Spring 5-15-2016

Essays on Economic Decision Making

Hee Chun Kim
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Economics

Dissertation Examination Committee:
Brian W. Rogers, Chair
Mariagiovanna Baccara
Baojun Jiang
Paulo Natenzon
Jonathan Weinstein

Essays on Economic Decision Making

by
Hee Chun Kim

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 2016
St. Louis, Missouri

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENT

I am greatly indebted to Professor Brian W. Rogers for his guidance and advise. His relentless support helped my academic and humane development. I also cannot thanksful enough to Professor Jonathan Weinstein and Paulo Natenzon for their passion and sincere support. I would not arrive at this stage without their insight and encouragement. I would like to thank Professor Mariagiovanna Baccara, Marcus Berliant, John Nachbar, Baojun Jiang, In-Koo Cho, and Ariel Rubinstein for their fruitful comments and encouragement. I am grateful to my colleagues in the Department of Economics for their time and effort to share their experience and idea. Mushe Harutyunian, Inkee Jang, Nicholas Agerakakis, and Jiemai Wu are deserve for my special thank. I should not forget friends in and out of the department for their support. Duk-Gyoo Kim, Minho Kim, Duksang Cho, Sunha Myong, Jaevin Park, Jungho Lee, Siewon Kim, Yao Yao, Sangyup Choi, Jongho Park, Joonkyu Choi, Kwangyong Park, and Youngdeok Hwang are also deserve for it. Finally, I would like to sincrely thank to my famliy for their love and support.

Dedicated to my parents and grandparents.

ABSTRACT OF THE DISSERTATION

Essays on Economic Decision Making

by

Hee Chun Kim

Doctor of Philosophy in Economics

Washington University in St. Louis, 2016

Doctor Brian W. Rogers, Chair

My dissertation consists of three chapters and I take different approach in each chapter to investigate economic decision making behavior. The first chapter analyzes individuals' strategic decision making when players have replaceable identities and private information in a repeated prisoner's dillmma game. The second chapter studies individuals' non-strategic decision making when she has incomplete information about her underlying preference in a sequential choice situation. The third chapter experimentally examines a link between an individual's strategic thinking and non-strategic decision making in a setting designed to elicit beliefs about independent random variables.

In the first chapter, I focus on strategic decision making of economic agents when they are replaceable in a repeated prisoner's dilemma. I assume that agents have different private information that restricts their set of actions, and that replacement of agent involves change of such private information. In this environment, some agents are required to signal their own private information to induce their opponent's cooperative response, which may induce Pareto improvement of their expected continuation payoffs. Except for a trivial equilibrium, we can have non-trivial equilibria supporting cooperative action as a part of the equilibrium play; however, different from the environment with two long-run agents, replaceable agents environment puts a restriction on an existence of the equilibrium in which agents share the risk of type uncertainty equally regardless of the past history. Because of replacement, agents can avoid a full cost of signaling by shifting it to their successor upon their own replacement. As replacement incurs such a situation with a strictly positive probability, the equilibrium cannot avoid failure.

In the second chapter I focus on an economic agent's optimal decision making in a non-strategic environment. Especially, I study a sequential choice problem where an agent's preferences evolve over time. I assume that an agent has an underlying preference, and she learns about her underlying preference depending on her choice histories. Given that an agent makes an optimal decision upon her current available menu, I characterize the sequential choice behavior that follows a Sequential Weak Axiom of Revealed Preference (WARP-S). Using this characterization, I provide criteria for sequential choice data that recovers agent's underlying preference.

In the third chapter I and my co-author, Duk-Gyoo Kim, focus on a link between an optimal decision making in a non-strategic environment and strategic environment. Our research investigates whether an individual decision maker follows own subjective optimization in a non-strategic decision making, and such a difference in subjective optimization is correlated with strategic decision making pattern. We conducted two separate sessions in the same subject. Each session is designed to identify subjects' behavioral pattern in strategic and non-strategic decision making environment respectively. From the data, we observed that subjects' behavioral pattern shows significant similarity in two sessions.

1 Chapter 1: Two-Sided Player Replacement and Threshold Equilibrium

1.1 Introduction

During World War I, French and German troops confronted on the five-hundred-mile line of the Western Front. Long and boring trench warfare taught soldiers the policy called "Live-and-Let-Live." According to the policy, both sides implictly agreed not to attack each other unltil they observed any aggressive behavior (Axelrod and Hamilton (1981), Ashworth (1980)). However, each army's headquarter wanted her troop to take the opponent's trench. To push their troops to the front, the headquarters dispatched new troop leaders, and some of them aggressively led an attack against the opponent's trench. That means, even though each troop faced the same opponent troop over night, it is possible that the opponent troop leader's attitude could change over that time period. For this reason, both troops could not be sure whether the Live-and-Let-Live policy would continue until the next morning. This anecdote depicts a situation in which a public or a nominal identity is separated from an actual identity.

