WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Economics

Dissertation Examination Committee:
David K. Levine, Chair
Bruce Petersen
Paulo Natenzon
Elizabeth Maggie Penn
Mariagiovanna Baccara

# Essays on Microeconomic Theory

by

Ceyhun Coban

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2015

Saint Louis, Missouri

# Acknowledgements

I would like to thank ...

# Contents

# Chapter 1

# Negative Information Sharing

## 1.1 Introduction

R&D is the engine of technological progress and growth, and it has a significant role in economics. However, there are many challenges in initiating and conducting R&D activities. One particular example is the *uncertainty*. R&D activity may lead to either a new product development or a complete failure. Moreover, it involves competition among many agents such as firms, academics, research institutions. Under the presence of uncertainty, competition may not be the best option for the development of a product. Therefore, various forms of cooperation exist in R&D markets. In this paper, we introduce a new form of cooperation through the sharing of *negative outcomes*. We consider a negative outcome as one either not worth publishing from an academic standpoint, or one not worth patenting from a corporate standpoint. We ask the following question: What are the incentives to cooperate through sharing negative outcomes?

We consider the problem of product innovation via multiple research lines in a competitive environment. R&D departments usually start working on various research lines at the same time, hoping at least one of them will lead to a successful outcome. However, most of these research lines usually end up without any meaningful results, which we denote here as a negative outcome. Some competing firms may accumulate many negative outcomes over time, however in the meantime others may decide to enter without any outcome. It is possible that some of the negative outcomes are also obtained by these newcomers. We call it *duplication* which is clearly a waste of resources. Moreover, this is an inefficiency from the society's point of view. In order to overcome this inefficiency, we propose to initiate a market for negative results where firms that own negative outcomes will be the sellers, and firms that are willing to obtain these outcomes will be the buyers. One interesting question is; what will the equilibrium price be?

There are many real-world applications that duplication has a negative impact. One particular example is trial-and-error type of R&D. Firms engaging in this type of R&D have to go through a protocol that can be described as follows; first identify all possible solutions as candidates to a particular problem. Then start with the most likely candidate and analyze whether it addresses the problem. If yes, then stop. If no, then move onto the second most likely candidate. Follow this process until finding out a solution that works successfully. Pharmaceutical research fits into this category very well. Most of the pharmaceutical companies use a trial-and-error type of R&D by identifying all potential candidate substances that may cure a particular disease and carrying out multiple research lines until they find the right one. Assuming that firms of comparable size have similar level of resources and knowledge base, firms working on the same research idea will go through a similar research process and produce mostly the same negative research outcomes over time. From an industry perspective, this means the resources are wasted through duplication of the same effort. In fact, in 2013, The Pharmaceutical Research and Manufacturers of America (PhRMA) released a statement on sharing clinical research data in order to avoid wasteful research and speed up the research process in pharmaceutical research.[1]

Another example is from basic science, namely material science. Similar to the pharmaceutical research, here firms/research institutes try different types of materials to address a problem. During the process, they produce many negative outcomes, basically materials that do not work for a solution, and these outcomes will be kept as private information even though they are valuable to other firms working on the same project. Moreover, any empirical academic research in basic science falls into this category as well. A scientist may work on a data set that does not support any meaningful relationship as solution to a particular problem. However, it is possible that the same data set had already been used and resulted in a negative outcome by another scientist. This duplication of efforts yielding the same negative outcomes is a serious problem in various research fields, and firms and society can benefit from cooperation through sharing these negative outcomes.

In this essay, I tried to investigate the conditions under which cooperation through sharing negative outcomes is possible. I built a simple model where two firms are competing to develop a new product in a R&D market. If only one of the firms obtains the successful outcome and develops the product, then it will be the only firm in production stage so that it monopolizes the product market. Simultaneous discovery is also possible so that if it happens, then firms will share the product market evenly. In order to keep the model simple, I do not specify a product market structure. However, I imposed that single discovery leads to monopoly and simultaneous discovery leads to duopoly in the product market. From the standard oligopoly theory, the industry profit under

[1] http://www.phrma.org/press-release/EFPIA-and-phrma-release-joint-principles-for-responsible-clinical-trial-data-sharing-to-benefit-patients

monopoly is greater than the industry profit under duopoly. Thus, in this model, the profit of a monopolist firm is larger than the total profits of the firms under duopoly.[2] Hence, this assumption makes single discovery more attractive and as a result it creates an opportunity cost for sharing negative outcomes due to the decrease in the likelihood of becoming the monopolist. On the other hand, the benefit of sharing negative outcomes is the increase in the likelihood of a discovery.

I obtain strikingly different results depending on the ability of detecting the validity of the shared outcomes. As a benchmark, first I investigate the case where firms are fully able to verify the outcomes. If the firms have the same number of negative outcomes (*symmetric firms*), then there is no need to have a monetary transaction, it is simply a barter. In this case, if the success probability is sufficiently low, then firms will agree to share negative outcomes. If one firm has more negative outcomes than the other (*asymmetric firms*), then the extra amount of negative outcomes have potential value. Again, if the success probability is sufficiently low, then there will be a range of prices where firms will agree to share. However, if the firms are unable to verify negative outcomes, then there is no way that firms will agree to share. Therefore, verifiability is a necessary but not a sufficient condition for sharing negative outcomes.

The model does not allow imitation and product market is shut down, thus in the end, this is an R&D race where winner takes the entire prize. In order to focus on uncertainty, both spillovers between firms and learning from failures are not allowed. With these restrictions, a static model becomes sufficient to explore the effect of uncertainty instead of a dynamic model.

Moreover, the investment decision is fixed, so the firms do not take into account how

[2]Moreover, we can interpret this situation as the prize for the single winner is larger than the total prize for two winners.

much they want to spend on R&D for any given value of the innovation. Thus, we can rule out the possible effects of investment on sharing decision. For the sake of simplicity, the model is based on one product and two firms. The underlying incentives would be the same if we increased the number of firms and the number of products, and their inclusion would only increase the amount of complexity.

Next section summarizes related literature. Section 3 presents the model and results. Section 4 includes the concluding remarks.

## 1.2 Related Literature

There are several lines of research that this paper relates to. The main contribution is to the literature on cooperation in R&D markets. Cooperation in R&D is analyzed in various forms, and the most distinguished feature of the models using these forms is how the research process is defined. Mainly, the literature is divided into two main categories; cooperation in process innovation and cooperation in product innovation. The very first form of cooperation in process innovation category is *cost sharing* models where the research process is taken by firms jointly in order to reduce the production cost of a particular good and each firm has right to use it in the production. The main goal is to share the cost of research process and avoid double spending (duplication). D'Aspremont and Jacquemin (1988) is a pioneering paper on this strand, focusing on spillovers rather than uncertainty where firms could benefit from each other's research activity. Kamien et al. (1992) also analyzes cost sharing models, however their model assumes no uncertainty and therefore their findings are not directly related to our results.

The second contribution is to the literature on *information sharing* in R&D. The literature mainly focuses on positive information sharing. In the positive information sharing models, the joint research process aims to produce a new or newer version of a particular good where each firm has the right to produce the new good. Marjit(1991) and Combs(1992) both analyze the effect of uncertainty on the incentives to cooperate by setting up simple models based on information sharing where there is no product market, spillovers, imitation and learning. However, their results contradict with each other. The cooperation occurs only when the success probability is sufficiently high in Combs(1992), not when the success probability is sufficiently low. The intuition is when the success probability is low, cooperation does not sufficiently pull the success probability to a higher level so that there is no point to try to innovate cooperatively. Even though their results are similar to this paper's, they do not consider the possibility of sharing negative outcomes. Kamien et al. (1992) focuses on both cost sharing and information sharing in research joint ventures, but still it does not consider sharing negative outcomes.

Silipo (2008) surveys the literature on cooperation in R&D, mainly focusing on the effects of uncertainty and spillovers over cooperation decision. Uncertainty is the key incentive to innovate. If the success probability of a product is low, firms may avoid the research regardless of the product's importance for the society (social welfare). Spillovers is another key issue. If the cost of product development research is high and if the final product can easily be imitated by another firm, then firms are unlikely to initiate a research project to develop it. Therefore, the survey compares several models in the literature and presents all the results in tables.[3] One important remark is that the incentives and the results in these models highly depend on how the research process is defined. Thus, they obtain strikingly different results.[4] Nevertheless, none of these models consider the possibility of negative information sharing. Another important distinction of these papers from this one is the verifiability of the outcomes. In all their

[3]See Table 1 (pg. 103) and Table 2 (pg. 105).

[4]See also Choi (1993), Amir et al. (2003), and Silipo and Weiss (2005).

models, the outcomes are verifiable, whereas in this paper I show that verifiability is critically important for cooperation.

There is only one closely related paper in the literature. Akcigit and Liu (2014) discusses the possibility of sharing dead-end research paths among competing firms. They argue that firms have a huge incentive to hide their dead-end findings so that their competitors go through the same path and waste resources on a research that leads to nowhere. However, this is clearly inefficient from the society's point of view. Another inefficiency is created by the nature of the research. Firms can never be sure about the outcome when they start working on the research. In a competitive environment, this may lead to early drop outs if an outcome is not received for a long time. Some research projects may take longer to obtain an outcome and if all firms drop out early, they may miss a potentially good outcome. By identifying these inefficiencies, this paper proposes a solution where firms are given the right incentives so that they agree to share their negative outcomes. With the right incentive scheme, wasteful duplicative research is avoided and an outcome is obtained optimally. As a result, social welfare will be increased and innovation pace will be accelerated. The important assumption for the aforementioned mechanism to work is that the outcomes should be verifiable. In this paper, we show that the mechanism will not work when this assumption is weakened or completely dropped.

## 1.3 Model

The model is constructed as follows. There are 2 risk neutral firms. There are a finite number of research lines. Only one will lead to a potentially successful outcome, the rest will not. Those that are not successful are denoted as *negative outcome*. Each firm accumulates a certain number of negative outcomes over time.[5] That means the likelihood of obtaining the successful outcome is determined by the accumulated number of negative outcomes. Let $q_i$ be the probability of obtaining successful outcome for firm $i$.[6] Clearly, $0 < q_i < 1$, for all $i$. Moreover, each firm knows the success probability of the other firm.

The game is organized as a two stage game. In the first stage, firms need to decide whether to share negative results or not. In the second stage, they pick a research line randomly[7]. At the end of the second stage, if only one firm obtains a successful R&D project, then it will be a monopolist in the production stage so that its payoff will be denoted by $\pi_m$ and the other firm's payoff would be 0. If both firms obtain the successful outcome, then they will split the duopoly profit $\pi_d$ evenly.[8] If both firms fail to obtain the successful outcome, then the game will move on by randomly undertaking a research line each period until the successful outcome is obtained. Since the number of research lines is finite, at least one firm will obtain the successful outcome eventually. Clearly, $\pi_m > \pi_d$.

The expected payoff of firm $i$ at the beginning of the game is denoted as follows;

$$\begin{aligned} V_i(q_i(0), q_j(0)) = q_i(0)(1 - q_j(0))\pi_m + q_i(0)q_j(0)\frac{\pi_d}{2} \\ + (1 - q_i(0))(1 - q_j(0))V_i(q_i(1), q_j(1)) - c \end{aligned} \tag{1.1}$$

where $q_i(0)$ and $q_j(0)$ denote the success probabilities of firms $i$ and $j$ at the beginning

[5]Another interpretation is that firms are endowed with different numbers of negative outcomes.

[6]Observe that the success probability increases when the number of negative outcomes increase.

[7]Firms are not allowed to decide on how much investment they want to make. Thus, it is better to call this situation as a fixed investment case.

[8]Here we shut down the product market and simply assigned valuations for different market structures (either monopoly or duopoly). It will be interesting to add the product market and see how it affects the incentives in R&D market.

of the game (first period), respectively and similarly $q_i(1)$ and $q_j(1)$ denote the success probabilities of firms $i$ and $j$ at the beginning of the second period, respectively.[9] The first term on the right hand side of the above equation is the expected payoff from the case where firm $i$ obtains success but the other firm not, second term is the expected payoff from both firms obtain success, the third term is the expected continuation payoff from the case where both firms fail to obtain successful outcome and the game continues until one obtains and the last term is the cost of undertaking a research line.

Before we proceed, we should remark that success probability of a firm or both firms will eventually will be 1 at some period $t$ and that means the only remaining research line is successful implying that the game will end at period $t+1$. Moreover, the success probability $q_i(t+1)$ for firm $i$ at the beginning of period $t+2$ is greater than the success probability $q_i(t)$ at the beginning of period $t+1$ since one more negative outcome is obtained at period $t+1$ and the game is moved on to the next period so that success probability is increased. Hence, we can conclude that success probability is increasing over time and eventually will be 1 at some point unless the successful outcome is obtained.

Another important remark is about the structure of the game. One may ask how the sharing takes place. In the end, this is an unusual situation and it is hard to determine the credibility of the outcome that is brought to the table. The obvious questions are; is it really an outcome that is obtained by working on the problem carefully or is it fabricated by the firm to increase the revenue? More importantly it can be used as a strategic device in order to misdirect the opponent. These questions all stem from the fact that the verifiability of negative outcomes would be difficult in the real world.[10] In

[9] More generally, $q_i(t)$ denotes the success probability of firm $i$ at the beginning of the period $t+1$.

[10] As a consequence, it is very likely that the firms will have problems in assessing the value and making a contractual agreement. In addition, a final agreement will even be more difficult if firms are

order to understand the importance of the verifiability issue, we first consider the case where the negative outcomes are perfectly verifiable. This will serve as a benchmark and helps us to make a comparison with the situation where the negative outcomes are not perfectly verifiable or unverifiable at all.

### 1.3.1 Benchmark Case: Verifiable negative outcomes

In this subsection, we assume that any outcome is perfectly verifiable. It means each firm is able to determine whether an outcome is actually negative immediately without incurring any cost. This may sound unrealistic since identification and reproduction of an outcome may take some time and money. However, we would like to see how hypothetical firms behave in this simple environment. We analyze this case in three parts; first the case where both firms have equal number of negative outcomes (*symmetry between firms*), and second, the case where one firm has no negative outcomes but the other has (*perfect asymmetry between firms*), and finally, the case where both firms hold negative outcomes but one has more than the other (*asymmetry between firms*).

#### Symmetry between firms

In this case, we assume that both firms hold the same number of negative outcomes. That means each firm is equally likely to obtain the successful outcome at the beginning of the game. Let $q_i^s(0)$ be the probability of drawing a potentially successful research line after sharing takes place for firm $i$.[11] So,

$$q_1^s(0) = q_2^s(0) = q^s(0)$$

uncertain about the success probabilities. This can be called *uncertainty on uncertainty.*

[11] In this case, sharing does not involve any monetary transaction, it is simply a barter between firms.

If firms do not agree to share, then they immediately make a draw. Let $q_i^{ns}(0)$ be the probability of drawing potentially successful research line for firm $i$ if there is no sharing. So,

$$q_1^{ns}(0) = q_2^{ns}(0) = q^{ns}(0)$$

Since $q_i^s(0) > q_i^{ns}(0)$, firms have a higher probability of choosing the potentially successful line in sharing scheme than non-sharing scheme. Then, the expected payoff of firm $i$ under the sharing scheme is denoted as follows;

$$\begin{aligned} V_i(q^s(0), q^s(0)) = q^s(0)(1 - q^s(0))\pi_m + q^s(0)q^s(0)\frac{\pi_d}{2} \\ + (1 - q^s(0))(1 - q^s(0))V_i(q^s(1), q^s(1)) - c \end{aligned} \tag{1.2}$$

Similarly, the expected payoff of firm $i$ under the non-sharing scheme is denoted as follows;

$$\begin{aligned} V_i(q^{ns}(0), q^{ns}(0)) = q^{ns}(0)(1 - q^{ns}(0))\pi_m + q^{ns}(0)q^{ns}(0)\frac{\pi_d}{2} \\ + (1 - q^{ns}(0))(1 - q^{ns}(0))V_i(q^{ns}(1), q^{ns}(1)) - c \end{aligned} \tag{1.3}$$

Note that the only difference between equation (2) and equation (3) is the success probabilities. Sharing will increase the success probability for both firms so that the only benefit is lower expected cost of obtaining success. However, the cost of sharing is the lower chance of being monopolist. Before we present our result that reflects the tendency between cost and benefit of sharing negative outcomes, we provide an example that illustrates how the success probabilities and continuation payoffs affect firms' decisions.

**Example 1.1** *There are four research lines, one is successful and the other three are not. There are two firms endowed with one negative distinct outcome each. The game is* *the same as described above. If firms decide not to share, then the success probabilities are*

$$q_1^{ns}(0) = q_2^{ns}(0) = \frac{1}{3}$$

*If they agree to share, then the new success probabilities are*

$$q_1^{s}(0) = q_2^{s}(0) = \frac{1}{2}$$

*Clearly, sharing increases the success probability. If both firms fail to obtain the successful outcome in both schemes, then the success probabilities in the next period will be*

$$q_1^{ns}(1) = q_2^{ns}(1) = \frac{1}{2}$$

*and*

$$q_1^{s}(1) = q_2^{s}(1) = 1$$

*Note that in the sharing scheme, firms reach the successful outcome at most in two periods whereas in the non-sharing scheme they reach at most in three periods. Given success probabilities, we obtain firms expected payoffs in sharing and non-sharing schemes from equations* (2) *and* (3) *as*

$$V(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4}\pi_m + \frac{1}{4}\frac{\pi_d}{2} + \frac{1}{4}V(1,1) - c \tag{1.4}$$

*and*

$$V(\frac{1}{3}, \frac{1}{3}) = \frac{2}{9}\pi_m + \frac{1}{9}\frac{\pi_d}{2} + \frac{4}{9}V(\frac{1}{2}, \frac{1}{2}) - c \tag{1.5}$$

*respectively. In equation* (4)*, the continuation payoff* $V(1,1)$ *basically tells us that if the game reaches to the next period, the success probability for both firms will be* 1 *implying that both firms will obtain the successful outcome and share* $\pi_d$ *evenly, and incur the cost* $c$*. Hence,*

$$V(1,1) = \frac{\pi_d}{2} - c$$

*By rearranging the equations* (4) *and* (5)*, we obtain the following condition that provides when sharing is possible;*

$$\frac{3}{11}(\pi_m - \pi_d) \leq c$$

From now on, we assume that the game also ends if neither firm obtains successful outcome.[12] Hence, there is no continuation payoff for the firms. Due to symmetry between firms, the success probabilities are the same in non-sharing scheme as well as sharing. Let's denote $q^{ns}$ as the success probability in non-sharing scheme and $q^s$ as the success probability in sharing scheme. Moreover, the expected payoffs will be the same, thus we modify the expected payoffs in equation (2) and (3) as follows;

$$V(q^s, q^s) = q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - c \tag{1.6}$$

and

$$V(q^{ns}, q^{ns}) = q^{ns}(1 - q^{ns})\pi_m + q^{ns} q^{ns} \frac{\pi_d}{2} - c \tag{1.7}$$

Given these expected payoffs, firms will agree to share whenever the expected payoff from sharing scheme is greater than or equal to the expected payoff from non-sharing scheme, i.e. $V(q^s, q^s) \geq V(q^{ns}, q^{ns})$. Therefore, the following result shows when sharing is possible.

