}(A)$ where $A = \{\zeta^1, \cdots, \zeta^k\}$.

### E.5.3 Voting as $S$ game

We show Corollary 5.4 by considering an appropriate $S$ game. Fix $p \in \Delta(\mathcal{X})^\circ$. Consider a following correspondence.

- $I_m = N = \{1, \cdots, n\}$.
- $S = \{(-E_p[\phi(x)x_1], \cdots, -E_p[\phi(x)x_n]) | \phi \in \Phi\}$.

That is, the nature chooses an agent and the statistician chooses a voting rule. Since the number of voting rules are finite, $S$ is finite. Moreover, each $s \in S^* = \Delta(S)$ is generated by a random voting rule $\phi \in co(\Phi)$. Therefore, we can write $\Xi = \Delta(N)$ and $S^* = \{(-E_p[\phi(x)x_1], \cdots, -E_p[\phi(x)x_n]) | \phi \in co(\Phi)\}$. In this formulation, for $\zeta = (w_1/\sum_{i=1}^n w_i, \cdots, w_n/\sum_{i=1}^n w_i)$ and $s \in S$,

$$\sum_{i \in N} \left(\frac{w_i}{\sum_{i=1}^n w_i}\right)(-E_p[\phi(x)x_i]) = \zeta \cdot s,$$

which implies that

$$\max_{\phi' \in \Phi} \sum_{i \in N} w_i E_p[\phi'(x)x_i] = \min_{\phi' \in \Phi} \sum_{i \in N} w_i(-E_p[\phi'(x)x_i]) = \min_{s \in S}(\zeta \cdot s).$$

Hence, (5.8) means that $\phi$, WMR with a weight vector $w \in \mathbb{R}^N$, is a Bayes point against $\zeta$. Then, by Theorem E.1, it is equivalent to that there is no $\phi' \in co(\Phi)$ such that

$$-E_p[\phi'(x)x_i] < -E_p[\phi(x)x_i] \Leftrightarrow E_p[\phi'(x)x_i] > E_p[\phi(x)x_i]$$

for all $i$, which means that $\phi$ is weakly efficient.

When $\zeta \in \Xi_+$, by Theorem E.2, $\phi$ is Bayes point against $\zeta$ and admissible, so that there is no $\phi' \in co(\Phi)$ such that

$$-E_p[\phi'(x)x_i] \leq -E_p[\phi(x)x_i] \Leftrightarrow E_p[\phi'(x)x_i] \geq E_p[\phi(x)x_i]$$

for all $i$ and inequality for some $i$, which means that $\phi$ is efficient.

## E.6 Non-robustness of the two-thirds rule

In this appendix, we see that the two thirds rule is not robust even when $n$ is not large. Suppose that $n = 9$ and voting rule $\phi^{2/3}$ is the two-thirds rule.[1] Let $X \subset \mathcal{X}$ be the set of voting profiles such that $|\{i|x_i = 1\}| = 5$ and, for each $i \in N$, $X_i \subset X$ be the set of voting profiles such that $x \in X$ and $x_i = -1$. Note that $|X| = {}_9C_5 = 126$ and $|X_i| = {}_8C_3 = 56$.[2] Then, for any $x \in X$, $\phi^{2/3}(x) = -1$.

Let us consider the uniform distribution over $X$. Then, for any $i \in N$, $p(\phi^{2/3}(x) = x_i) = p(x \in X_i) = 56/126 < 1/2$, which implies that $\phi^{2/3}$ is not robust.

[1] The following argument can be applied for any $n > 3$ and any supermajority rule which is not equivalent to SMR.

[2] $X_i$ is the set of all voting profiles such that $|\{j \neq i|x_i = 1\}| = 5$. Therefore, all combination of such profiles is ${}_8C_3 = 56$

# References

Abe, T., and Nakada, S. (2017). "Monotonic redistribution: Reconciling performance-based allocation and weighted division," *International Game Theory Review*, 19.

Ahn, D. (2007). "Hierarchies of ambiguous beliefs," *Journal of Economic Theory*, 136, 286–301.