We can find a similar discrepancy between a nominal identity and an actual identity in our modern industry. For example, suppose that two investment firms separately invest in a joint project. In the

market, each firm invests in the project through the account of each firm and fully observes another firm's decision. All the while, an identity of an actual trader who makes an investment decision is inside information for each firm. Actual traders who are in charge of the decision may have different attitudes for investment and will try to maximize their own expected payoffs. Moreover, we can also imagine that traders can be randomly replaced by another trader without notification to another firm. Such identity separation between firm and trader can result in poor cooperation between two firms and lower the social benefit from successful joint investment[1] [2].

Inspired by the above examples, this study analyzes the effect of identity separation on the equilibrium outcome in a repeated prisoner's dilemma with perfect monitoring of a full history of actions from both sides. Players are one of two types: a Stackelberg type, who always defects, or a Normal type, who maximizes expected payoffs. Players are randomly replaced with known probability at the end of every period, and the replacement is not publicly observed. Playing only a defective action, which is a unique Nash equilibrium strategy for a one-shot game, is a trivial equilibrium of the repeated game with replacement of players. My question is whether we can find a non-trivial equilibrium that supports cooperate action as a part of equilibrium play, and how the equilibrium play changes depending on the game environment. Interestingly, there exists the set of common discount factors and the replacement rates such that non-trivial equilibria exist, and any set of such equilibria contains a pure strategy renegotiation-proof equilibrium, which is not dominated by any other pure strategy equilibrium. Moreover, I show that any set of non-trivial threshold equilibria

[1] Recently such replacement of players was largely issued in online game services like "League of Legends" or "Starcraft." In these games, multiple ($3 \sim 5$) online players are matched as a team and defeat their opponent team. Individual players are rated based upon their past game play, so that highly rated players are welcomed by the other players. At the same time, highly rated players also prefer to play the game with similar or more highly rated players to raise or maintain their current rating. However, there are many reported cases when professional gamers are paid to raise the rating for another individual's account. In online communities, these accounts are called as "power leveler" or "fake ranker" and they are more likely to play deliberate noncooperative action (called "trolling") or use an illegal hacking program (called "helper" or "game hack"). Since players cannot directly observe other actual players behind the account, they may assume a possibility that different players share the same account or a player is replaced by another player. Due to separation between a virtual account and an actual player, service providers can fail to completely detect whose in charge of aggresive behavior. As a result, players in online games cannot completely believe other players' identity and are less likely to cooperatively play to avoid possible loss. For more information, please see the link below.

[2] http://forums.na.leagueoflegends.com/board/showthread.php?t=183195, http://2p.com/2598150_1/One-Of-The-Best-Korean-LOL-Players-Got-Banned-For-1000-Years-by-EdwardsLin.html (English) http://www.thisisgame.com/webzine/news/nboard/4/?n=50415 (Korean)

contains a pure strategy renegotiation-proof equilibrium, which is not dominated by any other pure strategy equilibrium.

To find non-trivial equilibrium, I define a threshold equilibrium where Normal players cooperate when public beliefs that both players are a Normal type is sufficiently high. That is, the Normal players play cooperate action as long as they have a higher probability of being both players' types to be a Normal type than a certain threshold level. Depending on behavioral pattern, I categorize (non-trivial) threshold equilibrium into three cases: (1) infinite punishment (IP) equilibrium, (2) finite punishment (FP) equilibrium, and (3) penance (PN) equilibrium. Infinite punishment equilibrium begins with a cooperation phase in which the Normal players play cooperate action as long as they only observe cooperate action in the previous period. Once they observe defect action, they turn into a punishment phase at which any player plays defect action forever. Finite punishment equilibrium allows players to return to the cooperation phase after the punishment phase of finite periods. The cooperation phase is the same as infinite punishment equilibrium. Once they observe any defect action (during the cooperation phase), they turn to the punishment phase and play defect action for finite periods, and then return to the cooperation phase. Penance equilibrium is different from finite punishment equilibrium in that players distinguish a way to punish a two-sided deviation from a one-sided deviation. For a two-sided deviation in which two players play defect action at the same period during the cooperation phase, they turn to the punishment phase of finite periods. After finite periods of defect action, both sides are supposed to return to the cooperation phase at the next period. For a one-sided deviation in which only one side, say the "guilty side," plays defect action while the other, say the "innocent side," plays cooperate action, they turn to a penance phase. At the penance phase, the innocent side player plays only defect action until she observes cooperate action from the guilty side. The guilty side player must play cooperate action immediately when she enters the game as a Normal type. Once the guilty side plays cooperate action, both sides begin again the cooperation phase from the next period.