**Proposition 1.1** *If* $q^s + q^{ns} \leq \frac{\pi_m}{\pi_m - (\pi_d/2)}$*, then both firms will agree to share. Moreover, if* $q^{ns} \leq \frac{\pi_d}{2\pi_m - \pi_d}$*, then firms always agree to share regardless of the value of* $q^s$*.*

**Proof.** Suppose $\frac{\pi_m}{\pi_m - (\pi_d/2)} \geq q^s + q^{ns}$. Then, by multiplying both sides with $(\pi_m - \frac{\pi_d}{2})$, we obtain

$$\pi_m \geq (q^s + q^{ns})(\pi_m - \frac{\pi_d}{2})$$

[12] By this assumption, there is uncertainty on obtaining successful outcome.

Similarly, by multiplying both sides of above inequality with $(q^s - q^{ns})$, we obtain

$$(q^s - q^{ns})\pi_m \geq (q^s - q^{ns})(q^s + q^{ns})(\pi_m - \frac{\pi_d}{2})$$

By rearranging the right hand side of above inequality, we obtain

$$(q^s - q^{ns})\pi_m \geq (q^s q^s - q^{ns} q^{ns})(\pi_m - \frac{\pi_d}{2})$$

By rearranging the terms of the above inequality as the ones with success probabilities in sharing scheme are on the left hand side and the ones with success probabilities in nonsharing scheme are on the right hand side, we obtain

$$(q^s - q^s q^s)\pi_m + q^s q^s \frac{\pi_d}{2} \geq (q^{ns} - q^{ns} q^{ns})\pi_m + q^{ns} q^{ns} \frac{\pi_d}{2}$$

Further regrouping of the similar terms of the above inequality leads to

$$q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - c \geq q^{ns}(1 - q^{ns})\pi_m + q^{ns} q^{ns} \frac{\pi_d}{2} - c$$

which is in fact the left hand side equals to $V(q^s, q^s)$ and the right hand side equals to $V(q^{ns}, q^{ns})$. Hence,

$$V(q^s, q^s) \geq V(q^{ns}, q^{ns})$$

Therefore, both firms will agree to share.

Now, suppose $q^{ns} \leq \frac{\pi_d}{2\pi_m - \pi_d}$. Then, by adding $q^s$ to both sides, we obtain

$$q^s + q^{ns} \leq q^s + \frac{\pi_d}{2\pi_m - \pi_d}$$

Since $q^s \leq 1$, then $q^s + \frac{\pi_d}{2\pi_m - \pi_d} \leq 1 + \frac{\pi_d}{2\pi_m - \pi_d}$ which leads to

$$q^s + q^{ns} \leq 1 + \frac{\pi_d}{2\pi_m - \pi_d}$$

By rearranging the terms on right hand side of the above inequality, we obtain

$$q^s + q^{ns} \leq \frac{2\pi_m}{2\pi_m - \pi_d}$$

Further rearrangement on the right hand side of above inequality leads to

$$q^s + q^{ns} \leq \frac{\pi_m}{\pi_m - (\pi_d/2)}$$

Hence, from the first part of the proof we can conclude that firms will agree to share regardless of the value of $q_s$. ■

This result reveals that if the initial success probability of firms is sufficiently low, then they always prefer to share their negative outcomes. However, if the initial success probability is high, then the new success probability from sharing must be under a certain level, otherwise the cost of losing the chance of being monopolist will exceed the gain from obtaining the successful outcome. Here the cost of undertaking a research line is irrelevant since firms are restricted to undertake a research line whether they share or not.

**Perfect asymmetry between firms**

We can think of this case as a scenario of a small start-up firm working on a research project and accumulating negative outcomes over time, and an established firm never worked on this project before but decides to work on it from now on.[13] Let's denote the former firm as firm 1 and the latter as firm 2. Since firm 2 has no negative results, only firm 1 may sell all its negative outcomes and receives a payment $p$. Let $q_i^s(0)$ be the probability of drawing a potentially successful research line after sharing takes place for firm $i$. So,

$$q_1^s = q_2^s = q^s$$

If firms do not agree to share, then they immediately make a draw. Let $q_i^{ns}$ be the probability of drawing potentially successful research line for firm $i$ if there is no sharing.

[13]There are many real-world situations similar to this scenario. For instance, the acquisitions of small startup firms by the established industry giants.

So,

$$q_1^{ns} = q_1^s$$

Since $q_1^s = q_1^{ns}$, then $q_1^s > q_2^{ns}$. Thus, firm 1 has a higher probability of choosing the potentially successful line than firm 2 implying that the expected payoff of firm 1 and 2 under the sharing scheme will be different, so they are denoted separately as follows;

$$V_1(q^s, q^s) = q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} + p - c \tag{1.8}$$

and

$$V_2(q^s, q^s) = q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - p - c \tag{1.9}$$

The only difference between equation (6) and (7) is the payment $p$ which is a transfer from firm 2 to firm 1. Similarly, the expected payoff of firm $i$ under the non-sharing scheme is denoted as follows;

$$V_i(q_i^{ns}, q_j^{ns}) = q_i^{ns}(1 - q_j^{ns})\pi_m + q_i^{ns} q_j^{ns} \frac{\pi_d}{2} - c \tag{1.10}$$

Given these expected payoffs, both firms will agree to share whenever $V_i(q^s, q^s) \geq V_i(q_i^{ns}, q_j^{ns})$ for each firm $i$. Therefore, the following result tells us when sharing is possible.

**Proposition 1.2** *If $q^s \leq \frac{\pi_m}{2\pi_m - \pi_d}$, then, for any $p \in [q^s(q^s - q_2^{ns})(\pi_m - \frac{\pi_d}{2}), (q^s - q_2^{ns})((1 - q^s)\pi_m + q^s \frac{\pi_d}{2})]$, both firms will agree to share.*

**Proof.** Suppose $\frac{\pi_m}{2\pi_m - \pi_d} \geq q^s$. Then, by multiplying both sides with $(2\pi_m - \pi_d)$, we obtain

$$\pi_m \geq (2\pi_m - \pi_d)q^s$$

By regrouping all the terms on the left hand side, we obtain

$$(1 - 2q^s)\pi_m + q^s\pi_d \geq 0$$

By multiplying both sides of above inequality with $(q^s - q_2^{ns})$, we obtain

$$(1-2q^s)(q^s - q_2^{ns})\pi_m + 2q^s(q^s - q_2^{ns})\frac{\pi_d}{2} \geq 0$$

By adding $q^s(q^s - q_2^{ns})\pi_m - q^s(q^s - q_2^{ns})\frac{\pi_d}{2}$ to both sides of above inequality, we obtain

$$(1-q^s)(q^s - q_2^{ns})\pi_m + q^s(q^s - q_2^{ns})\frac{\pi_d}{2} \geq q^s(q^s - q_2^{ns})\pi_m - q^s(q^s - q_2^{ns})\frac{\pi_d}{2}.$$

Hence, this inequality creates an interval for possible values of payment $p$. Thus, take any $p \in [q^s(q^s - q_2^{ns})(\pi_m - \frac{\pi_d}{2}), (q^s - q_2^{ns})((1-q^s)\pi_m + q^s\frac{\pi_d}{2})]$. Then,

$$p \geq q^s(q^s - q_2^{ns})\pi_m - q^s(q^s - q_2^{ns})\frac{\pi_d}{2}$$

By adding $q^s(1-q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - c$ to both sides of the above inequality, we obtain

$$q^s(1-q^s)\pi_m + q^s q^s\frac{\pi_d}{2} + p - c \geq q^s(1-q_2^{ns})\pi_m + q^s q_2^{ns}\frac{\pi_d}{2} - c$$

Since the left hand side of above inequality equals to $V_1(q^s, q^s)$ and the right hand side equals to $V_1(q_1^{ns}, q_2^{ns})$, we obtain

$$V_1(q^s, q^s) \geq V_1(q_1^{ns}, q_2^{ns})$$

implying that firm 1 will agree to share. Similarly, since

$$(1-q^s)(q^s - q_2^{ns})\pi_m + q^s(q^s - q_2^{ns})\frac{\pi_d}{2} \geq p,$$

by adding $(1-q^s)q_2^{ns}\pi_m + q^s q_2^{ns}\frac{\pi_d}{2} - p - c$ to both sides of above inequality, we obtain

$$q^s(1-q^s)\pi_m + q^s q^s\frac{\pi_d}{2} - p - c \geq (1-q^s)q_2^{ns}\pi_m + q^s q_2^{ns}\frac{\pi_d}{2} - c.$$

Since the left hand side of above inequality equals to $V_2(q^s, q^s)$ and the right hand side equals to $V_2(q_1^{ns}, q_2^{ns})$, we obtain

$$V_2(q^s, q^s) \geq V_2(q_1^{ns}, q_2^{ns})$$

implying that firm 2 will agree to share. Therefore, both firms will agree to share. ■

For the critical value $q^* = \frac{\pi_m}{2\pi_m - \pi_d}$, there exists a unique price $p^* = \frac{\pi_m}{2}\left(\frac{\pi_m}{2\pi_m - \pi_d} - q_2^{ns}\right)$ such that both firms will agree to share. However, if the success probability $q^s$ is smaller than the critical value $q^*$, then there will be a range of prices where sharing can take place at any price in that range.[14] We can interpret the result as follows: if the success probability of firm 1 is sufficiently high, then firm 1's willingness to share would be high in order to cover the opportunity cost of being monopolist, and firm 2 is unwilling to pay the price simply because the benefit of purchasing that information could not offset the cost. Hence, there will be no sharing.

### Asymmetry between firms

Now firm 2 is also endowed with negative outcomes. The setup of the game will be the same, but the sharing procedure will be modified as follows:
Let $q_i^{ns}$ denote the probability that a successful outcome is obtained by firm $i$ if there is no sharing. Assume that $q_1^{ns} > q_2^{ns}$ and this information is available to both firms. This means that firm 1 has more negative outcomes than firm 2. Since firm 1 has more negative outcomes, firm 1 receives a payment $p$ from firm 2 if sharing takes place. Let $q^s$ be the probability of drawing a success after sharing takes place. Note that this probability is not firm specific because by sharing both firms have all the available negative outcomes.

The changes in probabilities are denoted as follows: $\Delta q_i = q^s - q_i^{ns}$ denotes the change in probability of success for firm $i$ if sharing takes place. $\Delta q_1 q_2 = q^s q^s - q_1^{ns} q_2^{ns}$ denotes the change in probability if both firms obtain success. Now, the following result tells us when sharing is possible.

[14] The price can be determined based on the bargaining power of firms.

**Proposition 1.3** *Suppose* $\frac{\Delta q_1 q_2}{\Delta q_1 + \Delta q_2} \leq \frac{\pi_m}{2\pi_m - \pi_d}$. *Then, for any* $p \in [(\Delta q_1 q_2 - \Delta q_1)\pi_m - \Delta q_1 q_2 \frac{\pi_d}{2}, (\Delta q_2 - \Delta q_1 q_2)\pi_m + \Delta q_1 q_2 \frac{\pi_d}{2}]$*, both firms will agree to share.*

**Proof.** Suppose $\frac{\pi_m}{2\pi_m - \pi_d} \geq \frac{\Delta q_1 q_2}{\Delta q_1 + \Delta q_2}$. Then, by rearranging the inequality, we obtain

$$(\Delta q_1 + \Delta q_2)\pi_m \geq \Delta q_1 q_2 (2\pi_m - \pi_d)$$

By adding $\Delta q_1 q_2 \frac{\pi_d}{2} - (\Delta q_1 q_2 + \Delta q_1)\pi_m$ to both sides of above inequality, we obtain

$$(\Delta q_2 - \Delta q_1 q_2)\pi_m + \Delta q_1 q_2 \frac{\pi_d}{2} \geq (\Delta q_1 q_2 - \Delta q_1)\pi_m - \Delta q_1 q_2 \frac{\pi_d}{2}$$

implying that there is a range of values for payment $p$. Thus, take any $p \in [(\Delta q_1 q_2 - \Delta q_1)\pi_m - \Delta q_1 q_2 \frac{\pi_d}{2}, (\Delta q_2 - \Delta q_1 q_2)\pi_m + \Delta q_1 q_2 \frac{\pi_d}{2}]$. Then,

$$p \geq (\Delta q_1 q_2 - \Delta q_1)\pi_m - \Delta q_1 q_2 \frac{\pi_d}{2}$$

By replacing $\Delta q_1 q_2$ with $q^s q^s - q_1^{ns} q_2^{ns}$ and $\Delta q_1$ with $q^s - q_1^{ns}$ in the above inequality, we obtain

$$p \geq [q^s q^s - q_1^{ns} q_2^{ns} - (q^s - q_1^{ns})]\pi_m - [q^s q^s - q_1^{ns} q_2^{ns}]\frac{\pi_d}{2}$$

By rearranging the terms on the right hand side of the above inequality, we obtain

$$p \geq [q_1^{ns}(1 - q_2^{ns}) - q^s(1 - q^s)]\pi_m + [q_1^{ns} q_2^{ns} - q^s q^s]\frac{\pi_d}{2}$$

By adding $q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - c$ to both sides of the above inequality, we obtain

$$q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} + p - c \geq q_1^{ns}(1 - q_2^{ns})\pi_m + q_1^{ns} q_2^{ns} \frac{\pi_d}{2} - c$$

Since the left hand side of the above inequality equals to $V_1(q^s, q^s)$ and the right hand side equals to $V_1(q_1^{ns}, q_2^{ns})$, we obtain

$$V_1(q^s, q^s) \geq V_1(q_1^{ns}, q_2^{ns})$$

implying that firm1 will be better off by accepting the price $p$. Similarly, since

$$(\Delta q_2 - \Delta q_1 q_2)\pi_m + \Delta q_1 q_2 \frac{\pi_d}{2} \geq p,$$

by replacing $\Delta q_1 q_2$ with $q^s q^s - q_1^{ns} q_2^{ns}$ and $\Delta q_2$ with $q^s - q_2^{ns}$, we obtain

$$[(q^s - q_2^{ns}) - (q^s q^s - q_1^{ns} q_2^{ns})]\pi_m + [q^s q^s - q_1^{ns} q_2^{ns}]\frac{\pi_d}{2} \geq p$$

By rearranging the terms on the left hand side of the above inequality, we obtain

$$[q^s(1 - q^s) - (1 - q_1^{ns})q_2^{ns}]\pi_m + [q^s q^s - q_1^{ns} q_2^{ns}]\frac{\pi_d}{2} \geq p$$

By adding $(1 - q_1^{ns})q_2^{ns}\pi_m + q_1^{ns} q_2^{ns}\frac{\pi_d}{2} - p - c$ to both sides of the above inequality and rearranging them, we obtain

$$q^s(1 - q^s)\pi_m + q^s q^s \frac{\pi_d}{2} - p - c \geq (1 - q_1^{ns})q_2^{ns}\pi_m + q_1^{ns} q_2^{ns}\frac{\pi_d}{2} - c$$

Since the left hand side of the above inequality equals to $V_2(q^s, q^s)$ and the right hand side equals to $V_2(q_1^{ns}, q_2^{ns})$, we obtain

$$V_2(q^s, q^s) \geq V_2(q_1^{ns}, q_2^{ns})$$

implying that firm2 will be better off by accepting the price $p$. Therefore, both firms want to share. ■

Intuitively, this result tells us that sharing will only occur if the ratio of changes in the joint success probability to the sum of changes in individual success probabilities is below a certain threshold. Here, changes in joint success probability can be thought as a cost and changes in individual success probabilities as a benefit. Thus, if the cost/benefit ratio exceeds a certain threshold, sharing will not be beneficial to the firms.

### 1.3.2 Unverifiable negative outcomes

In this case, we are going to use the same setup in an even simpler way. There are 3 research lines. Only one will lead to a potentially successful outcome. Suppose each firm is endowed with one negative outcome and these negative outcomes are different (*symmetry between firms*).

The game is organized as follows; in the first stage, firms need to decide whether to share their negative outcomes or not. If they decide to share and the outcomes are revealed truthfully, then exchange occurs and as a result the successful outcome is obtained by both firms. However, they do not necessarily need to tell the truth. For example, one firm may declare another research line as negative outcome instead of true negative outcome in order to misdirect the opponent. If firms decide not to share, then game moves on to the second stage. In the second stage, firms pick a research line randomly. If any firm obtains the successful outcome, then the game ends. If no firm obtains the successful outcome, then they will undertake the only remaining research line which is of course the successful one. The cost of undertaking a research line is $c$. At the end of the game, if only one firm obtains the successful outcome, then it will be a monopolist in the production stage, so that its profit will be denoted by $\pi_m$. If both firms obtain the successful outcome, then they will split the duopoly profit, $\pi_d$ evenly.

Given the structure of the game, the following result shows us that sharing is impossible.