Alcalde, J., and Barberà, S. (1994). "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems," *Economic Theory*, 4, 417–435.

Aumann, R. J. (1962). "Utility theory without the completeness axiom," *Econometrica*, 30, 445–462.

Azrieli, Y., and Kim, S. (2014). "Pareto efficiency and weighted majority rules," *International Economic Reveiw*, 55, 1067–1088.

Azrieli, Y., and Teper, R. (2011). "Uncertainty aversion and equilibrium existence in games with incomplete information," *Games and Economic Behavior*, 73, 310–317.

Bade, S. (2008). "Housing Problems with Non-Rationalizable Behavior," Working paper.

Bade, S. (2016). "Pareto-optimal matching allocation mechanisms for boundedly rational agents," *Social Choice and Welfare*, 47, 501–510.

Béal, S., Casajus, A., Huettner, F., Rémila, E., and Solal, P. (2016). "Characterizations of weighted and equal division values," *Theory Decision*, 80, 649–667.

Bergemann, D., Brooks, B., and Morris, S. (2016). "Informationally robust optimal auction design," Working paper.

Bergemann, D., and Morris, S. (2005). “Robust mechanism design,” *Econometrica*, 73, 1771–1813.

Bergemann, D., and Schlag, K. (2011). “Robust monopoly pricing,” *Journal of Economic Theory*, 146, 2527–2543.

Bernheim, B. D., and Rangel, A. (2009). “Beyond Revealed Preferences: Choice Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 124, 51–104.

Blackwell, D., and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*, New York: John Wiley & Sons.

Bewley, T. F. (2002). “Knightian decision theory. Part I,” *Decisions in Economics and Finance*, 25, 79–110.

Blume, L. E. (1993). “The statistical mechanics of strategic interaction,” *Games and Economic Behavior*, 5, 387–424.

Bodoh-Creed, A. L. (2012). “Ambiguous beliefs and mechanism design,” *Games and Economic Behavior*, 75, 518–537.

Bose, S., Ozdenoren, E., and Pape, A. (2006). “Optimal auctions with ambiguity,” *Theoretical Economics*, 1, 411–438.

Bose, S., and Daripa, A. (2009). “A dynamic mechanism and surplus extraction under ambiguity,” *Journal of Economic Theory*, 144, 2084–2114.

Bose, S., and Renou, L. (2014). “Mechanism design with ambiguous communication devices,” *Econometrica*, 82, 1853–1872.

van den Brink, R., Funaki, Y., and Ju, Y. (2013). “Reconciling marginalism with egalitarianism: consistency, monotonicity, and implementation of egalitarian Shapley values,” *Social Choice and Welfare*, 40, 693–714.

Brighouse, H., and Fleurbaey, M. (2010). “Democracy and proportionality,” *Journal of Political Philosophy*, 18, 137–155.

Britton, E., Fisher, P., and Whitley, P. (1998). "The inflation report projections: Understanding the fan chart," *Bank of England Quarterly Bulletin*, 38, 30–37.

Carlsson, H., and Van Damme, E. (1993). "Global games and equilibrium selection," *Econometrica*, 61, 989–1018.

Carroll, G. (2015). "Robustness and linear contracts," *American Economic Review*, 105, 536–563.

Carroll, G. (2017). "Robustness and separation in multidimensional screening," *Econometrica*, 85, 453–488.

Chiesa, A., Micali, S., and Zhu, Z. A. (2015). "Knightian analysis of the vickrey mechanism," *Econometrica*, 83, 1727–1754.

Casajus, A. (2010). "Another characterization of the Owen value without the additivity axiom," *Theory and Decision*, 69, 523–536.

Casajus, A. (2011). "Differential marginality, van den Brink fairness, and the Shapley value," *Theory and Decision*, 71, 163-174.

Casajus, A. (2014). "The Shapley value without efficiency and additivity," *Mathematical Social Sciences*, 68, 1–4.

Casajus, A. (2015). "Monotonic redistribution of performance-based allocations: A case for proportional taxation," *Theoretical Economics*, 10, 887–892.