The first result we have is that we can find non-trivial threshold equilibrium, which includes cooperative action as a part of the equilibrium path. In the replaceable player environment, players

consider their survival to future with respect to the replacement rate. In other words, players discount their furture payoff upon their survival, so that replacement rate plays a similar role to the common discount factor (or time preference factor). This expectation of their future payoff may incentivize players to follow some equilibrium path if it provides them a higher expected contiuation payoffs than the trivial equilibrium's. Moreover, perfect monitoring helps players have a public belief, which characterzes the probability that both sides of player to be a Normal type player. Having a high enough public belief implies that each side of players believe their opponent as a Normal type and each are believed as a Normal type respectively. This public belief allows players to explictly compare their expected continuation payoffs (of some non-trivial equilibrium) with a minmax payoff of the trivial equilibrium. And then, we can characterize the minimum level of public belief as a "threshold" at which Normal players can play a cooperative action as an equilibrium play.

The second result is that the replacement rate governs the amount of uncertainty involved with private information and will affect the existence of non-trivial threshold equilibrium. At the high enough replacement rate and the time discount factor, players will find that signaling their type by cooperative action is too costly. When the replacement of player is too frequent, they cannot expect any cooperation will continue for long-enough periods. So that players would rather stay at the trivial equilibrium than playing a non-trivial equilibrium with some cost. At the intermediate level of two parameters, the infinite or finite punishment equilibrium can be supported. These equilibria allow players to share a risk from the type uncertainty equally in the form of fixed periods of the punishment phase. That is, regardless of "guiltiness" for causing the current punishment phase, both sides of players must play defective action for the same number of periods. In this aspect, we can consider them as the "equal treating" policy. However, as two parameters, the replacement rate and the common discount factor, approach zero, such an equal treating policy will not be supported and penance equilibrium will survive as a unique non-trivial threshold equilibrium.

Such failure of the equal treating policy comes from the replacement that separates a player from her behavior. Consider a Normal player who is newly replaced at the beginning of the punishment

phase. The player will only consider of her own expected continuation payoffs on the condition of her survival as a Normal type. That is, even if her signaling fails to earn a cooperative response from her opponent, she will not need to take full responsibility for such a failure. For this reason, the equal treating policy can face a failure when any newly replaced Normal player has a high enough belief in her opponent's type at the small enough replacement rate and the common discount factor. At the small enough parameters, the player may consider a trade-off between a cost of signaling and the amount of loss from the punishment phase. As the replacement rate becomes smaller, the length of the punishment phase will be longer because players need enough periods to guarantee the replacement of the player of a Stackelberg type with another Normal player after the beginning of the punishment phase. However, a cost of signaling will grow slower than the loss from the punishment phase since the replacement cuts down her cost of the "failed" signaling. As a result, the Normal player will find that the amount of payoff loss spent on the punishment phase is higher than a cost of signaling. To overcome such a systematic failure, players cannot avoid imposing full responsibility of a deviation (from the cooperation phase) on the guilty side by forcing a penance action. Penance equilibrium will not face such a systematic failure at any phase of the game, so that it will be supported as a unique non-trivial threshold equilibrium even in the asymptotic environment.

This change of equilibrium strategy distinguishes a two-sided replaceable player case from other cases. In the case of (non-replaceable) long-run players with type change, players can share relatively similar amounts of the risk even in the asymptotic environment (Ely and Valimaki (2012, 2013), Horner et al. (2011, 2013)). Such an equal treating policy can be maintained even in the asymptotic environment because players must endogenize a cost of failed signaling upon their furture play. For this reason, a full cost of signaling will be maintained as higher than a loss from the punishment phase. On the other hand, in a one-sided replaceable player case, a replaceable side player is asked to take most of the responsibility for the deviation, and a long-run player will refuse to share any risk until it becomes small enough (Mailath and Samuelson (2001, 2006)). Compared to these previous cases, a two-sided player replacement case shows transition between different

equilibrium paths of plays, and the transition is implied by a level of separation between a nominal and an actual identity.

The third result is that any nonempty set of non-trivial threshold equilibria always includes (non-trivial) renegotiation-proof equilibrium. Individual players in a two-sided replacement environment cannot expect when they will enter the game. Some players may enter the game in the middle of the cooperation phase and others may enter the game at the punishment phase. For this reason, even though players know that following the equilibrium path is sequentially rational behavior, they are tempted to renegotiate with their opponent to change a "phase." Confronting such possibility of a renegotiation, players require the equilibrium concept to be efficient and (internally) consistent. Efficiency implies that the expected continuation payoff of the equilibrium at each state should be at least equal to or better than that of another feasible equilibrium. Consistency implies that the equilibrium path should be maintained as it is predicted or specified as long as they are staying in the same game. To meet such considerations, I extend a notion of "renegotiation-proofness" from Pearce (1989) and Ray (1994) with modification to the current environment. I define renegotiation-proof equilibrium if it is not dominated by any other equilibria at any history. The threshold equilibrium is useful because we can find at least one non-trivial threshold equilibrium that satisfies such renegotiation-proof requirement among all pure-strategy (perfect Bayesian) equilibria. This finding supports threshold equilibrium as a reasonable choice when we confront the equilibrium selection problem in a two-sided replaceable player environment.