**Proposition 1.4** *It is never optimal for firms to share their negative results.*

**Proof.** If firms decide not to share their negative outcomes, then with probability 1/4 firm $i$ obtains the successful outcome but the other not so that the profit of firm $i$ is $\pi_m - c$, with probability 1/4 both firms obtain the successful outcome, thus the profit of both firms is $\pi_d/2 - c$, with probability 1/4 the other firm obtains the successful outcome but not firm $i$ so its profit is $-c$ and finally with probability 1/4 both firms could not obtain the successful outcome in first trial but in the second, thus the profits are $\pi_d/2 - 2c$. Hence, the overall expected profit for each firm is

$$\frac{\pi_m}{4} + \frac{\pi_d}{4} - \frac{5c}{4} \tag{1.11}$$

If firms decide to share their negative outcomes, then they need to decide whether to reveal it truthfully or to cheat. If both firms truthfully reveal their negative outcomes, then both will learn all the negative outcomes so that they know the remaining research line is successful. Hence, each firm's profit is

$$\frac{\pi_d}{2} - c \tag{1.12}$$

If one firm cheats and the other tells the truth, then with 1/2 probability cheating firm can achieve to misdirect the opponent so that it will be the only one that obtains successful outcome leading to the profit $\pi_m - c$ and truthful firm will get nothing so its profit is $-c$, however with 1/2 probability cheating firm reveals the same research line that the opponent has so that the opponent has a 50-50 chance of obtaining successful outcome. Then, the expected profit for the cheating firm is $1/2(\pi_m - c) + 1/2(\pi_d/2 - c)$ and the profit for the truthful firm is $\pi_d/4 - c$. Hence, the overall expected profits for firms when one cheats and the other tells the truth are

$$\frac{3\pi_m}{4} + \frac{\pi_d}{8} - c \tag{1.13}$$

and

$$\frac{\pi_d}{8} - c \tag{1.14}$$

respectively. If both firms cheat, then they will randomize between the remaining two research lines to reveal one as if it is the negative outcome. With probability 1/4 firm

$i$ reveals the research line that the opponent has and the opponent reveals the research line that neither has, then firm $i$ realizes that the revealed research line by the opponent is successful, but opponent cannot benefit from sharing so it randomizes between the remaining two research lines, so the expected profit of firm $i$ is $\frac{1}{2}(\pi_m - c) + \frac{1}{2}(\frac{\pi_d}{2} - c)$. With probability 1/4 each firm reveals the research line that the other has so that sharing does not bring any benefit. Therefore, the expected profit is the same as the case where there is no sharing, so $\pi_m/4 + \pi_d/4 - 5c/4$. With probability 1/4 firm $i$ reveals the research line that neither has, but the opponent reveals the one that firm $i$ has. Thus, only opponent benefits from the sharing. The expected profit for firm $i$ is $\pi_d/4 - c$. Finally, with probability 1/4 both firms reveal the research line that neither has. Thus, both will figure out that it is the successful one so that profits are $\pi_d/2 - c$. Hence, the overall expected profit for each firm when both cheats is

$$\frac{3\pi_m}{16} + \frac{5\pi_d}{16} - \frac{17c}{16} \tag{1.15}$$

It is easy to see that equation (13) is strictly greater than (12) and equation (15) is strictly greater than (14). Thus, cheating strictly dominates truthful revelation for each firm. Hence, if firms decide to share, they know that both cheated. However, if we compare the profits from non-sharing (11) and sharing (15), we see that (11) is strictly greater than (15) since $\pi_m > \pi_d$. Therefore, we can conclude that it is never optimal for firms to share their negative outcomes. ■

This result tells us that when firms are not able to verify the outcomes, there is no incentive to share their negative results. The reason is each firm has an incentive to misdirect the opponent so that it will be the only one who obtains successful outcome while the opponent was researching at the wrong direction. Hence, this result shows that verifiability is critically important on sharing negative outcomes.

## 1.4 Conclusion

In this paper, we show that it is possible to create incentives for firms to share their negative outcomes under some circumstances. The most important problem is the verifiability of negative outcomes. When there is no way to verify the negative outcomes, firms have no incentive to share their negative outcomes. Thus, verifiability is necessary but not sufficient for sharing negative outcomes. Once the verifiability issue is resolved, then sharing negative outcomes depends on how far firms close to obtaining successful outcome. If the gap between firms sufficiently low, then sharing will be possible. In other words, the benefit of sharing negative outcomes exceeds the cost of lowering the chance of being a monopolist.

One obvious question is why we do not see sharing negative outcomes in real-world. There is a market for patents, i.e successful outcomes, but not for negative outcomes. The reason might be that negative outcomes carry information on not only a particular research but also a complete plan or strategy of a firm. Revealing this information may yield to larger costs than expected.

The model that I present here is static in nature. Thus, it lacks of providing a dynamic interaction between firms and investigating its consequences. For example, it would be interesting to determine the timing of an agreement for sharing negative outcomes. Nevertheless, it proposes a new type of cooperation in R&D markets which hopefully implementable in near future. There are several challenges for its implementation. However, these challenges can be overcome by following certain policies. Moreover, this paper contributes to the discussion of competition vs. cooperation in R&D markets. Our findings suggest that in some circumstances, cooperation through sharing negative outcomes is beneficial for firms as well as society.

It would be interesting to investigate the relative performance of cooperation through sharing negative outcomes with other types of cooperation. Therefore, it would be easier to make policy recommendations. One interesting possible extension is adding spillovers into the model. In the presence of positive externalities, it is not straightforward to determine how incentives for cooperation would be affected.

# Bibliography

[1] Akcigit, U. and Liu, Q. (2014) ”The Role of Information in Innovation and Competition”, *mimeo, Columbia University and University of Pennsylvania.*

[2] Amir, R., Evstigneev, I., Wooders, J. (2003) ”Noncooperative versus cooperative R&D with endogenous spillover rates”, *Games and Economic Behavior* 42, 183207.

[3] Choi, J.P. (1993) ”Cooperative R&D with product market competition”, *International Journal of Industrial Organization* 11, 553571.

[4] Combs, K.L. (1992) ”Cost sharing versus multiple research projects in cooperative R&D”, *Economics Letters* 39, 353357.

[5] DAspremont, C. and Jacquemin, A. (1988) ”Cooperative and noncooperative R&D in duopoly with spillovers”, *American Economic Review* 78, 11331137.

[6] Kamien, I.M., Muller, E., Zang, I. (1992) ”Research joint ventures and R&D cartels”, *American Economic Review* 82, 12931306.

[7] Marjit, S. (1991) ”Incentives for cooperative and non-cooperative R and D in duopoly”, *Economics Letters* 37, 187191.

[8] Silipo, D.B. (2008) ”Incentives and forms of cooperation in research and development”, *Research in Economics* 62, 101-119.

[9] Silipo, D.B. and Weiss, A. (2005) ”Cooperation and competition in a duopoly R&D market”, *Research in Economics* 59, 4157.

# Chapter 2

# Random (probabilistic) assignment under generalized top-dominance condition

## 2.1 Introduction

The allocation of goods and services is a great challenge in economics. A market mechanism where the monetary transactions are allowed solves this problem well. However, without monetary transactions even a simple discrete resource allocation problem can be a real challenge. A discrete resource allocation problem consists of four components; a group of individuals, a group of distinct objects, individual preferences over these objects as strict orderings and most importantly a rule that matches each individual with one and only one object by taking all individual preferences into account. Consider the simplest discrete resource allocation problem; assigning two distinct objects to two individuals. Suppose both individuals prefer the same object to the other. In this case, it is impossible to determine who is going to receive the most desirable object according to these preferences? Without a priority structure, a fair outcome even for this simple problem is impossible. Thus, lotteries are generally used to restore fairness for these situations. In this simple case, a rule should allocate the same probability 1/2 to each individual.

A random (probabilistic) assignment rule assigns probabilities of receiving each object to individuals instead of actually matching the objects with individuals. To be able to determine which random assignment rule to use, one should look for several desirable properties that a random assignment rule must satisfy. Bogomolnaia and Moulin (2001) shows that it is impossible to find a random assignment rule that satisfies ordinal efficiency, no-envy and strategyproofness.[1] We introduce a domain restriction to overcome this impossibility result.

This impossibility result on discrete random allocation problem is interesting and important for several reasons. In real life, people face with allocation problems very frequently, one of them is cooperative housing. In the housing cooperation, individuals/members make equal contributions for the construction of houses and have same rights over the selection of houses without a priority structure in place. There are several methods (rules) used to solve the allocation problem in this kind of situations. The most common one is assigning numbers to each individual with a lottery first. Then the individual with the smallest number picks his/her top choice followed by the individual with the second smallest number to pick his/her top choice from the remaining houses and so on. This method is called random priority (also known as *random serial dictatorship*) assignment. It satisfies *ex-post efficiency* which means that once a priority order is determined, the resulting allocation is efficient. In other words, there

[1] We will describe these properties in detail later in this section.

is no other allocation that makes someone better-off without making anyone worse-off. Another method is to allocate objects first randomly by a lottery and then to apply the Gale's top trading cycle algorithm.[2] Abdulkadiroglu and Sonmez (1998) show these two methods are equivalent where they called the latter as *Core from random endowments*.

However, Bogomolnaia and Moulin (2001) introduce a new concept called *ordinal efficiency*. A random assignment is ordinally efficient if it is not stochastically dominated by another random assignment with respect to individual preferences over objects.[3] Bogomolnaia and Moulin (2001) show that ordinal efficiency implies ex-post efficiency, and random serial dictatorship (RSD) may fail to satisfy ordinal efficiency. Thus, they define a new random assignment rule, namely *probabilistic serial rule* (PS), which satisfies ordinal efficiency. Under PS, the probabilities are determined as follows; each individual starts consuming his/her favorite object first. The speed of consumption is equal for each individual. Once an object is exhausted, the amount consumed by an individual will be his/her probability of receiving that object. If an object is completely exhausted, individuals move on to the next best available object. This process goes on until all the objects are consumed. Finally, the amounts consumed by all individuals will represent the probability distribution of receiving each object (random allocation)

[2]Once each individual receives an object randomly, they point out the individual who is endowed with their most preferred object. This will create a cycle where one individual points to the next one and eventually the last individual points out the first individual. By allowing individuals to trade with each other, this cycle can be eliminated. Eventually, all objects will be traded by applying the same procedure for the remaining individuals and objects. This procedure of allocating objects is called top trading cycle algorithm. See Sharply and Scarf (1974).

[3]A random assignment stochastically dominates another random assignment if it assigns a higher probability of receiving one of the preferred objects among any number of objects to each individual than the other random assignment.

for each individual.[4] Thus, by construction, PS satisfies ordinal efficiency since it is not possible to increase the probabilities of receiving most preferred objects. The reason is each individual is already incentivized to increase the probability of receiving their most preferred objects in PS.

Hence, PS is superior to RSD in terms efficiency, while RSD is superior to PS in terms of strategyproofness which requires individuals to report their preferences truthfully. However, Bogomolnaia and Moulin (2001) state that there does not exist a random assignment rule that satisfies ordinal efficiency, strategyproofness and equal treatment of equals when the number of individuals and objects are greater than or equal to 4.[5]

This impossibility result attract our attention and we investigate if we can obtain a positive result by imposing a domain restriction. In other words, is it possible to find a random assignment rule that satisfies the desirable properties under a restricted domain? The answer is positive when we impose the top-dominance condition which is introduced by Alcalde and Barbera (1994). A domain satisfies *top dominance condition* if it is not possible to find an object that is superior to any other two objects which ordered differently in any pair of individual preferences. In other words, if top dominance condition is satisfied over a domain, then there does not exist any two ordering in this domain whose maximal (top) elements are the same. We know this condition is quite restrictive, but yet fairly reasonable when we focus on allocation problems such as cooperative housing assignment. If we go back to this problem, we see that top dominance condition is reasonable since individuals set their preferences by considering

[4] This is called simultaneous eating algorithm by Bogomolnaia and Moulin (2001).

[5] Equal treatment of equals requires the random allocation of any two individuals to be the same whenever they have the same preferences over objects. No-envy requires that no one prefers his/her random allocation to another individual's random allocation in a given random assignment. Note that equal treatment of equals is weaker than no-envy condition.

some features of houses. For example, in Japan people tends to prefer higher floors of a building to lower floors because of the earthquakes. When the maximal object is top floor, the rest of the preferences do not differ in individual preferences.

As a result, we show that under top dominance condition, PS satisfies strategyproofness and the conflict between ordinal efficiency and strategyproofness disappears. However, the domain that satisfies top dominance condition is not a maximal domain over which PS satisfies ordinal efficiency, strategyproofness and no-envy. We propose a richer domain than the one that satisfies top-dominance condition, namely generalized top-dominance condition. A domain satisfies *generalized top-dominance condition* if it is partitioned into two distinct subdomains where either one satisfies top-dominance condition while the other is empty or one satisfies top-dominance condition by adding an individual preference from the other subdomain if both are nonempty. We conjecture that a domain satisfying generalized top-dominance condition is the maximal domain under which PS is strategyproof.

The importance and significance of these results are based on its implementability. This paper is in an early stage on understanding the properties of these rules. Further research is needed to understand robustness of the implementation of these rules.

The paper proceeds as follows. Section 2 summarizes the related literature. Section 3 presents the model. Section 4 states our results. Section 5 makes some concluding remarks.

## 2.2 Related Literature

This paper contributes to the literature on discrete resource allocation problems, mainly focusing on random (probabilistic) assignments. Hylland and Zeckhauser (1979) is the first paper which considers the random allocation problem by using von Neumann-

Morgenstern utilities. Later, Zhou (1990) provides an impossibility result in a similar setting where individuals provide von Neumann-Morgenstern utilities over objects and show that there is no random assignment rule that satisfies ex-ante efficiency, strategyproofness and equal treatment of equals. However, the difficulty in obtaining cardinal preferences makes comparison of interpersonal utility complicated. Thus, Bogomolnaia and Moulin (2001) use ordinal preferences and define a random assignment rule, namely probabilistic serial rule. They show that there is no random assignment rule that satisfies ordinal efficiency, strategyproofness and equal treatment of equals when there are at least four objects.

One strand of the literature moved on to the direction of searching the axiomatic characterization of probabilistic serial rule. Kesten, Kurino, and Unver (2011) provide two characterizations of probabilistic serial rule; i) PS is the only random assignment rule that satisfies *non-wastefulness* and *ordinal fairness*, ii) PS is the only random assignment rule that satisfies ordinal efficiency, *ordinal envy-freeness*, and *upper invariance*. Hashimoto and Hirata (2011) provide a similar characterization of probabilistic serial rule in a setting where null object always exist. Bogomolnaia and Heo (2012) introduce a new weaker axiom called *bounded invariance* and provide a new characterization as probabilistic serial rule is the only random assignment rule that satisfies ordinal efficiency, strategyproofness and bounded invariance. Unlike these papers, our focus is on the strategyproofness condition and we try to determine the domains where probabilistic serial rule is strategyproof.

Kojima and Manea (2010) show that in sufficiently large markets, probabilistic serial rule is strategyproof. That means any gain from misrepresenting preferences disappear if the number of objects is sufficiently large. This partially solves the problem that we are interested, yet impossibility remains for small number of objects.

This paper is also related to the literature on domain restriction in resource allocation problems. Alcalde and Barbera (1994) apply a domain restriction to matching problems in order to obtain strategy-proof mechanisms. They introduce top-dominance condition and show that under this condition, there exist strategy-proof and stable mechanisms for matching problems, like marriage and college admissions.

A domain restriction to overcome an impossibility result is not always significant and interesting unlike a maximal domain result. A maximal domain tells how far one can expand the domain without falling into impossibility. Kojima (2007) presents two maximal domain restrictions on two-sided matching markets such that manipulations via capacities and pre-arranged matches are prevented. This paper follows a similar path for a different impossibility result.

## 2.3 The Model

We consider a finite set of individuals $N$ confronting a finite set of objects $A$ with the same cardinality, i.e., $|N| = |A|$. Let $L(A)$ be the set of complete, transitive and antisymmetric binary relations over $A$. Each individual $i \in N$ has *strict preferences* over the objects and denoted by $P_i$ where $P_i \in L(A)$.[6] Let $R_i$ be the weak counterpart of $P_i$ and $I_i$ be the indifferent part of $P_i$[7]. We write *preference profile* as a n-tuple $P = (P_1, P_2, ..., P_n) \in L(A)^N$.

An *assignment* is a one-to-one function $\phi$ from the set of individuals $N$ to the set of objects $A$. Since the number of individuals equals to the number of objects, $\phi$ is a bijection. In words, everyone obtains one and only one object. Given this requirement,

[6]In other words, each individual is able to rank all objects in a strict manner.

[7]As usual, for any distinct objects $x, y \in A$, $xP_iy$ denotes "$x$ is strictly preferred to $y$", $xR_iy$ denotes "$x$ is at least as good as $y$", and $xI_iy$ denotes "$x$ is indifferent to $y$".

there are many ways of assigning individuals to objects. Let $\Phi$ denotes the set of all assignments. When the number of individuals and objects are $n$, the total number of all possible assignments is $n!$, i.e., $|\Phi| = n!$. The reason is there are $n$ possible objects for the first individual, and after he/she obtains one, there are $n-1$ possibility for the second individual, and decreasing one by one there will be only one object left for the last individual. When we multiply all these possibilities, we obtain the total number of assignments which is $n.(n-1)....1 = n!$. If the number of individuals and objects increases, then the set of assignments increases much more rapidly.

There is a matrix representation for each assignment. Let $T : \Phi \longrightarrow \mathscr{M}_{n\times n}$ be a transformation defined as follows; for any given assignment $\phi \in \Phi$, let $T(\phi) = (t_{ij}^{\phi})$ be a $n \times n$ matrix where row index $i$ denotes individuals and column index $j$ denotes the objects. Clearly, the entries of matrix $T$ only take values either 0 or 1. So, $t_{ij}^{\phi} = 1$ if the individual $i$ receives the object $j$ under assignment $\phi$, and $t_{ij}^{\phi} = 0$ otherwise. Most importantly, each row and each column of $T(\phi)$ contains exactly one nonzero entry.[8] The following example will demonstrate the matrix representation of an assignment.