Casajus, A. (2016). "Differentially monotonic redistribution of income," *Economics Letters*, 141, 112–115.

Casajus, A., and Huettner, F. (2013). "Null players, solidarity, and the egalitarian Shapley values," *Journal of Mathematical Economics*, 49, 58–61.

Casajus, A., and Huettner, F. (2014). "Weakly monotonic solutions for cooperative games," *Journal of Economic Theory*, 154, 162–172.

Casajus, A., and Yokote, K. (2017a). "Weak differential marginality and the Shapley value," *Journal of Economic Theory*, 167, 274–284.

Casajus, A., and Yokote, K. (2017b). “Weak differential monotonicity, flat tax, and basic income,” *Economics Letters*, 151, 100–103.

Chakrabarti, S., and Gilles, R. P. (2007). “Network potentials,” *Review of Economic Design*, 11, 13–52.

Cherepanov, V., Feddersen, T., and Sandroni, A. (2013). “Rationalization,” *Theoretical Economics*, 8, 775–800.

Chun, Y. (1988). “The proportional solution for rights problems,” *Mathematical Social Sciences*, 15, 231–246.

Chun, Y. (1989). “A new axiomatization of the Shapley value,” *Games and Economic Behavior*, 1, 119–130.

Chun, Y. (1991). “On the symmetric and weighted Shapley values,” *International Journal of Game Theory*, 20, 183–190.

Chung, K. S., and Ely, J. C. (2007). “Foundations of dominant-strategy mechanisms,” *Review of Economic Studies*, 74, 447–476.

de Clippel, G. (2014). “Behavioral implementation,” *The American Economic Review*, 104, 2975–3002.

de Clippel, G. (2012). “Behavioral implementation,” Working paper.

Dasgupta, P., and Maskin, E. (2008). “On the robustness of majority rule,” *Journal of European Economic Association*, 6, 949–973.

Diamond, D. W., and Dybvig, P. H. (1983). “Bank runs, deposit insurance, and liquidity,” *The Journal of Political Economics*, 91, 401-419.

Di Tillio, A., Kos, N., and Messner, M. (2012). “The design of ambiguous mechanisms,” Working paper.

Domotor, Z. (1979). “Ordered sum and tensor product of linear utility structures,” *Theory and Decision*, 11, 375–399.

Dubins, L., and Freedman, D. (1981). “Machiavelli and the Gale-Shapley Algorithm,” *American Mathematical Monthly*, 88, 485–491.

Dybvig, P., and Ross, S. (2003). "Arbitrage, state prices and portfolio theory," in *Handbook of the Economics of Finance, Vol.1*, ed. by Constantinides, G., Harris, M., and Stulz, R. M. Amsterdam: Elsevier Publisher, 605–637.

Dybvig, P., and Ross, S. (2008). "Arbitrage," in *The New Palgrave Dictionary of Economics*, ed. by Durlauf, S., and Blume, L. London: Palgrave Macmillan, 188–197.

Einy, E., and Lehrer, E. (1989). "Regular simple games," *International Journal of Game Theory*, 18, 195–207.

Eliaz, K., and Ok, E. A. (2006). "Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences," *Games and Economic Behavior*, 56, 61–86.

Ellsberg, D. (1961). "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 75, 643–669.

Epstein, L. G. (1997). "Preference, rationalizability and equilibrium," *Journal of Economic Theory*, 73, 1–29.

Epstein, L. G., and Wang, T. (1996). "'Beliefs about beliefs' without probabilities," *Econometrica*, 64, 1343–1373.

Facchini, G., van Megen, F., Borm, P., and Tijs, S. (1997). "Congestion models and weighted Bayesian potential games," *Theory and Decision*, 42, 193–206.

Farrell, L. C., and Fudge, J. (2013). "An exploration of a quasi-stable online network: A longitudinal perspective," *Computers in Human Behavior*, 29, 681–686.

Fishburn, P. C. (1973). *The theory of social choice*. Princeton: Princeton University Press.