This paper proceeds as follows. In the following subsection, I discuss the related literature. Section 1 introduces a model and a game environment. Section 2 includes definition of threshold equilibrium and its relevant concepts. Section 3 characterizes three different forms of threshold equilibria according to their behavioral patterns. Section 4 shows existence of non-trivial threshold equilibrium in the asymptotic environment and existence of non-trivial renegotiation-proof equilibrium in the nonempty set of non-trivial threshold equilibria. Section 5 concludes the study and discusses future research.

1.2 Related Literature

1.2.1 Lifetime of Players

In a repeated game with perfect monitoring, a long-run and a short-run player is distinguished by their commitment to their future play. Long-run player is assumed to stay at the game until the end of it, so that they are responsible for their past action. For this reason, long-run player's identity is considered to be acknowledged. On the while, short-run player will not be assumed as the same player of the past history, so that they are considered as rather anonymous. This reason justifies short-run player's optimization on her current stage payoff.

In a non-cooperative repeated game, such lack of commitment plays central role for availability of equilibrium. Consider two-person prisoner's dilemma game. When both players are long-run players, enforcing certain equilibrium path by threatening each other with future punishment will work. Classic folk theorem results (Friedman (1971), Fudenberg and Levine (1983)) supports every equilibrium that achieves individually rational and feasible payoff as the discount factor approaches to unity. However, assuming short-run player on some side may change the range of available equilibrium payoff. When short-run player plays only one period of the game, any equilibrium that enforces cooperate action to short-run player will not work. A set of available equilibria will shrink to a trivial equilibrium that players only repeat a stage-game Nash equilibrium. Even though we allow short-run players to stay in the game for multiple periods, we may not achieve full cooperation as in the two long-run players case. When short-run player's finite life cycle is publicly known, what we can do is at most partial cooperation that brings cooperative outcome until a few periods before the end of each life cycle (Kreps et al (1982), Kreps and Wilson (1982)).

Now we extend our focus to the case in which short-run player has private information. We can consider several different form of private information; short-run player's own lifespan, payoff structure, a set of actions, etc. Mailath and Samuelson (2001, 2006) considered an imperfectly

monitored repeated game under which short-run players (firms) can have different type and randomly replaced by another short-run player. Each short-run player's type restricts a set of actions; "inept" firm only plays low effort, which corresponds defect action of the prisoner's dilemma game, while "competent" firm can take low and high effort, which corresponds defective and cooperate action respectively. A long-run player (a series of consumers) receives a noisy but distinguishable signal about the action of the replaceable players (firms). Except for signal, long-run players does not have any device to verify short-run player's identity or action. Moreover, short-run players are allowed to exchange lump-sum amount of payoff when they are replaced; several short-run players can compete to replace their predecessor with inheriting long-run player's belief about short-run player's type. In such setting, lump-sum exchange makes up lack of commitment for the short-run player. That is, the short-run player assumes that her effort level (action) will be compensated or punished through the lump-sum payment, which will be calculated based on the belief level they built, by her successor. This payment device works as a proxy for the short-run player's future expectation and supports long-run player's claim for the cooperate action. As a result, at low enough replacement rate, equilibrium encourages "competent" firm to play cooperate action to build own "reputation."

From this result, we may have a question about how replacement itself can affect equilibrium outcome without signaling device and/or side payment. Current study considers this question with assumption that players will not use any public signaling device that informs players' type or side payment device between actual players in the same side of account.

1.2.2 Stochastic Games and Replacement

Stochastic games that involves change of private information provides general guideline for role of replacement. Ely et al (2002; 2003), Hőrner et al (2011), Hőrner et al (2013) consider stochastic private information change while individual identity is fixed. Fixing individual identity allows players to take long-run player's role so that as they can achieve truthful equilibrium which en-

forces individual players to reveal their private type correctly. This outcome can be acquired only when they are equipped with the public/private signaling device that informs state of the world, which directly/indirectly contains the private type of players. Moreover, long-run players can internalize loss and benefit from their private type change, truthful equilibria supports players to share the risk of type change equally even in the asymptotic cases. That is, players expect that the amount of loss they have in some state will be canceled out by benefit in another state, they will not break equilibrium that shares risk of uncertainty equally among players. However, replaceable players cannot use such "long-run rebalancing," they are more likely to be tempted to exploit current private information.