**Example 2.1** *Let $N = \{1, 2, 3\}$ and $A = \{a, b, c\}$. Take an assignment $\phi$ where object $a$ is assigned to individual 2, object $b$ is assigned to individual 3, and object $c$ is assigned to individual 1. Then, the matrix representation of assignment $\phi$ is the following;*

$$T(\phi) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

[8] In mathematics, these type matrices are called *permutation matrices* where each row and column contains exactly one entry with 1 and others all 0s. We can think of each assignment as a permutation of a given assignment. Thus, the matrix representation of each assignment will be the permutation matrix of the initial assignment's matrix representation.

*Here, rows denote the individuals and columns denote the objects. By looking at the matrix, it is easy to see that individual 1 (first row) receives object c (third column), individual 2 (second row) receives object a (first column), and individual 3 (third row) receives object b (second column). Moreover, each row and each column contains exactly one entry with 1.*

An assignment is a deterministic way of assigning objects to individuals. However, there are many circumstances under which it is not certain who is going to receive what. In that case, we need to incorporate uncertainty to our framework. If the assignment is chosen randomly among all assignments, then trying to determine the chances of getting an object makes sense. This will lead us to define random assignments.[9] A random assignment is a probability distribution over all assignments. Basically, randomness is obtained by assigning weights to all deterministic assignments where the weights denote the likelihood of choosing that particular assignment. Since there are many ways of choosing those weights, there are many random assignments. Let $\Delta(\Phi)$ denotes the set of all random assignments. So, more formally, $\sigma \in \Delta(\Phi)$ is a *random assignment* if

$$\sigma = \sum_{k=1}^{n!} a_k \phi^k \text{ where } \sum_{k=1}^{n!} a_k = 1,\ a_k \geq 0,\ \phi^k \in \Phi,\ \forall k.$$

Hence, weights $a_k$ need to be nonnegative and add up to 1 in order to interpret them as probabilities. Moreover, a weight needs to be assigned to all deterministic assignments. Similar to deterministic assignments, a random assignment $\sigma \in \Delta(\Phi)$ also can be represented by a $n \times n$ matrix $T(\sigma) = (t_{ij}^{\sigma})_{i\in N, j\in A}$ such that

$$t_{ij}^{\sigma} = \sum_{k=1}^{n!} a_k t_{ij}^{\phi^k} \text{ where } \sum_{k=1}^{n!} a_k = 1,\ a_k \geq 0,\ \phi^k \in \Phi,\ \forall k.$$

Here, there are several remarks in order. First of all, for all individual $i \in N$, $T_i(\sigma)$ denotes the $i^{th}$ row of $T(\sigma)$, and it will be the random allocation of individual $i$. By

[9] In the literature, it is also called probabilistic assignment.

*random allocation*, we mean that a probability distribution over all objects in $A$. Second of all, each row and each column of $T(\sigma)$ adds up to 1, and it is denoted as weighted sum of $T(\phi^k)$[10];

$$T(\sigma) = \sum_{k=1}^{n!} a_k T(\phi^k)$$

The following example will demonstrate the matrix representation of a random assignment and show how it is written as a convex combination of deterministic assignments.

**Example 2.2** *Let $N = \{1,2,3\}$ and $A = \{a,b,c\}$. Take a random assignment $\sigma$ where object $a$ is assigned to individual 1 with probability 1/3, to individual 2 with probability 1/2 and to individual 3 with probability 1/6; object $b$ is assigned to individual 1 with probability 1/3, to individual 2 with probability 1/6 and to individual 3 with probability 1/2; object $c$ is assigned to individual 1 with probability 1/3, to individual 2 with probability 1/3 and to individual 3 with probability 1/3. Then, the matrix representation of this random assignment $\sigma$ is the following;*

$$T(\sigma) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/6 & 1/3 \\ 1/6 & 1/2 & 1/3 \end{bmatrix}.$$

*Here again, rows denote the individuals and columns denote the objects. By looking at the matrix $T(\sigma)$, it is easy to see that individual 1 is equally likely will obtain one of the three objects, hence its random allocation is $T_1(\sigma) = (1/3, 1/3, 1/3)$ (first row). Similarly, individual 2's random allocation is $T_2(\sigma) = (1/2, 1/6, 1/3)$ (second row) and*

[10]In mathematics, these type of matrices are called *doubly stochastic (bistochastic) matrices* where all the entries (they are all nonnegative) in each row and column add up to 1. Every doubly stochastic matrix can be written as a convex combination of permutation matrices due to well-known Birkhoff-von Neumann theorem.

*individual 3's random allocation is* $T_3(\sigma) = (1/6, 1/2, 1/3)$ *(third row). Moreover, each row and each column entries add up to 1. However, it is not straightforward to determine the weights of deterministic assignments that leads to the random assignment* $\sigma$*. There are 6 deterministic assignments that are stated with their matrix representations below;*

$$T(\phi^1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, T(\phi^2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, T(\phi^3) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$T(\phi^4) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, T(\phi^5) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, T(\phi^6) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

*Then, the weights are* $a_1 = 1/6$*,* $a_2 = 1/6$*,* $a_3 = 1/6$*,* $a_4 = 1/6$*,* $a_5 = 1/3$ *and* $a_6 = 0$*. It is easy to check that* $\sum_{k=1}^{n!} a_k T(\phi^k)$ *is equal to* $T(\sigma)$*.*

Up to this point, we defined deterministic assignments and random assignments and analyzed their relationship and representations. However, we have not specified how a random assignment is obtained. The only information needed to determine random assignments is individual preferences. Given any individual preferences over objects, we need to systematically obtain random assignments. The following definition shows us the formal way of carrying out this process.

A *random assignment rule* $f$ is a mapping from the set of all preference profiles $L(A)^N$ to the set of all random assignments $\Delta(\Phi)$. So, for any preference profile $P \in L(A)^N$, $f(P)$ is the random assignment that represents random allocations over objects with respect to individual preferences in $P$. A trivial random assignment rule is *constant rule* which assigns equal probability to receive each object for each individual no matter what their preferences are. Say there are $n$ individuals and $n$ objects. For any preference profile $P \in L(A)^N$ over these $n$ objects, constant rule $f^{CR}(P)$ assigns the following random allocations to the individuals;

$$T(f^{CR}(P)) = \begin{bmatrix} 1/n & ... & 1/n \\ ... & ... & ... \\ 1/n & ... & 1/n \end{bmatrix}$$

where the random allocation of each individual $i$ is $T_i(f^{CR}(P)) = (1/n, ..., 1/n)$. Constant rule is immune to information since it does not take into account the individual preferences in order to determine the random assignment. However, there are some random assignment rules that use information efficiently and as a result preferred widely. One such rule takes individual preferences first, and then determines a priority ordering randomly to decide who is going to make a choice first, second and so on. By knowing all the individual preferences, it is easy for each individual to determine the probabilities of receiving each object. Intuitively, if the preferences are similar, then the probability of receiving the better objects will be lower. If the preferences are diverse, then it is very likely that many individual will receive their favorite objects.

More formally, let $\tau : \{1, 2, ..., n\} \longrightarrow N$ be a bijection that leads to an ordering of individuals where $\tau(1)$ is the first, $\tau(2)$ is the second and so on. Let $\Sigma$ denote the set of such orderings. For any $\tau \in \Sigma$ and for any preference profile $P \in L(A)^N$, *priority assignment* $PA(P, \tau)$ is the corresponding assignment where $\tau(1)$ receives his top choice according to his preferences in $P$, $\tau(2)$ receives his top choice among the remaining objects and so on. The following definition states a random assignment rule in which the priority assignment is used and determined randomly.

**Definition 2.1** *A random assignment rule* $f : L(A)^N \longrightarrow \Delta(\Phi)$ *is called random serial* *dictatorship*[11] *(RSD) if*

$$T(f(P)) = \frac{1}{n!} \sum_{\tau \in \Sigma} T(PA(P, \tau)), \forall P \in L(A)^N.$$

It is better to denote random serial dictatorship rule as $f^{RSD}$. Basically, each priority assignment may occur equally likely in RSD rule and the probabilities of receiving any object may change depending on the individual preferences. The following example shows how random serial dictatorship rule works.

**Example 2.3** *Let $N = \{1, 2, 3\}$ and $A = \{a, b, c\}$. Take a preference profile $P \in L(A)^N$ where individual 1's preferences are $aP_1bP_1c$, individual 2's preferences are $bP_2aP_2c$ and finally individual 3's preferences are $bP_3cP_3a$. So the preference profile $P$ is the following;*

| $P_1$ | $P_2$ | $P_3$ |
|---|---|---|
| $a$ | $b$ | $b$ |
| $b$ | $a$ | $c$ |
| $c$ | $c$ | $a$ |

*There are 6 priority orderings. The first priority ordering is*

$$(\tau^1(1) = 1, \tau^1(2) = 2, \tau^1(3) = 3)$$

*where first priority is given to individual 1, second priority to individual 2 and third priority to individual 3. In short, we denote it $\tau^1 = (1, 2, 3)$. Then, the other orderings are $\tau^2 = (1, 3, 2)$, $\tau^3 = (2, 1, 3)$, $\tau^4 = (2, 3, 1)$, $\tau^5 = (3, 1, 2)$, and $\tau^6 = (3, 2, 1)$. According to these priority ordering $\tau^k$'s, priority assignment $PA(P, \tau)$ will generate 6 deterministic*

[11] The name indicates that once a priority is determined, the first individual can be seen as a dictator since he can obtain whatever object he wants. Moreover, this a serial dictatorship in the sense that after first individual second individual can be seen as a dictator, then the third individual and so on. In the literature, random serial dictatorship rule is also known as random priority assignment rule.

*assignments. Their matrix representations are*

$$T(PA(P,\tau^1)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, T(PA(P,\tau^2)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, T(PA(P,\tau^3)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$T(PA(P,\tau^4)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, T(PA(P,\tau^5)) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, T(PA(P,\tau^6)) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

*Since the priority ordering is determined randomly, any of these assignments has an equal chance of being realized which is 1/6. Then the random assignment rule $f^{RSD}$ will take the convex combination of these matrices by multiplying each of them with 1/6 and generate the following matrix representation of random assignment $f^{RSD}(P)$;*

$$T(f^{RSD}(P)) = \begin{bmatrix} 5/6 & 0 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

The first desirable property on random assignment rules is the efficiency. Given individual preferences, can a random assignment rule generate efficient random assignments? Here the problem is how to define efficiency concept over random assignments. Standard efficiency definition requires that an allocation is efficient if there is no other allocation that makes someone better-off without making anyone worse-off. We could define *ex-ante efficiency* if we were able to know the *cardinal preferences* of individuals. By calculating expected utilities, we could compare random allocations for each individual. However, we are only informed about ordinal preferences. Thus, instead of ex-ante efficiency, we can only use ex-post efficiency. For any preference profile $P \in L(A)^N$, a random assignment $f(P) \in \Delta(\Phi)$ is *ex-post efficient* if it can be written as a convex combination of priority assignments;

$$T(f(P)) = \sum_{\tau \in \Sigma} a_\tau T(PA(P,\tau)) \text{ where } \sum_{\tau \in \Sigma} a_\tau = 1,\ a_\tau \geq 0,\ \forall \tau \in \Sigma.$$

Given individual preferences, the random assignment rule will generate a random assignment and each realization of this random assignment has to be efficient in the sense that no one could be better-off without making anyone worse-off. Since the priority assignment will generate an efficient outcome where everybody chooses successively the best option from the available objects, then there is no way to make someone better-off without making anyone worse-off. Thus, a random assignment rule $f$ is ex-post efficient if for any preference profile $P \in L(A)^N$, random assignment $f(P) \in \Delta(\Phi)$ is ex-post efficient. By definition, random assignment rule $f^{RSD}$ is ex-post efficient.

Ex-post efficiency does not tell anything about the relationship of two random allocations. However, individuals can in fact make comparisons of random allocations without using expected utility theory. For each individual, we define *stochastic dominance (sd) relation* over random allocations with respect to two different random assignment as follows; for any given preference profile $P \in L(A)^N$ and given two distinct random assignments $\sigma, \sigma' \in \Delta(\Phi)$,

$$T_i(\sigma) sd(P_i) T_i(\sigma') \Longleftrightarrow \{\sum_{k=1}^{j} t_{ik}^{\sigma} \geq \sum_{k=1}^{j} t_{ik}^{\sigma'}, \forall j = 1, ..., n\}, \forall i \in N$$

where $k$ denotes top $k$ preferred objects according to individual $i$'s preferences $P_i$. In words, if a random allocation assigns higher probabilities to receiving one of the most preferred objects than another random allocation, then the former is preferred to the latter allocation. Thus, for any given preference profile $P \in L(A)^N$, a random assignment $\sigma \in \Delta(\Phi)$ *stochastically dominates* another random assignment $\sigma' \in \Delta(\Phi)$ if

$$T_i(\sigma) sd(P_i) T_i(\sigma'), \forall i \in N$$

Given the definition of stochastic domination relation, it is easy to define a new efficiency concept. A random assignment $\sigma \in \Delta(\Phi)$ is *ordinally efficient* if it is not stochastically dominated by any other random assignment. Similarly, a random assignment rule $f$ is ordinally efficient if for any preference profile $P \in L(A)^N$, random assignment $f(P)$ is ordinally efficient.

Unfortunately, random serial dictatorship rule is not ordinally efficient. The following example is taken from Bogomolnaia and Moulin (2001) that shows how RSD fails to satisfy ordinal efficiency.[12]

**Example 2.4** *Let $N = \{1, 2, 3, 4\}$ be the set of individuals and $A = \{a, b, c, d\}$ be the set of objects. Take the preference profile $P$ as*

| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|
| $a$ | $a$ | $b$ | $b$ |
| $b$ | $b$ | $a$ | $a$ |
| $c$ | $c$ | $d$ | $d$ |
| $d$ | $d$ | $c$ | $c$ |

*The matrix representation of random assignment $f^{RSD}(P)$ as follows;*

$$T(f^{RSD}(P)) = \begin{bmatrix} 5/12 & 1/12 & 5/12 & 1/12 \\ 5/12 & 1/12 & 5/12 & 1/12 \\ 1/12 & 5/12 & 1/12 & 5/12 \\ 1/12 & 5/12 & 1/12 & 5/12 \end{bmatrix}.$$

[12]When the number of objects are 2 and 3, ex-post efficiency coincides with ordinal efficiency.

*Now, consider the following matrix representation of the random assignment $\sigma$;*

$$T(\sigma) = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}.$$

*Here, every individual will prefer the random allocations from $\sigma$ to the random allocations from $f^{RSD}(P)$. Thus, $\sigma$ stochastically dominates $f^{RSD}(P)$ implying that RSD is not ordinally efficient.*

In the seminal work of Bogomolnaia and Moulin (2001), a new random assignment rule is introduced. This new rule follows an algorithm, called *simultaneous eating algorithm*, which described as follows; each agent starts consuming his/her most preferred object (suppose as if the object is divisible) with equal speed $\omega_i(t)$. Once an object is consumed, they moved to their second preferred object if it is not consumed at all. This process will continue until all the objects are consumed.

**Definition 2.2** *A random assignment rule $f : L(A)^N \longrightarrow \Delta(\Phi)$ is called probabilistic serial rule (PS) if for any preference profile $P \in L(A)^N$,*

$$\omega_i(t) = 1, \forall i \in N, \forall t \in [0, 1].$$

The following example shows how PS works.

**Example 2.5** *Take the preference profile $P$ in Example 3;*

| $P_1$ | $P_2$ | $P_3$ |
|---|---|---|
| $a$ | $b$ | $b$ |
| $b$ | $a$ | $c$ |
| $c$ | $c$ | $a$ |

*According to eating algorithm, individual 1 starts consuming object a and others start consuming their favorite object b. At time $t = 1/2$, object b is completely exhausted, but half of object a is remaining (the probability shares from b is $p_{1b} = 0$, $p_{2b} = 1/2$ and $p_{3b} = 1/2$). Starting from this point, individual 2 consumes a and individual 3 consumes c. At time t=3/4, object a is completely exhausted (observe that it only takes 1/4 time to finish the remaining of a). So, the probability shares from object a is $p_{1a} = 1/2 + 1/4 = 3/4$, $p_{2a} = 1/4$ and $p_{3a} = 0$. At this point, 1/4 of c is consumed. Since it is the only object left, all individuals consume it from this point on. Finally at time $t = 1$, c is completely exhausted since it takes only 1/4 time to finish the 3/4 of c. So, the probability shares from object c is $p_{1c} = 1/4$, $p_{2c} = 1/4$ and $p_{3c} = 1/4 + 1/4 = 1/2$. Hence, the matrix representation of resulting random assignment from probabilistic serial rule is*

$$T(f^{PS}(P)) = \begin{bmatrix} 3/4 & 0 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

*Observe that outcome of PS and RSD are different.*

There are two other desirable properties that a random assignment rule should satisfy. The first one is about fairness. A random assignment rule should assign random allocations in a way that no one envies somebody else's random allocation. More formally, a random assignment rule $f : L(A)^N \longrightarrow \Delta(\Phi)$ is *envy free (no-envy)* if for any preference profile $P \in L(A)^N$,

$$\forall i, j \in N, T_i(f(P)) sd(P_i) T_j(f(P)).$$

Each individual's random allocation stochastically dominates any other individual's random allocation for any given preference profile. There is also a weaker version of this condition. A random assignment rule $f : L(A)^N \longrightarrow \Delta(\Phi)$ satisfies *equal treatment of* *equals* if for any preference profile $P \in L(A)^N$,

$$\forall i, j \in N, P_i = P_j \Longrightarrow T_i(f(P)) = T_j(f(P)).$$

This condition requires that a random assignment rule has to assign the same random allocation to the individuals whenever their individual preferences are the same in any given preference profile.