Fishburn, P. C., and Gehrlein, W. V. (1977). "Collective rationality versus distribution of power for binary social choice functions," *Journal of Economic Theory*, 15, 72–91.

Fleurbaey, M. (2008). "One stake one vote," Working paper.

Furusawa, T., and Konishi, H. (2007). "Free trade networks," *Journal of International Economics*, 72, 310–335.

Gale, D. (1960). *The Theory of Linear Economic Models*. Chicago: University of Chicago Press.

Gale, D., and Shapley, L.S. (1962). "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9–15.

Gilboa, I., and Marinacci, M. (2013). "Ambiguity and the Bayesian paradigm," in *Advances in Economics and Econometrics*, ed. by Acemoglu, D., Arellano, M., and Dekel, E. Cambridge: Cambridge University Press.

Gilboa, I., and Schmeidler, D. (1989). "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

Goldstein, I. (2005). "Strategic complementarities and the twin Crises," *The Economic Journal*, 18, 368–390.

Goldstein, I., and Pauzner, A. (2005). "Demand-deposit contracts and the probability of bank runs," *the Journal of Finance*, 60, 1293–1327.

Goyal, S., and Joshi, S. (2006). "Bilateralism and free trade," *International Economic Review*, 47, 749–778.

Halpern, J. Y., and Kets, W. (2015). "Ambiguous language and common priors," *Games and Economic Behavior*, 90, 171–180.

Harsanyi, J. (1955). "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility," *Journal of Political Economics*, 63, 309–321.

Hart, S., and Mas-Colell, A. (1989). "Potential, value, and consistency," *Econometrica*, 57, 589–614.

Hofbauer, J., and Sorger, G. (1999). "Perfect foresight and equilibrium selection in symmetric potential games," *Journal of Economic Theory*, 85, 1–23.

Hsie, C. S., König, M. D., and Liu, X. (2016). "Network formation with local complements and global substitutes: The case of R&D networks," Working paper.

Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley & Sons.

Iachan, F. S., and Nenov, P. T. (2015). "Information quality and crises in regime-change games," *Journal of Economic Theory*, 158, 739–768.

Iijima, R., and Kamada, Y. (2016). "Social Distance and Network Structures," *Theoretical Economics*, 12, 655–689.

Inostroza, N., and Pavan, A. (2017). "Persuasion in global games with application to stress testing," Working paper.

Jackson, M. O. (2003). "The Stability and Efficiency of Economic and Social Networks," in *Advances in Economic Design*, ed. by Sertel, M., and Koray, S. Heidelberg: Springer-Verlag, 319–361.

Jackson, M. O. (2008). *Social and economic networks, 3rd edition.* Princeton: Princeton University Press.

Jackson, M. O., and Watts, A. (2002). "The evolution of social and economic networks," *Journal of Economic Theory*, 106, 265–295.

Jackson, M. O., and Wolinsky, A. (1996). "A strategic model of social and economic networks," *Journal of Economic theory*, 71, 44–74.

Jackson, M. O., and Zenou, Y. (2014). "Games on networks," in *Handbook of game theory, Vol.4*, ed. by Young, P., and Zamir, S. Amsterdam: Elsevier Publisher, 95–163.

Joosten, R. (1996). *Dynamics, equilibria and values*, dissertation, Maastricht University.

Joosten, R. (2016). "More on linear-potential values and extending the 'Shapley family' for TU-games," Working paper.

Kalai, G., Rubinstein, A., and Spiegler, R. (2002). "Rationalizing choice functions by multiple rationales," *Econometrica*, 70, 2481–2488.

Kalai, E., and Samet, D. (1987). "On weighted Shapley values," *International Journal of Game Theory*, 16, 205–222.

Kaminsky, G. L., and Reinhart, C. M. (1999). "The twin crises: the causes of banking and balance-of-payments problems," *American Economic Review*, 89, 473–500.

Kajii, A., and Morris, S. (1997). "The robustness of equilibria to incomplete information," *Econometrica*, 65, 1283–1309.