Mailath and Samuelson (2000) considers the stochastic game that one-sided replacement of player involves the private type change. They considered the perfect (Bayesian) equilibrium under imperfect monitoring and side payment. Given that the signaling device provides statistically distinguishable information for the uncertain type of players, they can asymptotically achieve cooperative outcome as a part of equilibrium path. However, side payment between players in the same side plays connects different identity as if the same long-run player. That is, side payment directly transfers the accumulated/exploited value of the previous history between different individual identities, so that a string of replaceable players play a role of a single long-run player. As a result, they are allowed to achieve the similar (efficient) outcome from fixed individual identity even with replacement. However, only a one side of player has private information and long-run player, such reputation cannot be maintained. In Cripps et al (2004, 2007), whether the uninformed side player's belief about the informed side player's type is public (Cripps et al, 2004) or private (Cripps et al, 2007), the informed side player's reputation (as different type) will disappear eventually. That is, one-sided replacement of player also supports truthful equilibrium because players can specify who is responsible for deviation from equilibrium path. Different from one-sided replacement literature, this study focuses on a two-sided replaceable players environment in which players cannot completely specify responsibility of actual players. Two-sided replaceable case supports equilibrium that only a one side of player takes all responsibility of the past devia-

tion. Except this, two-sided replaceable case also supports equilibrium that two sides of players are equally share responsibility of the past deviation when the loss of deviation is too huge. In one-sided replacement case, asking share of such loss will not be accepted to innocent player side.

1.2.3 Renegotiation-Proofness

In repeated game, any equilibrium that contains more than a single phase cannot avoid the problem that whether each phase's equilibrium strategy is robust to coalition among players. That is, if players are open to renegotiation, they may try to avoid punishment phase which will cost not only the punished one but also punishers. For this matter, Farrell and Maskin (1989) considered the weak renegotiation-proofness (WRP) under the complete information. WRP requires that the expected payoff from each phase should not be strictly dominated by that from the other phases. They defined players' state as the phase and restricted equilibrium strategy depends on such state. For example, if some player can strictly improve payoff by changing to a new phase (state), then there should be another player who weakly prefer current phase to the new phase. Pearce (1989), Ray (1994), Benoit and Krishna (1993), Van Damme (1989) presented more general notion of renegotiation-proofness that defines each finite history as each single state. In Pearce (1989), it was shown that renegotiation-proof equilibrium's payoff can asymptotically achieve the payoff at the Pareto frontier. Ray (1994) expanded the result that such payoff may exists as singleton sets or a continuous set on the Pareto frontier. Especially when we focus on the perfect Bayesian equilibrium, we can exploit and expand the notion of Pearce (1989) into the incomplete information game. Different from Pearce (1989) and Ray (1994), this study presents the possibility that the incomplete information itself can restrict payoff from the Pareto frontier without signaling device and/or side payment.

	H	L
H	(x, x)	(z, w)
L	(w, z)	(y, y)

Table 1.1: A Payoff Structure of A Prisoner's Dilemma game

1.3 Model

1.3.1 Basic Game Environment

I consider a stage game g with two sides of players, $i = 1, 2$. The stage game is a symmetric simultaneous two-person prisoner's dilemma game of Figure 0. Payoff structure is assume to be $w > x > y > z$ and $2x > w + z$. Each side of individual players has type $T \in \{Normal, Stackelberg\}$. The type of each individual player will be fixed until the player quits the game. An individual player of the side i plays the game g with the player of the opponent side j ($\neq i$). To denote the opponent side j in terms of the opponent side of i, I will denote him/her side as -i. $t \in \{0, 1, 2, \cdots\}$ is a discrete calendar time period. A_i^t is a finite set of actions for the player of the side i at time t and $a_i^t \in A_i^t$ is a pure action of player of side i. Simply I denote them as the set of the player i's actions and the player i's action at t respectively. $A^t \equiv A_1^t \times A_2^t$ is a finite set of pure action profiles at t $a^t = (a_1^t, a_2^t)$. A Normal type (henceforth, type N) player has the set of actions $\{H, L\}$ and a Stackelberg (henceforth, type S) player has a singleton action set $\{L\}$. In other words, players can distinguish the type N player from the type S player only by the observation of action H. We call that H and L as a cooperative and defect action respectively.

In the repeated game, each player can be independently replaced with a strictly positive and fixed probability at the end of each period. Formally, $\lambda \in (0, 1)$ is a common probability of replacement of players. $\theta \in (0, 1)$ is a probability that newly replaced player is the type N. $\delta \in (0, 1]$ is a common time preference factor for future payoffs. I denote a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$ consists of the stage-wise game[3] g, the replacement rate λ, the distribution of the type N player θ, and the time preference δ. In this environment, I focus on the equilibrium

[3]I assume that the stage-wise game g contains the set of players' type.

which only relies on the individual players' incentive. Since I depict the situation in which the individual player is the decision maker, it is proper to consider only the individual players' incentive. For this reason, I handle actions, beliefs, and histories in the perspective of individual players at the corresponding period. So I simply denote 'a player of side i' as 'a player i.'

Before we define strategy in the next subsection, we need to consider a form of information that players have at the beginning of each period. I assume that players share all history of past actions and strategies of both sides. That is, even though players are replaceable over period, they can access to information about past actions and strategies played by their predecessors. I assume that players cannot observe replacement of player(s) of the opponent side and players are allowed to adopt any public signaling devices that gives information about two players' type. In other words, a full history of the past actions and strategies is the only available public information. Formally, $h^t = (a_1^1, a_2^1, \cdots, a_1^{t-1}, a_2^{t-1})$ is a *history* at the beginning of period t. $\mathscr{H}$ is the set of all finite histories. For each t, the set of all finite histories upto time t $\mathscr{H}^t$ is the subset of $\mathscr{H}$.