The last desirable property is about the strategic aspect of a random assignment rule. It may be possible for some individual to report their preferences untruthfully. In order to prevent this type of situations, a random assignment should satisfy the following condition. A random assignment rule $f : L(A)^N \longrightarrow \Delta(\Phi)$ is *strategyproof* if

$$\forall i \in N, \forall P, P' \in L(A)^N, T_i(f(P)) sd(P_i) T_i(f(P_{-i}, P'_i)).$$

This condition requires that no individual can increase the likelihood of receiving his/her favorite objects by misrepresenting his/her preferences. In the next section, we will analyze the consequences of these properties when they imposed on a random assignment rule.

## 2.4 The Results

Without any domain restriction, it is easy to see that random serial dictatorship (RSD) and probabilistic serial rule (PS) coincides when the number of individuals and objects are 2. Hence, we can conclude that random assignment rule PS is strategyproof. Thus, from now on we only consider cases where the number of individuals is greater than equal to 3, i.e., $|N| \geq 3$. Moreover, when the domain is extremely restricted, i.e $|D| = 2$, random assignment rule PS is again strategyproof no matter what the number of individuals and objects are.

The intuition behind this result is the following; since there are at least two individuals with the same preferences, it is impossible to gain probability increase on favorite objects by reporting the other preference available. There is not much variety in individual preferences to create a possibility of gaming the rule.

**Proposition 2.1** *Let $|N| = |A| = n$ where $n \geq 3$. Take any $D \subseteq L(A)$ with $|D| = 2$. Then, random assignment rule PS is strategyproof over $D^N$.*

Thus, from now on we consider only domains that contain at least three orderings. However, if there is enough diversification in the domain, we already showed that three ordering is sufficient, then random assignment rule PS may fail to satisfy strategyproofness. Here is an example that shows this claim.

**Example 2.6** *Let $N = \{1, 2, 3\}$ and $A = \{a, b, c\}$. Take the domain*

$$D = \left\{ \begin{array}{ccccc} a & & b & & b \\ b & , & a & , & c \\ c & & c & & a \end{array} \right\}$$

*Even though this domain is restricted (there are only three orderings out of possible six orderings), it is possible to misrepresent the preferences and increase the probability of getting better objects. To see this, take the preference profile $P$ as*

$$\begin{array}{ccc} \underline{P_1} & \underline{P_2} & \underline{P_3} \\ a & b & b \\ b & a & c \\ c & c & a \end{array}$$

*So, individual 1's preferences are $aP_1bP_1c$, individual 2's preferences are $bP_2aP_2c$ and finally individual 3's preferences are $bP_3cP_3a$. Then the matrix representation of random* *assignment rule $f^{PS}$ for this preference profile $P$ is*

$$T(f^{PS}(P)) = \begin{bmatrix} 3/4 & 0 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

*Hence, the random allocation for individual 1 is $T_1(f^{PS}(P)) = (3/4, 0, 1/4)$. However, individual 1 can obtain a better random allocation by misrepresenting his preferences. If he misrepresents his preferences as $bP'_1aP'_1c$, then the new preference profile $P'$ would be*

| $P'_1$ | $P_2$ | $P_3$ |
|---|---|---|
| $b$ | $b$ | $b$ |
| $a$ | $a$ | $c$ |
| $c$ | $c$ | $a$ |

*and the associated matrix representation of random assignment rule $f^{PS}$ outcome would be*

$$T(f^{PS}(P')) = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 0 & 1/3 & 2/3 \end{bmatrix}.$$

*Hence, the new random allocation for individual 1 is $T_1(f^{PS}(P')) = (1/2, 1/3, 1/6)$. The probability of getting object $a$ or $b$ for individual 1 in $P$ is $2/3$; however, in the new preference profile $P'$ it is $5/6$. Therefore, random assignment $f^{PS}(P)$ cannot stochastically dominates random assignment $f^{PS}(P')$ implying that random assignment rule $f^{PS}$ is not strategyproof over $D^N$.*

At this point, a natural question arises; under what domain restriction, the random assignment rule PS is strategyproof? Alcalde and Barbera (1994) proposed a domain restriction, namely top-dominance, in two-sided matching markets which leads to the existence of strategyproof and stable matching rules. Here is the definition of top-dominance condition;

**Definition 2.3** *A domain $D \subseteq L(A)$ with $|D| \geq 3$ satisfies top-dominance condition if $\forall P, P' \in D$, $\forall x, y \in A$ such that $xPy$ and $yP'x$, then there is no $z \in A$ such that $zPx$ and $zP'y$.*

In words, it is impossible to find two orderings which agree on the top object. Thus, every ordering must have different object on top in a top-dominated domain. This domain is quite restricted yet random assignment rule PS is strategyproof.

The reason why PS is strategyproof under top-dominance condition follows from the fact that top-dominance rule does not allow two different individual preferences with the same top object. Hence, this narrows down the domain in a way that it makes impossible to gain any probability increase over favorite objects by misreporting preferences.

The proof shows case-by-case the impossibility of any gain from misrepresenting preferences. Given any preference profile, either all individual have the same preferences or two has the same, one different or all different. If they are all different, then each individual obtains their top object for sure, so there is no need to misreport preferences. If they are all the same, then in order to have a gain from misrepresenting the individual has to forgone consuming his/her favorite object to increase the probability of receiving second best object. However, the other two agents start consuming the second best object (it is their second best too) immediately after they finish the best object so that there will not be sufficient probability increase. The same logic applies to the last case, yet it requires more technical work.

**Proposition 2.2** *Let $N$ be the set of individuals and $A$ be the set of objects with $|N| = |A| = n \geq 3$. Let domain $D \subseteq L(A)$ with $|D| \geq 3$ satisfies top-dominance condition.*

*Then, random assignment rule $f^{PS} : D^N \longrightarrow \Delta(\Phi)$ is strategyproof.*

**Proof.** Let $N$ be the set of individuals and $A$ be the set of objects with $|N| = |A| = n$. Take any domain $D \subseteq L(A)$ with $|D| \geq 3$ which satisfies top-dominance condition. Take any preference profile $P \in D^N$. We want to show that for all individual $i \in N$ and for all individual preferences $P_i' \in D$ with $P_i' \neq P_i$,

$$T_i(f^{PS}(P)) \; sd(P_i) \; T_i(f^{PS}(P_i', P_{-i})).$$

Take any individual $i \in N$, and label it as $i = 1$. Then label his individual preferences as

$$a_1 P_1 a_2 P_1 a_3.$$

Take any individual preference $P_i' \in D$. Denote the random allocations of individual 1 under the random assignment rule $f^{PS}$ as $T_1(f^{PS}(P)) = (p_{1a_1}, p_{1a_2}, p_{1a_3})$ and $T_1(f^{PS}(P_1', P_{-1})) = (p'_{1a_1}, p'_{1a_2}, p'_{1a_3})$ when the preference profiles are $P$ and $P' = (P_1', P_{-1})$, respectively. Let $argmax(P_i)$ is the maximal (most preferred) object at individual preference $P_i$.

Case1: Suppose $argmax(P_i) \neq argmax(P_j)$, for all distinct $i, j \in \{1, 2, 3\}$. Then, $p_{1a_1} = 1$ since everyone would consume its maximal object completely, implying that no other random allocation stochastically dominates the random allocation $T_1(f^{PS}(P)) = (1, 0, 0)$, so we are done.

Case2: Suppose $argmax(P_i) = a_1$, for all individual $i \in \{1, 2, 3\}$. So, all individual preferences are the same since top-dominance condition does not allow any two different individual preferences with the same maximal object. Thus, for any other individual preferences $P_1' \in D$, $argmax(P_1') \neq a_1$. Suppose there exist a individual preference $P_1'$ in domain $D$ such that $argmax(P_1') = a_2$. Otherwise, with an individual preference where any other object is maximal, it is never optimal to create an increase in the probability of obtaining a better object other than $a_1$ while decreasing the probability of receiving $a_1$.

Thus, with individual 1's preferences $P_1'$, the probability of receiving $a_1$ is $p'_{1a_1} = 0$ and the probability of receiving $a_2$ is

$$p'_{1a_2} = \frac{1}{n-1} + \frac{n-2}{n(n-1)} = \frac{2}{n}$$

where the first term is the share that individual 1 obtained from $a_2$ when all other individuals were consuming $a_1$ (they exhausted $a_1$ at time $\frac{1}{n-1}$) and the second term is the share that individual 1 obtained from $a_2$ when all individuals were consuming the leftover amount $\frac{n-2}{n-1}$ of $a_2$. On the other hand, in preference profile $P$, all individual preferences are the same. As a result, each individual has equal share from each object. Hence, the probability of receiving any object is $p_{1a_1} = ... = p_{1a_n} = \frac{1}{n}$. Then,

$$p_{1a_1} = \frac{1}{n} > 0 = p'_{1a_1}$$

and

$$p_{1a_1} + p_{1a_2} = \frac{2}{n} \geq \frac{2}{n} = p'_{1a_1} + p'_{1a_2}$$

implying that there is no gain from misrepresenting preferences for individual 1 in terms of an increase in the probability of receiving second best object. Similarly, it is straightforward to see that if we continue to compare probabilities of receiving third best, fourth best and so on, we will obtain that there will be no gain from misrepresenting preferences. Hence,

$$\textstyle\sum_{s=1}^{k} p_{1a_s} \geq \sum_{s=1}^{k} p'_{1a_s},\ \forall k \in \{3, ..., n\}$$

implying that no other random allocation stochastically dominates the random allocation $T_1(f^{PS}(P))$.

Case3: Suppose case1 and case2 do not hold. Then,

Claim1: $p_{1a_1} > p'_{1a_1}$.

Proof: Suppose not. So, $p_{1a_1} \leq p'_{1a_1}$. Let $t(a_1)$ be the time at which $a_1$ exhausted, and $n(a_1,t)$ is the number of individuals who eat object $a_1$ at time $t$. At preference profile $P$, individual 1 is eating $a_1$ during the whole interval $[0, t(a_1))$, so the probability of receiving object $a_1$ is $p_{1a_1} = t(a_1)$. At preference profile $(P'_1, P_{-1})$, individual 1 is eating object $a_1$ on a subset of interval $[0, t'(a_1))$. Hence, $t(a_1) \leq t'(a_1)$ since $p_{1a_1} \leq p'_{1a_1}$.

Since it is shown by Bogomolnaia and Moulin(2001) that $N(a_1,t) \subseteq N'(a_1,t)$ for $t \in [0, t(a_1))$, and combining with

$$\int_0^{t(a_1)} n(a_1,t)dt = \int_0^{t'(a_1)} n'(a_1,t)dt = 1,$$

and $n(a_1,t), n'(a_1,t)$ are nondecreasing in $t$, then $t(a_1) = t'(a_1)$. Hence, $p_{1a_1} = p'_{1a_1}$. However, this implies that individual 1 is eating $a_1$ during the whole interval $[0, t'(a_1))$ at preference profile $(P'_i, P_{-i})$, establishing that $argmax(P'_1) = a_1$. Since $P'_i \neq P_i$, then there exists $i, j \neq 1$ such that $a_i P_1 a_j$ and $a_j P'_1 a_i$. Hence this leads to a contradiction due to the violation of top-dominance condition. Therefore, $p_{1a_1} > p'_{1a_1}$.

Claim2: $p_{1a_1} + p_{1a_2} \geq p'_{1a_1} + p'_{1a_2}$.

Proof: Suppose not. Then,

$$p_{1a_1} + p_{1a_2} < p'_{1a_1} + p'_{1a_2}.$$

Since $p_{1a_1} > p'_{1a_1}$ from claim1, then $d = p_{1a_1} - p'_{1a_1} > 0$ and $p_{1a_2} + d < p'_{1a_2}$. Then, following the same steps from claim1, we obtain

$$p_{1a_1} + p_{1a_2} \geq p'_{1a_1} + p'_{1a_1}.$$

Similarly, if we continue to compare probabilities of receiving third best, fourth best and so on, we will obtain that there will be no gain from misrepresenting preferences.

Hence,

$$\sum_{s=1}^{k} p_{1a_s} \geq \sum_{s=1}^{k} p'_{1a_s}, \forall k \in \{3, ..., n\}.$$

Therefore, random allocation of individual 1 under $f^{PS}(P)$ stochastically dominates any other random allocation under $f^{PS}(P'_1, P_{-1})$, i.e.,

$$T_1(f^{PS}(P))\ sd(P_1)\ T_1(f^{PS}(P'_1, P_{-1})),$$

implying that PS is strategyproof. ■

Then this positive result leads us to overcome the impossibility theorem stated in Bogomolnaia and Moulin(2001).

**Corollary 2.1** *Let $D \subseteq L(A)^N$ satisfies top-dominance condition. Then, random assignment rule PS is ordinally efficient, strategyproof and satisfies no-envy.*

However, we observe that a domain that satisfies top-dominance condition is not maximal meaning that it is possible to find a larger domain that contains a domain which satisfies top-dominance condition and PS will be strategyproof over that larger domain. The following example illustrates this fact.

**Example 2.7** *Let $N = \{1, 2, 3\}$ and $A = \{a, b, c\}$. Take the domain*

$$D = \left\{ \begin{array}{ccccc} a & & b & & c \\ b & , & c & , & b \\ c & & a & & a \end{array} \right\}$$

*This domain satisfies top-dominance condition, hence PS is strategyproof over $D^N$. If we add another ordering, it fails to satisfy top-dominance condition. Take the domain*

$$D' = \left\{ \begin{array}{ccccccc} a & & b & & c & & a \\ b & , & c & , & b & , & c \\ c & & a & & a & & b \end{array} \right\}$$

*So, $D \subset D'$. Moreover, PS is strategyproof over ${D'}^N$.*

This example shows us that top-dominance condition is not sufficient for strategyproofness. Thus, we extend this result by showing that it is possible to find a domain richer than the domain that satisfies top-dominance condition. The solution we propose is the generalization of top-dominance condition, and definition is given as follows;

**Definition 2.4** *A domain $D \subseteq L(A)^N$ with $|D| \geq 3$ satisfies generalized top-dominance condition if it is partitioned into two distinct subdomains $D_1$ and $D_2$ with satisfying one of the following cases;*

- *if $D_1 = \emptyset$, then $D_2$ satisfies top-dominance condition.*
- *if $D_1 \neq \emptyset$ and $D_2 \neq \emptyset$, then $\forall P_i \in D_1$, $D_2 \cup \{P_i\}$ satisfies top-dominance condition.*

We conjecture that a domain satisfying generalized top-dominance condition will be the maximal domain over which random assignment rule PS is strategyproof.

**Conjecture 2.1** *The maximal domain that random assignment rule PS is ordinally efficient, strategyproof and satisfies no-envy is the domain $D \subseteq L(A)^N$ that satisfies generalized top-dominance condition.*

## 2.5 Conclusion

In this paper, we deal with the problem of discrete resource allocation where there is equal number of individuals and objects, and the task is to assign each individual one and only one object. Monetary transfers are not allowed and there is no priority structure among individuals. In order to restore fairness, it is better to use lotteries, yet individual preferences over objects need not to be the same.[13] Thus, by taking into ac-

[13]A lottery works well when all the individuals have the same preferences over objects.

count individual preferences, one has to define assignment rules that assign these objects to individuals in a systematic ways. They are called random (probabilistic) assignment rules. A random assignment rule should satisfy certain desirable properties. First of all, it should allocate probabilities of receiving objects efficiently such that there is not any other random assignment that makes someone better-off without making anyone worse-off (ordinal efficiency). Second of all, it should not create any envy between individuals such that no one should prefer any other random allocation (no-envy). Finally, individuals should not try to gain advantage by misreporting their preferences so that every one reveals his/her preferences truthfully (strategyproofness). However, Bogomolnaia and Moulin (2001) states that there is no random assignment rule that satisfies all these three properties. We provide a domain restriction to overcome the impossibility result stated above. In fact, under a domain which satisfies top-dominance condition rather than full domain, probabilistic serial rule will be strategyproof so that it will overcome the impossibility.[14]

There are some weaknesses in implementation of probabilistic serial rule. It is quite complicated to tell individuals how their preferences are used and PS generates an outcome. On the other hand, random serial dictatorship is easy to implement and as a result it is widely used in real world. Another weakness is on the top-dominance condition. Alcalde and Barbera (1994) provides many example where top-dominance is used or can be used, however, they are vague and does not reflect the importance of the condition. Thus, we tried to generalize the top-dominance condition in a way that it may be the maximal domain where PS is strategyproof.

Kojima and Manea (2010) show that in a sufficiently large market PS is approximately strategyproof. That means when the number of individuals and objects are large,

[14]PS already satisfies ordinal efficiency and no-envy under full domain.

trying to gain some advantage from misrepresenting preferences will be redundant. It would be interesting to know if a group of individuals coordinate their preferences and then misrepresent them accordingly. In that case, the domain restriction that we introduced here may prevent that kind of behavior. In this paper, we did not allow indifferences among objects in individual preferences. It will be interesting to analyze consequences of allowing indifferences in individual preferences. We leave this as a future work.

# Bibliography


[1] Abdulkadiroglu, A. and Sonmez, T. (1998) "Random serial dictatorship and the core from random endowments in house allocation problems", *Econometrica* 66: 689-701.

[2] Alcalde, J. and Barbera, S. (1994) "Top dominance and the possibility of strategy-proof stable solutions to matching problems", *Economic Theory* 4: 417-435.

[3] Bogomolnaia, A. and Heo, E. J. (2012) "Probabilistic assignment of objects: Characterizing the serial rule", *Journal of Economic Theory*, 147: 20722082.

[4] Bogomolnaia, A. and Moulin, H. (2001) "A new solution to the random assignment problem", *Journal of Economic Theory* 100: 295-328.

[5] Hashimoto, T. and Hirata, D. (2011) "Characterizations of the probabilistic serial mechanism", Unpublished paper.

[6] Hylland, A. and Zeckhauser, R (1979) "The efficient allocation of individuals to positions", *Journal of Political Economy*, 87: 293-314.

[7] Kesten, O., Kurino, M., and Unver, M. U. (2011) "Fair and efficient assignment via the probabilistic serial mechanism", Unpublished paper.

[8] Kojima, F. (2007) "When can manipulations be avoided in two-sided matching markets?-Maximal domain results", *The B.E. Journal of Theoretical Economics* Vol. 7 Iss. 1(contributions), Art. 32.