Kajii, A., and Ui, T. (2005). "Incomplete information games with multiple priors," *Japanese Economic Review*, 157, 332–351.

Kandori, M., Mailath, G. J., and Rob, R. (1993). "Learning, mutation, and long run equilibria in games," *Econometrica*, 61, 29–56.

Kawagoe, T., and Ui, T. (2015). "Global games and ambiguous information: an experimental study," Working paper.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. New York: Houghton Mifflin.

Kojima, F., and Troyan, P. (2011). "Matching and Market Design: An Introduction to Selected Topics," *Japanese Economic Review*, 62, 82–98.

Korpela, V. (2012). "Implementation without rationality assumptions," *Theory and Decision*, 72, 189–203.

Laskar, D. (2014). "Ambiguity and perceived coordination in a global game," *Economics Letters*, 122, 317–320.

Morris, S., and Shin, H. S. (1998). "Unique equilibrium in a model of self-fulfilling currency attacks," *American Economic Review*, 88, 587–597.

Lo, K. C. (1998). "Sealed bid auctions with uncertainty averse bidders," *Economic Theory*, 12, 1–20.

Lopomo, G., Rigotti, L., and Shannon, C. (2009). "Uncertainty in mechanism design," Working paper.

Manzini, P., and Mariotti, M. (2007). "Sequentially rationalizable choice," *The American Economic Review*, 97, 1824–1839.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). "Revealed attention," *The American Economic Review*, 102, 2183–2205.

Maskin, E. (1999). "Nash equilibrium and welfare optimality," *The Review of Economic Studies*, 66, 23–38.

Matsui, A., and Matsuyama, K. (1995). "An approach to equilibrium selection," *Journal of Economic Theory*, 65, 415–434.

Mandler, M. (2005). "Harsanyi's utilitarianism via linear programming," *Economics Letters*, 88, 85–90.

May, K. (1952). "A set of independent necessary and sufficient conditions for simple majority decisions," *Econometrica*, 20, 680–684.

Monderer, D., and Shapley, L. S. (1996). "Potential games," *Games and Economic Behavior*, 14, 124–143.

Morris, S., and Shin, H.S. (2000). "Rethinking multiple equilibria in macroeconomic modeling," *NBER Macroeconomics Annual*, 15, 139–161.

Morris, S., and Shin, H. S. (2002). "Social value of public information," *The American Economic Review*, 92, 1521–1534.

Morris, S., and Shin, H.S. (2003). "Global games: theory and applications," in *Advances in Economics and Econometrics*, ed. by Dewatripont, M., Hansen, P. L., and Turnovsky, S. J. Cambridge: Cambridge University Press, 56–114.

Morris, S., and Shin, H.S. (2004). "Coordination risk and the price of debt," *European Economic Review*, 48, 133–153.

Myerson, R. B. (1991). *Game theory: analysis of conflict*. Cambridge: Harvard University Press.

Myerson, R. B. (1977). "Graphs and cooperation in games," *Mathematics of operations research*, 2, 225–229.

von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Nishimura, H., and Ok, E. A. (2016). "Utility representation of an incomplete and nontransitive preference relation," *Journal of Economic Theory*, 166, 164–185.

Nitzan, S., and Paroush, J. (1981). "The characterization of decisive weighted majority rules," *Economics Letters*, 7, 119–124.

Nowak, A., and Radzik, T. (1994). "A Solidarity Value for n-Person Transferable Utility Games," *International Journal of Game Theory*, 23, 43–48.

Nowak, A., and Radzik, T. (1995). "On axiomatizations of the weighted Shapley values," *Games and Economic Behavior*, 8, 389–405.

Obstfeld, M. (1996). "Models of currency crises with self-fulfilling features," *European Economic Review*, 40, 1037–1047.

Ok, E. A. (2002). "Utility representation of an incomplete preference relation," *Journal of Economic Theory*, 104, 429–449.

Palsule-Desai, O. D., Tirupati, D., and Chandra, P. (2013). "Stability issues in supply chain networks: Implications for coordination mechanisms," *International Journal of Production Economics*, 142, 179–193.