We can imagine the environment as following: consider each side of the game as the board of a firm, and two firms invest to the joint project. Individual players are individual traders for the firm hired by the board to delegate the firm's invenstment decision. The board may have different perspectives about the joint project, so that they hire a normal trader with θ or a aggresive trader with probability 1 - θ. Firms guarantee anonymity of staffs and individual traders make decisions based upon their attitude. Individual traders observe all the past histories of two firms, but they may not know the private information about the past traders. Replacement decision occurs independent of the individual traders' attitudes and actual outcomes of their decisions. For example, the personnel appointment process of the board, or a personal relation between the board member and the trader could be reasons for the replacement. However, the traders will not be fired because of the outcomes of their decisions or their attitude about the investment.

1.1. Strategy and Belief

I define a (public) *strategy* of the type N player i α_i: $\mathscr{H} \rightarrow [0, 1]$ as a probability to play action H. Since only type N players are able to play action H, the strategy is only available for type N player. $\mathscr{A}$ is the set of all strategies. For specific history h^t, I denote $\alpha_i^t \equiv \alpha_i(h^t)$. Similarly, I define the strategy of the player -i $\alpha_{-i}$and a strategy profile $\alpha = (\alpha_1, \alpha_2)$ respectively. $u_i(a^t)$ is a payoff function of the player i with respect to the action profile $a^t = (a_1^t, a_2^t)$. Then, $u_i(L, H) = w > u_i(H, H) = x > u_i(L, L) = y > u_i(H, L) = z$ and $2x > w + z$.

μ_i: $\mathscr{H} \rightarrow [0, 1]$ is the player i's *belief* that player -i is the type N player with respect to history. For specific history h^t, I denote $\mu_i^t \equiv \mu_i(h^t)$. μ_i^0 is the player i's initial belief about the player -i's type. Similarly, the player -i's belief and initial belief about the player i's type is denoted by μ_{-i} and μ_{-i}^0 respectively. Also I define a belief pair and belief pair at history h^t $\mu = (\mu_1, \mu_2)$ and $\mu^t = (\mu_1^t, \mu_2^t)$ respectively. I assume that the initial belief pair is a common knowledge and belief pair updated from the previous period is inherited as a common knowledge. That is, players in every period know the belief pair updated from the played actions and specified strategies in the previous stage.

Specifically, I assume that the belief pair follows the Bayesian belief update rule with respect to the strategy profile α. Without loss of generality, I consider the update rule (or the transition rule) for μ from h^{t-1} with respect to strategy α. Then, player i will update his/her belief to $\mu_i^t = (1\text{-}\lambda) + \lambda\theta = 1\text{-}\lambda(1\text{-}\theta)$ if he/she observes player -i's previous action $a_{-i}^{t-1} = H$. On the while, if the player i observes $a_{-i}^{t-1} = L$, then μ_i^t will be updated to

$$\mu_i^t = \lambda\theta + (1-\lambda)\left[\frac{\mu_i^{t-1}\left(1-\alpha_{-i}^{t-1}\right)}{\left(1-\mu_i^{t-1}\right)+\mu_i^{t-1}\left(1-\alpha_{-i}^{t-1}\right)}\right].$$

I can define the Bayesian belief update rule for the player -i's belief at h^{t-1} in a similar way. Notice that $a_{-i}^{t-1} = H$ immediately jumps up the player i's belief to $\mu_i^t = 1\text{-}\lambda(1\text{-}\theta)$, the highest level of belief that players can have. I denote *the maximum belief* $\overline{\mu} \equiv 1\text{-}\lambda(1\text{-}\theta)$. On the other side, $\mu_i^t = \lambda\theta$ is the lowest level of belief that players can have in case of $\alpha_{-i}^{t-1} = 1$ while he/she observes $a_{-i}^{t-1} = L$.

Similarly I denote *the minimum belief* $\underline{\mu} \equiv \lambda\theta$. Then, I can restrict a set of possible beliefs $M \equiv [\underline{\mu}, \overline{\mu}] \subset [0, 1]$.