[9] Kojima, F. and Manea, M. (2010) "Incentives in the probabilistic serial mechanism", *Journal of Economic Theory*, 145: 106-123.

[10] Shapley, L. S. and Scarf, H. E. (1974) "On cores and indivisibility", *Journal of Mathematical Economics*, 1: 23-37.

[11] Zhou, L. (1990) "On a Conjecture by Gale about One-Sided Matching Problems", *Journal of Economic Theory* 52 (1): 123-135.

# Chapter 3

# Social Choice without the Pareto Principle under Weak Independence[1]

## 3.1 Introduction

In a society, some decisions are taken collectively as a group rather than individually. In collective decision making, the obvious starting point is how to aggregate a group of individuals' preferences over a set of alternatives/objects/issues to decide what the group will choose to do. If every single individual's preference is important in the group decision making process, then there has to be a systematic way of making a choice/preference for the society. An *aggregation rule* provides a systematic way by taking individual preferences as inputs and produces a social preference/choice as the output. Then, the important question is; would it be always possible to find an aggregation rule that

[1]This is the extended version of Coban, C. and Sanver, M.R. (2014) "Social Choice without the Pareto Principle under Weak Independence", *Social Choice and Welfare* 43: 953-961.

satisfies certain desirable properties?

Arrow tried to answer this question in his seminal work *Social Choice and Individual Values*[2] by imposing the following desirable properties on the aggregation rule; an aggregation rule should produce a social outcome for all logically possible combinations of individual preferences (*full domain*), an aggregation rule should produce a ranking between any two alternatives and rank all the alternatives at least as a weak order (*collective rationality*), an aggregation rule should rank one alternative on top of another whenever all individuals strictly prefer the former to the latter (*Pareto condition*), an aggregation rule should rank two alternatives only by considering their individual rankings (*independence of irrelevant alternatives*) and finally an aggregation rule should not give the decision making power to a single individual (*nondictatoriality*). As a conclusion, he shows that it is impossible to find an aggregation rule that satisfies all of these properties when there are at least three alternatives.

This impossibility result attracted many scholars to search for aggregation rules that might satisfy somewhat weaker versions of those desired properties. The independence of irrelevant alternatives (IIA) condition is blamed on leading to the impossibility and many studies have been presented various weakenings of IIA in the context of Arrovian framework. This paper takes one of those weakenings, namely *weak IIA*[3], and characterizes the class of social welfare functions[4] (SWFs) that satisfies this condition.

Preference aggregation problem is very important in many social situations, such as *voting*. In a voting situation, there is a certain number of candidates and voters. In an

[2] I am referring to the second edition. In the first edition, he uses slightly different conditions instead of some conditions mentioned above.

[3] Weak IIA requires that the social ordering of any two alternatives cannot be reversed in any two distinct preference profiles where the individual orderings of that pair of alternatives are the same in both profiles.

[4] A social welfare function is an aggregation rule which satisfies collective rationality.

ideal situation, voters have full information about all candidates and also know how to evaluate them. From an individual decision making perspective, the obvious question is how voters rank candidates. It is also interesting to look at what makes a voter rank one candidate over another and to understand how voters form their preferences. In this paper, our focus is on collective decision making instead of individual decision making. In collective decision making, there are many voting methods that are being used. As an example in plurality rule, which is the one of the most common methods to select political leaders in US and around the world, each voter votes for one candidate, and the candidate with the highest number of votes wins. However, this rule is problematic that it only takes into account each individual's top candidate rather than taking full individual preferences over candidates. A better social outcome might be achieved by taking each individual's preferences over all candidates. Thus, it is important to compare aggregation rules in terms of whether they satisfy certain properties or not.

Another interesting group decision making problem is the committee decisions. Suppose there is a committee that has a certain number of individuals, and it needs to decide whom to hire as a university president among three candidates. One way to make a social decision is to tell committee members to rank each pair of candidates and to choose the winner by making pairwise comparisons. Using an aggregation rule like this may end up choosing candidate 1 over candidate 2 and candidate 2 over candidate 3 and finally candidate 3 over candidate 1 which violates the collective rationality requirement.[5] In response to this, one might suggest to ask committee members assign numbers to candidates as 2 for the most preferred, 1 for the second most and 0 for the least preferred. Then, the aggregation rule sums up the numbers for each candidate and ranks the candidate with the highest number first, second highest second so on.[6]

[5] In the literature, this is known as Condorcet paradox.

[6] In the literature, this is known as Borda rule.

However, this aggregation rule violates IIA in the sense that the social ranking of any two alternatives depends on the individual rankings of all other alternatives. These two examples show us that there is a tension between collective rationality (transitivity of social preferences) and IIA.[7]

In this paper, we consider the preference aggregation problem in a society with a certain number of individuals and at least three alternatives over which each individual has strict preferences.[8] We are mainly interested in social welfare functions that satisfy weak IIA condition. We try to overcome the conflict between collective rationality (transitivity of social preferences) and IIA by using weak IIA. Given a social welfare function that satisfies weak IIA, we show that it can be written as an IIA aggregation rule whose cycles are converted to indifference classes. Conversely, for any IIA aggregation rule, we apply a transformation by removing cycles and the end result will be a weakly IIA SWF. Hence, we obtain a fairly large class of SWFs. In fact, by weakening Pareto condition to weak Pareto[9], we obtain a similar result.

One might argue the significance of our result by claiming that weak IIA SWFs will produce many social indifferences among alternatives. However, compared to dictatoriality, it provides an improvement. We did not have an attempt to overcome Arrow's Impossibility by simultaneously weakening both IIA and Pareto condition. Our main interest is the role of IIA on the impossibility result and the consequences of weakening IIA condition. The end result was valuable in the sense that we obtain a full characterization

[7]An aggregation rule based on pairwise comparison will satisfy IIA but fail to satisfy collective rationality. On the other hand, an aggregation rule based on scoring will satisfy collective rationality but not IIA.

[8]We did not allow individuals to be indifferent among alternatives in order to keep the analysis simple.

[9]Weak Pareto requires that a social ranking of any two alternatives cannot be reversed whenever all individuals strictly prefer one over the other.

of a class of SWFs. An important concern over this application is the implementation of weak IIA SWFs. It is possible to use such aggregation rules if a milder independence is needed in a situation where the social decision over any two alternatives should not be affected by the presence of other alternatives. However, it is important to note that weak IIA does not provide much independence from other alternatives since it takes the presence of other alternatives into account.

The paper is organized as follows; section 2 summarizes the related literature. Section 3 presents the basic notions. Section 4 states our results. Section 5 makes some concluding remarks.

## 3.2 Related Literature

This paper contributes to the literature on preference aggregation in Arrovian framework. To overcome Arrow's impossibility result, weaker versions of Arrow's original conditions are introduced. Our paper is closely related to weakening of one of these conditions, namely IIA. However, the Arrovian impossibility is remarkably robust against weakenings of IIA.[10] For example, letting $k$ stand for the number of alternatives that the society confronts, Blau (1971) proposes the concept of *$m$-ary independence* for any integer between 2 and $k$. A SWF is *$m$-ary independent* if the social ranking of any set of alternatives with cardinality $m$ depends only on individuals' preferences over that set. Clearly, when $m = 2$, $m$-ary independence coincides with IIA. Moreover, every SWF trivially satisfies $m$-ary independence when $m = k$. It is also straightforward to see that $m$-ary independence implies $n$-ary independence when $m < n$. Nevertheless, Blau

[10] In fact, it is robust against weakenings of other conditions as well: Wilson (1972) shows that the Arrovian impossibility essentially prevails when the Pareto condition is not used. Ozdemir and Sanver (2007) identify several restricted domains which exhibit the Arrovian impossibility.

(1971) shows that $m$-ary independence implies $n$-ary independence when $n < m < k$ as well. Thus, weakening IIA by imposing independence over sets with cardinality more than two does not allow to escape the Arrovian impossibility, unless independence is imposed over the whole set of alternatives - a condition which is satisfied by the definition of a SWF.

Campbell and Kelly (2000a, 2007) further weaken $m$-ary independence by requiring that the social preference over a pair of alternatives depends only on individuals' preferences over some proper subset of the set of available alternatives. This condition, which they call *independence of some alternatives* (ISA) is considerably weak. As a result, non-dictatorial SWF that satisfies Pareto principle and ISA -such as the "gateau rules" identified by Campbell and Kelly (2000a)- do exist. On the other hand, "gateau rules" fail neutrality and as Campbell and Kelly (2007) later show, an extremely weaker version of ISA disallows both anonymity and neutrality within the Arrovian framework.

Denicolo (1998) identifies a condition called *relational independent decisiveness* (RID). He shows that although IIA implies RID, the Arrovian impossibility prevails when IIA is replaced by RID. These papers introduce and analyze other types of weakenings of IIA rather than the weakening that we applied and therefore they are indirectly related with our paper.

Campbell (1976) proposes a weakening of IIA which requires that the social decision between a pair of alternatives cannot be reversed at two distinct preference profiles that admit the same individual preferences over that pair. We refer to this condition as *weak IIA*.[11] Baigent (1987) shows that every Paretian and weak IIA SWF must be dictatorial in a sense which is close to the Arrovian meaning of the concept - hence a version of

[11] See Campbell (1976) for a discussion of the computational advantages of weak IIA. Note that when social indifference is not allowed, IIA and weak IIA are equivalent.

the Arrovian impossibility.[12] None of these two papers tried to characterize the class of weak IIA SWFs.

In brief, the literature which explores the effects of weakening IIA on the Arrovian impossibility presents results of a negative nature. We revisit the literature in order to contribute by a positive result. Under the weakening proposed by Campbell (1976) and Baigent (1987), we characterize the class of weak IIA SWFs and show that this is a fairly large class which is not restricted to SWFs where the decision power is concentrated on one given individual. In fact, this class contains SWFs that are both anonymous and neutral.

This paper also contributes to the literature on simultaneous weakenings of two or more original Arrow conditions. In fact, the positive result that we obtained prevails when a weak version of the Pareto condition is imposed. Nevertheless, a recent paper Cato (2014) shows that possibility is limited in the sense that if Pareto condition is replaced by *(strict) non-imposition* instead of weak Pareto, the result will be dictatoriality.

Moreover, our result is a contrast to the results of Wilson (1972) and Barberà (2003) who show that the Pareto condition has little impact on the Arrovian impossibility which is essentially a tension between IIA and the range restriction imposed over SWFs. Moreover, we establish that there is no tension between weak IIA and the transitivity of the social outcome.

[12] Baigent (1987) claims this impossibility in an environment with at least three alternatives. Nevertheless, Campbell and Kelly (2000b) show the existence of Paretian and weak IIA SWF when there are precisely three alternatives. They also show that the impossibility announced by Baigent (1987) prevails when there are at least four alternatives and even under restricted domains.

## 3.3 Basic Notions

We consider a finite set of individuals $N$ with $\#N \geq 2$, confronting a finite set of alternatives $A$ with $\#A \geq 3$. An *aggregation rule* is a mapping $\upsilon : L(A)^N \to C(A)$ where $L(A)$ is the set of complete, transitive and antisymmetric binary relations over $A$ while $C(A)$ is the set of complete binary relations over $A$. We interpret $P_i \in L(A)$ as the *preference* of individual $i \in N$ over alternatives in $A$.[13] That means each individual is able to compare all available alternatives in $A$ and order them from best to worst according to his/her preferences. This ordering is strict in the sense that there are no indifferences among any alternatives. We write $P = (P_1, ..., P_{\#N}) \in L(A)^N$ for a *preference profile* and $\upsilon(P) \in C(A)$ reflects the social preference obtained by the aggregation of $P$ through $\upsilon$. For all possible combination of individual preferences, aggregation rule $\upsilon$ generates a complete social preference. That means any two alternative is socially comparable for any given preference profile $P$ under $\upsilon$. Note that $\upsilon(P)$ need not be transitive implying that an alternative, say $x$, is socially preferred to another alternative, say $y$, and $y$ is also socially preferred to another distinct alternative, say $z$, but $z$ is socially preferred to $x$ which is a cycle.[14] Moreover, as $\upsilon(P)$ need not be antisymmetric, we write $\upsilon^*(P)$ for its strict counterpart.[15]

Before we proceed any further, we can summarize the preference aggregation problem as a tuple $(N, A, (P_i)_{i \in N}, \upsilon)$ where there is a finite set of individuals $N$, a finite set of alternatives $A$, individual preferences $P_i$ as strict orderings over all alternatives in $A$

[13]As usual, for any distinct $x, y \in A$, we interpret $xP_iy$ as $x$ being preferred to $y$ in view of $i$.

[14]More formally, if there are more than three alternatives, say $x_1, ..., x_k$, where $x_1$ is preferred to $x_2$ and $x_2$ is preferred to $x_3$ and so on till $x_{k-1}$ is preferred to $x_k$ and in addition $x_k$ is preferred to $x_1$, then this creates a cycle of preferences not an ordering. Imposing transitivity on preferences prevents this type of problems.

[15]So for any distinct $x, y \in A$, we have $x\ \upsilon^*(P)\ y$ whenever $x\ \upsilon(P)\ y$ and not $y\ \upsilon(P)\ x$.

and finally an aggregation rule $\upsilon$ that takes whole individual preferences in a preference profile $P$ as inputs and produces a complete social preference as the output. The aggregation problem is static since individual set and alternative set are fixed, and preferences are taken as given. From this point on, we impose some "desirable" conditions on aggregation rule $\upsilon$ and analyze the consequences.

An aggregation rule $\upsilon$ is *independent of irrelevant alternatives* (IIA) iff given any distinct alternatives $x, y \in A$ and any preference profiles $P, P' \in L(A)^N$ with $xP_iy \Longleftrightarrow xP'_iy$ for all individual $i \in N$, we have $x\ \upsilon(P)\ y \Longleftrightarrow x\ \upsilon(P')\ y$. In words, the social ranking between any two alternatives only depends on the individual rankings of those particular alternatives. Thus, all the other alternatives are irrelevant on determining the social ranking between these two alternatives.[16] Given its restrictive nature, this condition is the most controversial among other conditions in Arrovian framework. We discuss and provide evidence on this claim further in the text. We write $\Phi$ for the set of aggregation rules which satisfy IIA.

For any distinct alternatives $x, y \in A$, the set of complete and transitive preferences over $\{x, y\}$ are $\{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}, xy\}$[17], and the set of complete, transitive and antisymmetric preferences over $\{x, y\}$ are $\{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}\}$. An *elementary aggregation rule* is a mapping $\upsilon_{\{x,y\}} : \{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}\}^N \rightarrow \{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}, xy\}$ where $\{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}\}^N$ is the domain of preference profiles over $\{x, y\}$. Thus, for each pair of alternatives, one can define an elementary

[16] Moreover, IIA is also known as binary independence in the literature. The reason is independence notion is defined over pairs.

[17] We interpret $\begin{matrix} x \\ y \end{matrix}$ as $x$ being preferred to $y$; $\begin{matrix} y \\ x \end{matrix}$ as $y$ being preferred to $x$; and $xy$ as indifference between $x$ and $y$.

aggregation rule which maps individual preferences over that pair onto complete and transitive social preferences over that specified pair.[18] Any family $\{\upsilon_{\{x,y\}}\}$ of elementary aggregation rules indexed over all possible distinct pairs $x, y \in A$ induces an aggregation rule $\upsilon$ as follows: For each preference profile $P \in L(A)^N$ and each distinct alternatives $x, y \in A$, let $x\ \upsilon(P)\ y \iff \upsilon_{\{x,y\}}(P^{\{x,y\}}) \in \{\begin{matrix} x \\ y \end{matrix}, xy\}$ where $P^{\{x,y\}} \in \{\begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}\}^N$ is the restriction of preference profile $P \in L(A)^N$ over $\{x, y\}$.[19] In words, the social preference between two alternatives $x, y$ under the aggregation rule $\upsilon$ will be the same as the social preference between the same alternatives under the elementary aggregation rule $\upsilon_{\{x,y\}}$. Note that $\upsilon = \{\upsilon_{\{x,y\}}\}$ is IIA. Moreover, any IIA aggregation rule $\upsilon$ can be expressed in terms of a family $\{\upsilon_{\{x,y\}}\}$ of elementary aggregation rules.[20] The intuition behind this result is that IIA requires considering individual preferences only on a particular pair while deciding on a social preference on that pair which implies that an IIA aggregation rule is at the end a combination of elementary aggregation rules over all pairs.

We need elementary aggregation rules in the characterization of weakly IIA Social welfare functions. A *Social Welfare Function* (SWF) is an aggregation rule whose range is restricted to $CT(A)$ where $CT(A)$ is the set of complete and transitive binary relations over $A$. It is easy to see that a social welfare is more demanding than an aggregation rule since it requires an ordering of the social preferences.[21] In fact, requiring transitivity of social preferences will be crucial when we impose other desirable conditions. We will define the rest of the conditions for social welfare functions.

[18]Note that all preferences over two alternatives are always transitive.

[19]So for any $i \in N$, we have $P_i^{\{x,y\}} = \begin{matrix} x \\ y \end{matrix} \iff xP_iy$.

[20]This claim is first stated by Gibbard (1968).

[21]Completeness and transitivity of social preferences together are called collective rationality or full rationality in the literature.

A SWF $\alpha : L(A)^N \to CT(A)$ is *Paretian* iff given any distinct alternatives $x, y \in A$ and any preference profile $P \in L(A)^N$ with $xP_iy$ for all individual $i \in N$, we have $x$ $\alpha^*(P)$ $y$. In words, if everyone in the society prefers one alternative over the other, then society must have the same preference. This is the most obvious and straightforward Arrovian condition. No one will argue that Pareto condition is undesirable or restrictive. Even though it seems a milder condition, along with IIA it is quite restrictive.

A SWF $\alpha : L(A)^N \to CT(A)$ is *dictatorial* iff there exists an individual $i \in N$ such that $xP_iy$ implies $x$ $\alpha^*(P)$ $y$ for all preference profile $P \in L(A)^N$ and for all distinct alternatives $x, y \in A$. This condition requires that the decision power must not be held by a single individual. The reason is everyone's preferences should be taken into account in the aggregation process. Similar to Pareto condition, nondictatoriality condition seems less demanding. However, the Arrovian impossibility, as we consider, states that a SWF $\alpha : L(A)^N \to CT(A)$ is Paretian and IIA if and only if $\alpha$ is dictatorial. In other words, it is impossible to obtain a SWF that satisfies IIA, Pareto condition and nondictatoriality.