Pearce, D. G. (1984). "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52, 1029–1050.

Rae, D. W. (1969). "Decision-rules and individual values in constitutional choice," *American Political Science Review*, 63, 40–56.

Roth, A. (1979). "Proportional solutions to the bargaining problem," *Econometrica*, 47, 775–778.

Roth, A. (1982). "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

Roth, A. (2008). "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions," *International Journal of Game Theory*, 36, 537–569.

Roth, A., and Sotomayor, M. A. O. (1992). *Two-sided matching: A study in game-theoretic modeling and analysis*. New York: Cambridge University Press.

Salo, A., and Weber, M. (1995). "Ambiguity aversion in first-price sealed-bid auctions," *Journal of Risk and Uncertainty*, 11, 123–137.

Schmeidler, D. (1989). "Subjective probability and expected utility without additivity," *Econometrica*, 57, 571–587.

Schmitz, P. W., and Tröger, T. (2012). "The (sub-)optimality of the majority rule," *Games and Economic Behavior*, 74, 651–665.

Shapley, L. (1953a). *Additive and non-additive set functions*, dissertation, Princeton University.

Shapley, L. (1953b). "A Value for n-Person Games," in *Contributions to the Theory of Games II*, ed. by Kuhn, H. W., and Tucker, A. W. Princeton: Princeton University Press, 307–317.

Shapley, L., and Shubik, M (1969). "On the core of an economic system with externalities," *American Economic Review*, 59, 678–684.

Shapley, L., and Scarf, H. (1974). "On cores and indivisibility," *Journal of mathematical economics*, 1, 23–37.

Stauber, R. (2011). "Knightian games and robustness to ambiguity," *Journal of Economic Theory*, 146, 248-274.

Slade, M. E. (1994). "What does an oligopoly maximize?," *Journal of Industrial Economics*, 42, 45–61.

Staudigl, M. (2011). "Potential games in volatile environments," *Games and Economic Behavior*, 72, 271–287.

Straffin, P. D. (1977). "Majority rule and general decision rules," *Theory and Decision*, 8, 351–360.

Suzumura, K. (1976). "Remarks on the theory of collective choice," *Economica*, 43, 381–390.

Taylor, A., and Zwicker, W. (1992). "A characterization of weighted voting," *Proceedings of the American Mathematical Society*, 115, 1089–1094.

Taylor, M. (1969). "Proof of a theorem on majority rule," *Behavioral Science*, 14, 228–231.

Turocy, T. (2008). "Auction choice for ambiguity-averse sellers facing strategic uncertainty," *Games and Economic Behavior*, 62, 155–179.

Ui, T. (2000). "A Shapley value representation of potential games," *Games and Economic Behavior*, 31, 121–135.

Ui, T. (2001). "Robust equilibria of potential games," *Econometrica*, 69, 1373–1380.

Ui, T. (2015). "Ambiguity and risk in global games," Working paper.

Ui, T., and Yoshizawa, Y. (2015). "Characterizing social value of information," *Journal of Economic Theory*, 158, 507–535.

Yamashita, T. (2015). "Implementation in weakly undominated strategies: Optimality of second-price auction and posted-price mechanism," *Review of Economic Studies*, 82, 1223–1246.

Yokote, K. (2015). "Weak addition invariance and axiomatization of the weighted Shapley value," *International Journal of Game Theory*, 44, 275-293.

Yokote, K., and Casajus, A. (2017). "Weak differential monotonicity, flat tax, and basic income," *Economics Letters*, 151, 100–103.

Young, P. (1985). "Monotonic solutions of cooperative games," *International Journal of Game Theory*, 14, 65–72.

Young, P. (1993). "The evolution of conventions," *Econometrica*, 61, 57–84.

Wald, A. (1950). *Statistical Decision Functions.* New York: Chapman & Hall.

Weymark, J. (1993). "Harsanyi's social aggregation theorem and the weak Pareto principle," *Social Choice and Welfare*, 10, 209–221.