$\mathscr{U}_1(\alpha,\mu; h^t)$ is a stage-wise expected payoff of the player 1 with respect to the strategy profile α and the belief pair μ with respect to history h^t. For simplicity, I denote $\mathscr{U}_1(\alpha,\mu; h^t) \equiv \mathscr{U}_1(\alpha^t,\mu^t)$. Then we have

$$\begin{aligned}
\mathscr{U}_1((H,\alpha_2^t),(\mu_1^t,\mu_2^t)) &= \mu_1^t \cdot \left(u_1(H,H)\cdot\alpha_2^t + u_1(H,L)\cdot(1-\alpha_2^t)\right) + (1-\mu_1^t)\cdot u_1(H,L) \ , \\
\mathscr{U}_1((L,\alpha_2^t),(\mu_1^t,\mu_2^t)) &= \mu_1^t \cdot \left(u_1(L,H)\cdot\alpha_2^t + u_1(L,L)\cdot(1-\alpha_2^t)\right) + (1-\mu_1^t)\cdot u_1(L,L) \ , \\
\mathscr{U}_1(\alpha^t,\mu^t) &= \alpha_1^t \cdot \mathscr{U}_1((H,\alpha_2),(\mu_1^t,\mu_2^t)) + (1-\alpha_1^t)\cdot \mathscr{U}_1((L,\alpha_{-i}),(\mu_1^t,\mu_2^t)) \ .
\end{aligned}$$

Player 2's expected payoff is similarly defined. From the stage-wise expected payoff, I define an expected continuation payoffs for the player i in the repeated game $G(g, \lambda, \theta, \delta)$ with respect to strategy α and belief pair μ at history h^t;

$$\begin{aligned}
V_i(\alpha,\mu;h^t) &= \mathscr{U}_i(\alpha^t,\mu^t) + \sum_{s=1}^{\infty}((1-\lambda)\delta)^s \left[\sum_{\forall h^{t+s}\, s.t.\, Pr_\alpha(h^{t+s}|h^t)>0} Pr_\alpha(h^{t+s}|h^t)\mathscr{U}_i(\alpha^{t+s},\mu^{t+s}) \right] \\
&= \mathscr{U}_i(\alpha^t,\mu^t) + (1-\lambda)\delta\mathbb{E}_{h^{t+1}\sim Pr_\alpha(h^{t+1}|h^t)}\left[V_i(\alpha,\mu;h^{t+1})\right],
\end{aligned}$$

where $Pr_\alpha(h^{t+s} \mid h^t)$ is an ex-ante transition probability from the history h^t to h^{t+s} with respect to the strategy profile α. From this formation, I define the (subgame) perfect Bayesian equilibrium (α, μ) as following.

Definition 1. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$. Suppose that G has a initial belief pair μ^0. Then, the perfect Bayesian equilibrium consists of a strategy α: $\mathscr{H} \to [0, 1]^2$ and a belief system μ: $\mathscr{H} \to M^2$ such that, $\forall i$, $\forall h^t$,

(i) $V_i(\alpha,\mu;h^t) \geq V_i((\alpha_i',\alpha_{-i}),\mu;h^t)\ \forall\alpha_i' \in \mathscr{A}$,

(ii) $\mu_i^{t+1} = Pr[T_{-i}^{t+1} = N \mid \alpha, \mu, h^t]\ \forall t$, where T_{-i}^{t+1} is the type of player -i at $t+1$.

1.4 Threshold Equilibrium

1.4.1 Threshold Strategy

A threshold strategy of the repeated game $G(g, \lambda, \theta, \delta)$ with respect to a threshold $\tau^* = (\tau_A^*, \tau_B^*)$ is the pure strategy (subgame) perfect Bayesian equilibrium equipped with the strategy α_{τ^*} and corresponding belief system μ_{τ^*}.

Definition 2. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$. I define a *Threshold Strategy* $\alpha_{\tau^*} : \mathscr{H} \rightarrow \{0, 1\}^2$ with respect to a threshold $\tau^* = (\tau_A^*, \tau_B^*)$ such that, for each $i = 1, 2$ and $\forall h^t$

$$\alpha_{i,\tau^*}(h^t) = \begin{cases} 1 & if\ (\mu_i^t, \mu_{-i}^t) \geq (\tau_A^*, \tau_B^*) \\ 0 & otherwise \end{cases}$$

I assume that threshold strategy includes a binary state process such that each belief pair is translated into the binary (Markovian) state and then each state is mapped into pure action strategy. Since I assumed stage game g to be symmetric prisoner's dilemma game, two sides of players will have symmetric form. Without loss of generality, I will abuse a notation $\alpha^* \equiv \alpha_{\tau^*}$ and $\mu_{\tau^*} \equiv \mu^*$. We can formally define the threshold equilibrium as following;

Definition 3. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$. Suppose that G has a initial belief pair $\mu^0 = (\mu_i^0, \mu_{-i}^0)$. Then, a *threshold equilibrium* (TEQ) with threshold $\tau^* = (\tau_i^*, \tau_{-i}^*)$ consists of a threshold strategy $\alpha^* : \mathscr{H} \rightarrow \{0, 1\}^2$ and a belief system $\mu^* : \mathscr{H} \rightarrow M^2$ such that, $\forall i$, $\forall h^t$,

(i) $V_i(\alpha^*, \mu^*; h^t) \geq V_i((\alpha_i', \alpha_{-i}^*), \mu^*; h^t)\ \forall \alpha_i' \in \mathscr{A}$,

(ii) $\mu_i^{*,t+1} = Pr[T_{-i}^{t+1} = N |\ \alpha^*, \mu^*, h^t]\ \forall t$.