In order to have a better understanding what leads to impossibility, one of the Arrovian conditions either dropped or replaced with a weaker version in the literature. Soon, it is realized that the real problem is the tension between transitivity of social preferences and IIA. IIA forces a group of individuals being decisive over a pair of alternatives.[22] By IIA, once a group is decisive over a pair of alternatives, then the same group is decisive over all pairs. However, this will create cycles in social preferences along with Pareto principle. Hence, the group shrinks to a single individual which will be the dictator. Given Arrovian impossibility is remarkably robust against weakenings

[22] A group of individual is decisive over a pair of alternatives if the social preference over that pair is the same as the individuals' preferences in that group whenever they all have the same preferences over that pair.

of Arrovian conditions, what would happen if two or more conditions are simultaneously weakened? The next section seeks an answer to this question.

## 3.4 Results

Baigent (1987) proves a version of the Arrovian impossibility where IIA and dictatoriality are replaced by their following weaker versions: A SWF $\alpha$ is *weak IIA* iff given any distinct alternatives $x, y \in A$ and any distinct preference profiles $P, P' \in L(A)^N$ with $xP_iy \iff xP'_iy$ for all individual $i \in N$, we have $x\ \alpha^*(P)\ y \Rightarrow x\ \alpha(P')\ y$. Weak IIA is less demanding than IIA in the sense that it allows social indifference among two alternatives in one profile whenever one alternative is strictly socially preferred to another in the other profile where the individual rankings of these two alternatives are the same in both profiles. Note that weak IIA and IIA coincide when indifferences are ruled out from the social preference. However, in terms of informational requirement, weak IIA is open to use relative orderings of a pair in individual preferences on deciding their social ranking whereas IIA does not take into account those orderings at all. In that sense, one might argue how the social decision on a pair of alternatives independent from other alternatives and those alternatives are relevant or irrelevant.

A SWF $\alpha$ is *weakly dictatorial* iff there exist an individual $i \in N$ such that $xP_iy$ implies $x\ \alpha(P)\ y$ for all preference profile $P \in L(A)^N$ and for all distinct alternatives $x, y \in A$. In words, this condition requires that there is an individual whose preferences over any two alternatives cannot be reversed as a social ranking over those two alternatives. Still the dictator has decision power on social ranking over all alternatives.

When IIA is replaced with weak IIA and dictatoriality is replaced with weak dictatoriality (simultaneous weakening of two conditions), Baigent (1987) establishes that every Paretian and weak IIA SWF is a weak dictatorship. Nevertheless, we remark that, unlike the original version of the Arrovian impossibility, the converse statement of Baigent (1987) is not true: Although every weak dictatorship is weak IIA, there exists weak dictatorships that are not Paretian.[23] Following this remark, we allow ourselves to state a slight generalization of this theorem of Baigent (1987), corrected by Campbell and Kelly (2000b)[24]:

**Theorem 3.1** *Let $\#A \geq 4$. Within the family of Paretian SWFs, a SWF $\alpha : L(A)^N \to CT(A)$ is weak IIA iff $\alpha$ is weakly dictatorial.*

We are interested in obtaining a characterization of all weak IIA SWFs. However, this results only provides a characterization of weak IIA SWFs within the family of Paretian SWFs. Thus, first we explore the effect of being confined to the class of Paretian SWFs.

The strict counterpart of a complete binary relation $S$ over all alternatives in $A$ is denoted as $S^*$. Let $\rho : C(A) \longrightarrow 2^{CT(A)}$ stand for the correspondence which transforms each complete binary relation $S$ over $A$ into a collection of complete and transitive binary relations over $A$ such that $\rho(S) = \{R \in CT(A) : xSy \Longrightarrow xRy \ \forall x, y \in A\}$.[25] Basically, the transformation $\rho$ generates all complete and transitive binary relations that contains complete binary relation $S$, not just the *transitive closure* of $S$ which is the intersection of all the complete and transitive binary relations in $\rho(S)$. Moving from complete binary relations to complete and transitive binary relations is necessary since we deal with social welfare functions which require the transitivity of social preferences.

In order to have a clearer understanding of how exactly the transformation $\rho$ is

[23]For example the SWF $\alpha$ where $x \ \alpha(P) \ y$ for all distinct alternatives $x, y \in A$ and for all preference profile $P \in L(A)^N$ is a weak dictatorship but not Paretian.

[24]See Footnote 3.

[25]Note that the transformation is from a complete binary relation to a set of complete and transitive binary relations.

carried out, we recall that every complete binary relation $S$ over $A$ induces an ordered list of "cycles".[26] A nonempty subset $Y \subseteq A$ is a *cycle* (with respect to $S \in C(A)$) iff $Y$ can be written as $Y = \{y_1, ..., y_{\#Y}\}$ such that $y_i S y_{i+1}$ for all $i \in \{1, ..., \#Y - 1\}$ and $y_{\#Y} S y_1$. The *top-cycle* of a nonempty subset $X \subseteq A$ with respect to $S \in C(A)$ is a cycle $K(X, S) \subseteq X$ such that $yS^*x$ for all $y \in K(X, S)$ and for all $x \in X \backslash K(X, S)$.[27] In words, any alternative in the top-cycle is strictly preferred to alternatives that are not in the top-cycle. That implies one can rank (order) cycles over alternative set from top to bottom. So, let $A_1 = K(A, S)$ and recursively define $A_i = K(A \backslash \bigcup_{k=1}^{i-1} A_k, S)$ for all $i \geq 2$. Given the finiteness of $A$, there exists an integer $k$ such that $A_{k+1} = \emptyset$. So every complete binary relation $S$ over $A$ induces a unique ordered partition $(A_1, A_2, ....., A_k)$ of $A$. It follows from the definition of the top-cycle that whenever $i < j$, we have $xS^*y$ for all $x \in A_i$ and for all $y \in A_j$. The following example shows how cycles are constructed and ordered according to given definitions.

**Example 3.1** *Let $N = \{1, 2, 3, 4\}$ be the set of individuals and $A = \{a, b, c, d, e, f\}$ be the set of alternatives. Consider the following preference profile $P$ which contains four individual preferences over six alternatives:*

[26] We use the definition of "cycle" as stated by Peris and Subiza (1999).

[27] The top-cycle, introduced by Good (1971) and Schwartz (1972), has been explored in details. Moreover, Peris and Subiza (1999) extend this concept to weak tournaments. In their setting, as $K(X, S)$ is a cycle, there does not exist $Y \subset K(X, S)$ with $yS^*x$ for all $y \in Y$ and for all $x \in K(X, S) \backslash Y$.

| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|
| $a$ | $f$ | $c$ | $c$ |
| $b$ | $b$ | $a$ | $a$ |
| $c$ | $e$ | $d$ | $b$ |
| $d$ | $a$ | $e$ | $e$ |
| $e$ | $d$ | $b$ | $d$ |
| $f$ | $c$ | $f$ | $f$ |

*Suppose the aggregation rule $\upsilon$ is the pairwise majority rule and applied to the profile $P$. Then, we obtain a complete binary relation $\upsilon(P)$. Three individuals (1,3,4) prefer $a$ to $b$ against one individual (2) who prefers $b$ to $a$, thus $a\ \upsilon^*(P)\ b$. Two individuals (1 and 2) prefer $b$ to $c$ but the other two individuals (3 and 4) prefer $c$ to $b$ which is a tie, so $b\ \upsilon(P)\ c$ and $c\ \upsilon(P)\ b$. Similarly, two individuals (1 and 2) prefer $a$ to $c$ but the other two individuals (3 and 4) prefer $c$ to $a$ which is again a tie, so $a\ \upsilon(P)\ c$ and $c\ \upsilon(P)\ a$. Hence, $a\ \upsilon^*(P)\ b\ \upsilon(P)\ c\ \upsilon(P)\ a$ is a cycle and $x\ \upsilon^*(P)\ y$, for all $x \in \{a, b, c\}$, for all $y \in \{d, e, f\}$. Hence, $\{a, b, c\}$ is the top-cycle of the alternative set $A$ and denoted by $A_1$. Recursively, among the remaining alternatives, $d\ \upsilon(P)\ e\ \upsilon(P)\ d$ is the top-cycle with $x\ \upsilon^*(P)\ f$, for all $x \in \{d, e\}$. Thus, $A_2 = \{d, e\}$. There is only one remaining alternative, namely $f$, therefore $A_3 = \{f\}$. Finally, no alternatives left, thus $A_4 = \emptyset$. The following figure shows the ranking of cycles;*

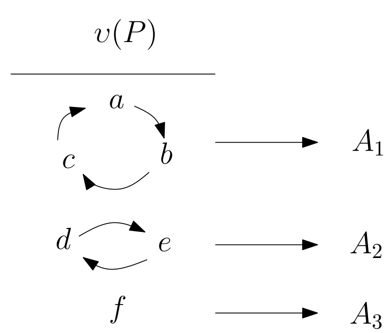


Figure 1: Ranking of the cycles

So, it is straightforward to observe that an alternative is strictly preferred to another alternative whenever it is located in a higher cycle. However, it is interesting to know what would happen to the comparison between alternatives in the same cycle when the transformation is carried out. The following lemma provides an answer to this question. It basically states that all the alternatives will be indifferent in any given cycle when the transformation is applied. The intuition behind the result is simple; cycles create intransitivity and the only way to restore transitivity is to eliminate cycles by creating indifferences among alternatives in each cycle. Since cycles are already ordered, then transitivity is obtained.

The proof of the lemma uses the fact that not allowing indifferences among the alternatives in a cycle in the transformation process will lead to a contradiction by violating transitivity.

**Lemma 3.1** *Take any complete binary relation $S \in C(A)$ which induces the ordered partition $(A_1, A_2, ....., A_k)$. Given any $A_i$ and any distinct alternatives $x, y \in A_i$, we have $xRy$ and $yRx$ for all $R \in \rho(S)$.*[28]

**Proof.** Take any complete binary relation $S \in C(A)$ which induces the ordered partition $(A_1, A_2, ....., A_k)$. Take any cycle $A_i$, any alternatives $x, y \in A_i$ and any complete and transitive binary relation $R \in \rho(S)$. If there is only one alternative in $A_i$, then $x$ and $y$ coincide, hence $xRy$ and $yRx$ holds by the completeness of $R$. If there are two alternatives in $A_i$, then $xSy$ and $ySx$ since $A_i$ is a cycle, which implies $xRy$ and $yRx$ since $R \in \rho(S)$. We complete the proof by considering the case where there are more than three alternatives in $A_i$. Let $A_i = \{x_1, x_2, ....., x_k\}$. Suppose, without loss of generality, $x_1Rx_2$ and *not* $x_2Rx_1$. This implies $x_1S^*x_2$ since $R \in \rho(S)$. Moreover, since $A_i$ is a cycle, there exists another alternative $x \in A_i$ such that $x_2Sx$. Let, without

[28]The properties of top-cycles has been well discussed in the literature (See Deb (1977)).

loss of generality, $x = x_3$, so $x_2 S x_3$. Thus $x_2 R x_3$ holds by definition of $\rho$ which implies $x_1 R x_3$ and *not* $x_3 R x_1$ by the transitivity of $R$. Again by definition of $\rho$, we have $x_1 S^* x_3$. Since $A_i$ is a cycle, there exists $j \in \{4, ....., k-1\}$ such that $x_3 S x_j$. Suppose, without loss of generality, $j = 4$. So $x_3 S x_4$, hence $x_3 R x_4$, implying $x_1 R x_4$ and *not* $x_4 R x_1$, which in turn implies $x_1 S^* x_4$. So, iteratively, for all $i \in \{4, ...., k-1\}$, we have $x_i S x_{i+1}$, which implies $x_i R x_{i+1}$ and moreover $x_1 R x_{i+1}$ and *not* $x_{i+1} R x_1$. Hence, $x_1 S^* x_{i+1}$. But since $A_i$ is a cycle, we have $x_k S x_1$. So $x_k R x_1$ holds by definition of $\rho$. As we also have $x_i R x_{i+1}$, for all $i \in \{1, ..., k-1\}$, $x_2 R x_1$ holds by transitivity of $R$, which leads to a contradiction. Therefore, $xRy$ and $yRx$ for all $x, y \in A_i$, for all $R \in \rho(S)$. ■

Thus for any complete binary relation $S \in C(A)$ which induces the ordered partition $(A_1, A_2, ....., A_k)$ and any complete and transitive binary relation $R \in CT(A)$, we have $R \in \rho(S)$ if and only if for any alternatives $x, y \in A$ $(i)$ $x, y \in A_i$ for some $A_i$ implies $xRy$ and $yRx$ and $(ii)$ $x \in A_i$ and $y \in A_j$ for some $A_i$, $A_j$ with $i < j$ implies $xRy$.[29]

The following example will be helpful to understand how the transformation is carried out.

**Example 3.2** *Let $N = \{1, 2, 3, 4\}$ be the set of individuals and $A = \{a, b, c, d, e, f\}$ be the set of alternatives. Consider the following preference profile $P$ that we already used in Example 1:*

[29] As a remark, these results can also be obtained by using graph theory. Any complete binary relation $S$ over $A$ is a directed graph (digraph) $G(S)$ defined as follows; the elements of $A$ becomes the vertices of digraph $G(S)$ and the edges are determined by $S$. ($xSy$ means that there is an edge from $x$ to $y$ and another one from $y$ to $x$.) Then, the ordered partitions of the transitive closure of the complete relation $S$ are the strong components of the digraph $G(S)$. Hence, it is already well known as an elementary result that the set of these strong components are linearly ordered and the maximal strong component is equivalent to the notion of top-cycle.(See Bang-Jensen and Gutin (2007))

| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|
| $a$ | $f$ | $c$ | $c$ |
| $b$ | $b$ | $a$ | $a$ |
| $c$ | $e$ | $d$ | $b$ |
| $d$ | $a$ | $e$ | $e$ |
| $e$ | $d$ | $b$ | $d$ |
| $f$ | $c$ | $f$ | $f$ |

*Similarly, suppose the aggregation rule $\upsilon$ is the pairwise majority rule and applied to the profile $P$. Then, from Example 1 we obtain the cycles as $A_1 = \{a, b, c\}$, $A_2 = \{d, e\}$, and $A_3 = \{f\}$. Given these ordered partitions, cycles, we can obtain the following complete and transitive binary relations by applying the transformation $\rho$;*

| $R_1$ | $R_2$ | $R_3$ | $R_4$ |
|---|---|---|---|
| $abc$ | $abc$ | $abcde$ | $abcdef$ |
| $de$ | $def$ | $f$ | |
| $f$ | | | |

*Here, $xyz$ means $x, y$ and $z$ are all indifferent. In the process of creating transitive binary relations from complete binary relations, all the alternatives in the same cycle must be indifferent from Lemma 1 and all the alternatives in lower cycles can be moved upward to higher cycles which results in more indifferences. For instance, $R_2$ is obtained by moving cycle $A_3$ to the cycle $A_2$. In fact, the extreme point is the indifference of all alternatives which is denoted by $R_4$.*

We now proceed towards characterizing the family of weak IIA SWFs. Take any aggregation rule $\upsilon \in \Phi$. By composing $\upsilon$ with the transformation $\rho$, we get a *social welfare correspondence* $\rho \circ \upsilon$: $L(A)^N \longrightarrow 2^{CT(A)}$ which assigns to each preference profile $P \in L(A)^N$ a non-empty subset $\rho(\upsilon(P))$ of $CT(A)$. Possibly there are more than one complete and transitive binary relation in $\rho(\upsilon(P))$. Clearly, every singleton-valued selection of $\rho \circ \upsilon$ is a SWF.[30] We define a set containing all single-valued selection of $\rho \circ \upsilon$.[31] Let $\Sigma^{\upsilon} = \{\alpha : L(A)^N \to CT(A) \mid \alpha$ is a singleton-valued selection of $\rho \circ \upsilon$ $\}$. As a next step, we define the set as the union of the sets that contains all aggregation rules transformed under $\rho$. So, we write $\Sigma = \cup_{\upsilon \in \Phi} \Sigma^{\upsilon}$. Interestingly, the class of quasi IIA SWFs coincides with the set $\Sigma$.

The intuition behind the result is the following; imposing IIA to any aggregation rule will create cycles. One way to remove cycles and restore transitivity is to set up indifferences among alternatives in the same cycle whenever there are any. The immediate consequence of this process is to loose IIA but gain a weaker version of IIA, namely weak IIA. In fact, the following result shows that this is the only way to obtain weak IIA social welfare functions.

The only if part of the proof requires a construction of an elementary aggregation rule since we need to show the existence of an aggregation rule that coincides with the given SWF after the transformation $\rho$ is applied. Once we construct the elementary aggregation rule, it is straightforward to see that the family of these elementary aggregation rules constitutes the desirable aggregation rule. Hence, its transformation under $\rho$ will coincide with the given SWF. The if part shows that a social welfare function which is obtained by removing the cycles of an IIA aggregation rule has to satisfy weak IIA. The reason is while creating transitivity by removing cycles, the process does not reverse any strict preferences among alternatives, it only change some of them to indifferences. As a result, this will violate IIA but not weak IIA. The proof creates a contradiction in the outcome of an elementary aggregation rule if we suppose that the social welfare function does not satisfy weak IIA.

[30]We say that $\alpha : L(A)^N \to CT(A)$ is a singleton-valued selection of $\rho \circ \upsilon$ iff $\alpha(P) \in \rho \circ \upsilon(P)$ $\forall P \in L(A)^N$.

[31]In other words, all SWFs produced from aggregation rule $\upsilon$ under the transformation of $\rho$.

**Theorem 3.2** *A SWF $\alpha : L(A)^N \to CT(A)$ is weak IIA iff $\alpha \in \Sigma$.*

**Proof.** To establish the "only if" part, let $\alpha : L(A)^N \to CT(A)$ be a weak IIA SWF. For any distinct alternatives $x, y \in A$, we define an elementary aggregation rule $\upsilon_{\{x,y\}} : \{ \begin{smallmatrix} x \\ y \end{smallmatrix}, \begin{smallmatrix} y \\ x \end{smallmatrix} \}^N \to \{ \begin{smallmatrix} x \\ y \end{smallmatrix}, \begin{smallmatrix} y \\ x \end{smallmatrix}, xy\}$ as follows:

For any preference profile $r \in \{ \begin{smallmatrix} x \\ y \end{smallmatrix}, \begin{smallmatrix} y \\ x \end{smallmatrix} \}^N$,

$$\upsilon_{\{x,y\}}(r) = \begin{cases} \begin{matrix} x \\ y \end{matrix} & \text{if} \quad x\ \alpha^*(P)\ y \text{ for some } P \in L(A)^N \text{ with } P^{\{x,y\}} = r \\ \\ \begin{matrix} y \\ x \end{matrix} & \text{if} \quad y\ \alpha^*(P)\ x \text{ for some } P \in L(A)^N \text{ with } P^{\{x,y\}} = r \\ \\ xy & \text{if } x\ \alpha(P)\ y \text{ and } y\ \alpha(P)\ x \text{ for all } P \in L(A)^N \text{ with } P^{\{x,y\}} = r \end{cases}$$

Since $\alpha$ is weak IIA, $\upsilon_{\{x,y\}}$ is well-defined. Thus, we can define an aggregation rule $\upsilon$ as $\upsilon = \{\upsilon_{\{x,y\}}\} \in \Phi$. We now show $\alpha(P) \in \rho(\upsilon(P))$ for all preference profile $P \in L(A)^N$. Take any preference profile $P \in L(A)^N$ and any distinct alternatives $x, y \in A$. First let $x\ \upsilon^*(P)\ y$. So $\upsilon_{\{x,y\}}(P^{\{x,y\}}) = \begin{smallmatrix} x \\ y \end{smallmatrix}$. By definition of $\upsilon_{\{x,y\}}$, we have $x\ \alpha^*(Q)\ y$ for some $Q \in L(A)^N$ with $Q^{\{x,y\}} = P^{\{x,y\}}$ which implies $x\ \alpha(P)\ y$ since $\alpha$ is weak IIA. If $y\ \upsilon^*(P)\ x$, then one can similarly obtain $y\ \alpha(P)\ x$. Now, let $x\ \upsilon(P)\ y$ and $y\ \upsilon(P)\ x$. So, $\upsilon_{\{x,y\}}(P^{\{x,y\}}) = xy$ which, by definition of $\upsilon_{\{x,y\}}$, implies $x\ \alpha(Q)\ y$ and $y\ \alpha(Q)\ x$ for all preference profiles $Q \in L(A)^N$ with $Q^{\{x,y\}} = P^{\{x,y\}}$, hence $x\ \alpha(P)\ y$ and $y\ \alpha(P)\ x$. Thus, $x\ \upsilon(P)\ y \Longrightarrow x\ \alpha(P)\ y$ for all alternatives $x, y \in A$, establishing $\alpha(P) \in \rho(\upsilon(P))$.

To establish the "if" part, take any SWF $\alpha \in \Sigma$. So there exists an aggregation rule $\upsilon \in \Phi$ such that $\alpha(P) \in \rho(\upsilon(P))$ for all preference profiles $P \in L(A)^N$. Suppose $\alpha$ is not weak IIA. So, there exist two alternatives $x, y \in A$ and there exist two preference profiles $P, Q \in L(A)^N$ with $P^{\{x,y\}} = Q^{\{x,y\}}$ such that $x\ \alpha^*(P)\ y$ and $y\ \alpha^*(Q)\ x$. By the definition of $\rho$ we have $x\ \upsilon^*(P)\ y$ and $y\ \upsilon^*(Q)\ x$ which implies $\upsilon_{\{x,y\}}(P^{\{x,y\}}) = \begin{array}{c} x \\ y \end{array}$ and $\upsilon_{\{x,y\}}(Q^{\{x,y\}}) = \begin{array}{c} y \\ x \end{array}$, giving a contradiction since $P^{\{x,y\}} = Q^{\{x,y\}}$, thus showing that $\alpha$ is weak IIA. ■

The immediate implication of Theorems 1 and 2 together is that removing the Pareto condition has a dramatic impact, since the class $\Sigma$ of weak IIA SWFs is fairly large and allows those where the decision power is not necessarily concentrated on a single individual. This positive result prevails even when the following weak Pareto condition is imposed: A SWF $\alpha$ is *weakly Paretian* iff given any distinct alternatives $x, y \in A$ and any preference profile $P \in L(A)^N$ with $xP_iy$ for all individual $i \in N$, we have $x\ \alpha(P)\ y$. In words, weak Pareto requires that social ranking of two alternatives cannot be reversed as opposed to individual rankings of those alternatives when all individuals have the same ranking. Similarly, an aggregation rule $\upsilon \in \Phi$ is *weakly Paretian* iff for any distinct alternatives $x, y \in A$ and any preference profile $r \in \{\begin{array}{c} x \\ y \end{array}, \begin{array}{c} y \\ x \end{array}\}^N$ with $r_i = \begin{array}{c} x \\ y \end{array}$ for all individual $i \in N$, we have $\upsilon_{\{x,y\}}(r) \in \{\begin{array}{c} x \\ y \end{array}, xy\}$. Let $\Phi^*$ stand for the set of weak Paretian and IIA aggregation rules and $\Sigma^* = \cup_{\upsilon \in \Phi^*} \Sigma^{\upsilon}$ similar to the set $\Sigma$ where the only difference is in aggregation rules.

The following result is a straightforward extension of the main result. Moreover, instead of requiring aggregation rule to be Paretian, weak Pareto will be sufficient to obtain weak Paretian and weak IIA social welfare functions. Weak Pareto principle is easily carried out from aggregation rule to social welfare function and vice versa.

The proof is very similar to the proof of Theorem 2. The only if part uses the same elementary aggregation rule that we constructed in the only if part of the proof of Theorem 2. The only addition is to show that SWF is weak Paretian. Given a weak Paretian aggregation rule, it is straightforward to prove the same condition has to satisfy by the SWF as well. Again, "if" part of the proof is the same as the "if" part of the proof of Theorem 2 except showing that SWF is weak Paretian.

**Theorem 3.3** *A SWF $\alpha : L(A)^N \to CT(A)$ is weak Paretian and weak IIA iff $\alpha \in \Sigma^*$.*

**Proof.** To show the "only if" part, take any SWF $\alpha : L(A)^N \to CT(A)$ which is weak Paretian and weak IIA. For any distinct alternatives $x, y \in A$, we define the same elementary aggregation rule $\upsilon_{\{x,y\}} : \{ \begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix} \}^N \to \{ \begin{matrix} x \\ y \end{matrix}, \begin{matrix} y \\ x \end{matrix}, xy\}$ that we used in Theorem 3.2. Since $\alpha$ is weak IIA, $\upsilon_{\{x,y\}}$ is well-defined. Thus, the family of these elementary functions induces an IIA aggregation rule $\upsilon$ where $\upsilon = \{\upsilon_{\{x,y\}}\} \in \Phi$. Suppose, $\upsilon$ is not weak Paretian. So, there exist two alternatives $x, y \in A$ and there exists a preference profile $P \in L(A)^N$ with $xP_iy$ for all $i \in N$ such that $y\ \upsilon^*(P)\ x$, implying $\upsilon_{\{x,y\}}(P^{\{x,y\}}) = \begin{matrix} y \\ x \end{matrix}$. By definition of $\upsilon_{\{x,y\}}$, we have $y\ \alpha^*(Q)\ x$ for some $Q \in L(A)^N$ with $Q^{\{x,y\}} = P^{\{x,y\}}$, contradicting that $\alpha$ is weak Paretian, which establishes $\upsilon = \{\upsilon_{\{x,y\}}\} \in \Phi^*$. Therefore, following the "only if" part of Theorem 3.2 we obtain $\alpha(P) \in \rho(\upsilon(P))$ for all preference profiles $P \in L(A)^N$.

To show the "if" part, take any SWF $\alpha \in \Sigma^*$. So there exists a weak Paretian and IIA aggregation rule $\upsilon \in \Phi^*$ such that $\alpha(P) \in \rho(\upsilon(P))$ for all preference profiles $P \in L(A)^N$. Take any distinct alternatives $x, y \in A$ and any preference profile $P \in L(A)^N$ with $xP_iy$ for all individual $i \in N$. By the weak Pareto condition of $\upsilon$, we have $\upsilon_{\{x,y\}}(P^{\{x,y\}}) \in$

$\{ \begin{smallmatrix} x \\ y \end{smallmatrix} , xy\}$, hence $x\ \upsilon(P)\ y$, which implies $x\ \alpha(P)\ y$ by the definition of transformation $\rho$. Thus, SWF $\alpha$ is weak Paretian. The "if"part of Theorem 2 establishes that $\alpha$ is weak IIA, completing the proof. ■

We close the section by giving an example covered by Theorem 2 and Theorem 3. In fact, at any preference profile $P \in L(A)^N$, for any aggregation rule $\upsilon \in \Phi$, take the transitive closure of the social preference $\upsilon()$ as the selection of $\rho \circ \upsilon$[32].

**Example 3.3** *Let $N = \{1, 2, 3, 4\}$ be the set of individuals and let $A = \{a, b, c, d\}$ be the set of alternatives. Let $\upsilon$ be the pairwise majority rule. Consider the following preference profile $P$ which contains four individual orderings of four alternatives:*

| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|
| $a$ | $a$ | $b$ | $d$ |
| $b$ | $b$ | $a$ | $c$ |
| $c$ | $c$ | $c$ | $b$ |
| $d$ | $d$ | $d$ | $a$ |

Since $\upsilon$ is the pairwise majority rule, $a\ \upsilon(P)\ b$ and $b\ \upsilon(P)\ a$, and $x\ \upsilon^*(P)\ y$, for all alternatives $x \in \{a, b\}$, for all alternatives $y \in \{c, d\}$, and $c\ \upsilon^*(P)\ d$. Thus, the ordered partitions, cycles, are obtained as $A_1 = \{a, b\}$, $A_2 = \{c\}$, $A_3 = \{d\}$, and $A_4 = \emptyset$. Then, if we take the transitive closure of the social preference $\upsilon()$ as the selection of $\rho \circ \upsilon$, we obtain a SWF $\alpha$ which satisfies weak IIA by Theorem 2. For this particular preference profile $P$, the social outcome is

[32] By "taking the transitive closure", we mean to replace cycles with indifference classes. Formally speaking, writing $(A_1, A_2, ....., A_k)$ for the ordered partition induced by $\upsilon(P) \in \Theta$ at $P \in L(A)^N$, take $\alpha(P) \in \rho(\upsilon(P))$ where $x\ \alpha^*(P)\ y$ for all alternatives $x \in A_i$ and for all alternatives $y \in A_j$ with $i < j$. One can see Sen (1986) for a general discussion of the "closure methods".

$$\begin{array}{c} \underline{\alpha(P)} \\ a, b \\ c \\ d \end{array}$$

Moreover, it is easy to see that under this SWF $\alpha$, no individual has veto power.[33]

## 3.5 Conclusion

Within the scope of the preference aggregation problem, we contribute to the understanding of the well-known tension between requiring the pairwise independence of the aggregation rule and the transitivity of the social preference. As Wilson (1972) shows, a SWF $\alpha : L(A)^N \to CT(A)$ is *non-imposed*[34] and IIA if and only if $\alpha$ is dictatorial or *antidictatorial*[35] or *null*[36]. Thus, aside from these, any aggregation rule which is IIA allows non-transitive social outcomes. In case these outcomes are rendered transitive according to one of the prescriptions made by $\rho$, we attain a SWF which fails IIA but satisfies weak IIA. In fact, as Theorem 3.2 states, the class of weak IIA SWFs coincides with those which can be attained through a selection made out of the social welfare correspondence obtained by the composition of an aggregation rule that is IIA with $\rho$. This

[33]Bordes (1976) and Baigent and Klamler (2004) considers this particular rule for other reasons. However, none of them provide a full characterization of it.

[34]$\alpha : L(A)^N \to CT(A)$ is *non-imposed* iff $\forall x, y \in A\ \exists P \in L(A)^N$ with $x\ \alpha(P)\ y$. In words, every alternative is socially preferred to any other alternative in at least one profile of individual preferences.

[35]$\alpha$ is *anti-dictatorial* iff $\exists i \in N$ such that $xP_iy \Longrightarrow y\ \alpha^*(P)\ x\ \forall P \in L(A)^N, \forall x, y \in A$. This is the reversed version of dictatoriality where there is an individual whose individual preferences between any two alternatives is exactly the opposite of the social preference no matter what the other individual preferences are.

[36]$\alpha : L(A)^N \to CT(A)$ is *null* iff $x\ \alpha(P)\ y\ \forall x, y \in A$ and $\forall P \in L(A)^N$. In words, every alternative is socially indifferent no matter what the individual preferences are.

can be interpreted as a positive result, since the class of weak IIA SWFs is fairly rich and not restricted to those where the decision power is concentrated on one individual. In fact, this class contains SWFs that are both anonymous and neutral.[37] Moreover, as Theorem 3.3 states, this positive result prevails when a weaker version of the Pareto condition is imposed. Thus, we can conclude that the transitivity of the social outcome can be achieved at a cost of reducing IIA to weak IIA and compromising of the strength of the Pareto condition.

This paper does not attempt to overcome the Arrovian impossibility. Even though by simultaneously weakening IIA and Pareto condition we obtain a fairly large class of SWFs, it is not obvious to determine how large this class is and any of these SWF can be implementable. Moreover, there is a weakness on the weaker versions of IIA and Pareto condition. Weak IIA uses all the available information meaning that the other alternatives are not irrelevant on determining the social ranking between any two alternatives. In that sense, independence notion becomes ambiguous. Similarly, weak Pareto condition allows social indifference among two alternatives even when all individuals strictly prefer one alternative over another. The justification for this type of social preference is lacking. We need a better understanding of these weaker versions and more work is required on this direction.

Another way of looking at the problem is to conceive it as determining the possible "stretchings" of the null rule (which is well-known to be IIA) without violating weak IIA. So it is worth exploring "how far" weak IIA SWFs are from the null rule. This exploration requires to ask for the minimization of the imposed social indifference. The answer is straightforward for a given aggregation rule $\upsilon \in \Phi$: Taking the transitive

[37] As a matter of fact, the SWF in Example 2 of Campbell and Kelly (2000b), which shows the failure of Theorem 3.1 for $\#A = 3$, belongs to this class.

closure of the social preference is the selection of $\rho \circ \upsilon$ which minimizes the imposed social indifference. Nevertheless, the choice of the (non-dictatorial) $\upsilon$ that minimizes the imposed social indifference remains as an interesting open question.[38] In addition to this, allowing indifferences in individual preferences is worth to investigate and we leave it as a further work.

[38] We conjecture, by relying on Dasgupta and Maskin (2008), that this will be the pairwise majority rule.

# Bibliography


[1] Arrow, K. J. (1951) *Social Choice and Individual Values*, John Wiley, New York.

[2] Baigent, N. (1987) "Twitching weak dictators", *Journal of Economics*, Vol. 47, No. 4: 407 - 411.

[3] Baigent, N. and C. Klamler (2004) "Transitive closure, proximity and intransitivities", *Economic Theory*, 23: 175-181.

[4] Bang-Jensen, J. and G. Gutin (2007) Digraphs: Theory, Algorithms and Applications, Springer-Verlag, New York.

[5] Barberà, S. (2003) "A theorem on preference aggregation", WP166 CREA-Barcelona Economics.

[6] Blau, J. H. (1971) "Arrow's Theorem with weak independence", *Economica*, Vol. 38, No. 152: 413 - 420.

[7] Bordes, G. (1976) "Consistency, Rationality, and Collective Choice", *The Review of Economic Studies*, Vol. 43, No. 3: 451-457.

[8] Campbell, D. E. (1976) "Democratic preference functions", *Journal of Economic Theory*, 12: 259 - 272.

[9] Campbell, D. E. and J. S. Kelly (2000a) "Information and preference aggregation", *Social Choice and Welfare*, 17: 3 - 24.

[10] Campbell, D. E. and J. S. Kelly (2000b) "Weak independence and veto power", *Economics Letters*, 66: 183 - 189.

[11] Campbell, D. E. and J. S. Kelly (2007) "Social welfare functions that satisfy Pareto, anonymity and neutrality but not IIA",*Social Choice and Welfare*, 29: 69 - 82.

[12] Dasgupta, P. and E. Maskin (2008) "On the robustness of majority rule", *Journal of the European Economic Association*, 6(5): 949-973.

[13] Deb, R. (1972) "On Schwartz's Rule", *Journal of Economic Theory*, 16: 103 -

110.

[14] Denicolo, V. (1998) ”Independent decisiveness and the Arrow Theorem”, *Social Choice and Welfare*, 15: 563 - 566.

[15] Gibbard, A. (1968) ”Social Choice and the Arrow Conditions”, *unpublished manuscript*.

[16] Good, I. J. (1971) ”A note on Condorcet sets”, *Public Choice*, 10: 97 - 101.

[17] Ozdemir, U. and R. Sanver (2007) ”Dictatorial domains in preference aggregation”, *Social Choice and Welfare*, 28: 61 - 76.

[18] Peris, J. E. and B. Subiza (1999) ”Condorcet choice correspondences for weak tournaments”, *Social Choice and Welfare*, 16: 217 - 231.

[19] Schwartz, J. (1972) ”Rationality and the myth of maximum”, *Nous*, Vol. 6, No. 2: 97 - 117.

[20] Sen, A. (1986) ”Social choice theory”, in *Handbook of Mathematical Economics (Volume 3),* Eds. KJ Arrow and MD Intriligator, North-Holland.

[21] Wilson, R. (1972) ”Social Choice Theory without the Pareto principle”, *Journal of Economic Theory*, 5: 478 - 486.