Threshold equilibrium has an implication to renegotiation-proofness in the repeated game with the two-sided replacement environment. On a one hand, the environment contains strictly positive

amount of uncertainty about the opponent's type at every period. On the other hand, perfect monitoring on history of actions, players must hold public belief about each player's type and the type only can be verified by the cooperate action.

From a definition of threshold equilibrium, we induce an individual rationality (IR) condition. For simplicity, I will use a notation for the replacement rate weighted discount factor $\eta \equiv \delta(1\text{-}\lambda)$. I will also use a notiation for threshold equilibrium (α^*, τ^*) to specify a threshold τ^*[4] and simplify the expected payoff $V_i(\alpha^*; \mu^{*,t}) \equiv V_i(\alpha^*, \mu^*; h^t)$.

Definition 4 (Individual Rationality)**.** Fix $G(g,\delta,\lambda,\theta)$. Suppose a set of the threshold equilibria is nonempty. Then, a threshold equilibrium (α^*, τ^*) satisfies *individual rationality* if $V_i(\alpha^* | \mu^{*,t}) \geq \frac{b}{1-\eta}$ $\forall h^t$ and $\forall i$.

This individual rationality bounds the least expected payoff for threshold equilibrium payoff, which is came from a trivial equilibrium in which both players play L irrelevant to the history. With consideration of individual rationality, I redefine a non-trivial threshold equilibrium.

Definition 5 (Non-Trivial Threshold Equilibrium)**.** Fix $G(g,\delta,\lambda,\theta)$. Suppose a set of the threshold equilibria is nonempty. Then, a threshold equilibrium (α^*, τ^*) such that $\tau_A^*, \tau_B^* \leq \overline{\mu}$ resepctively and satisfies individual rationality is a $non-trivial\ threshold\ equilibrium$.

1.4.2 Markovian State Space

In this subsection, I construct a *Markovian state space* (henceforth, state space) that translates level of belief into discrete state.

Definition 6. Fix $G(g, \lambda, \theta, \delta)$. I define a $lower-side\ belief\ \mu_{HL}^k$ at k such that

[4]For belief system that follows Bayesian updating rule, I will implicitly assume that to be attached to threshold strategy α^*. In other words, a threshold equilibrium (α^*, τ^*) is equivalent to a PBE (α^*, μ^*) consists of threshold strategy $\alpha^* \equiv \alpha_{(\tau_i^*, \tau_{-i}^*)}$ and corresponding belief system $\mu^* \equiv \mu_{(\tau_i^*, \tau_{-i}^*)}$.

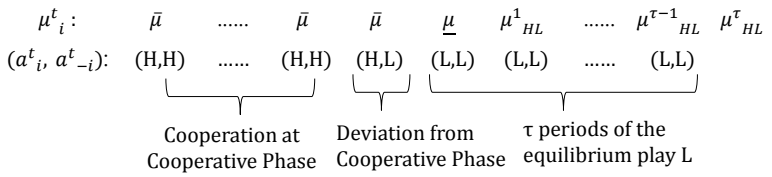


Figure 1.1: A lower-side belief

$$\mu_{HL}^k = \begin{bmatrix} \underline{\mu} & 1-\underline{\mu} \end{bmatrix} \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \\ \underline{\mu} & 1-\underline{\mu} \end{bmatrix}^{k-1} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix},$$

and a $upper-side\ belief\ \mu_{LH}^k$ at k such that

$$\mu_{LH}^k = \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \end{bmatrix} \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \\ \underline{\mu} & 1-\underline{\mu} \end{bmatrix}^{k-1} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix},$$

for any $t > 0$.

Moreover, I define $\mu_{HL}^0 = \underline{\mu}$ and $\mu_{LH}^0 = \overline{\mu}$.

A lower-side belief truncates k - period history of equilibrium play (L, L) starting from $\underline{\mu}$. Consider a history $h^{k+1} = ((L, L))^k$ from the initial belief $\underline{\mu}$ and assume that h^{k+1} is induced from the threshold strategy α. Then, the Bayesian update rule transits the player i's belief (after observing h^k) to μ_{HL}^k. For $k = 1$, the player i's belief after observing one period of (L, L) is $\underline{\mu}\cdot\overline{\mu}$ + (1 - $\underline{\mu}$)$\cdot\underline{\mu} = \mu_{HL}^1$. Similarly, for $k = 2$, observation of two period history $((L, L), (L, L))$ updates his/her belief to $\mu_{HL}^1\cdot\overline{\mu}$ + (1 - μ_{HL}^1)$\cdot$(1 - $\underline{\mu}$) = μ_{HL}^2. By extending this logic, I derive a lower-side belief μ_{HL}^k for k periods of equilibrium play (L, L). An upper-side belief similarly describes k periods history of equilibrium play (L, L) starting from $\overline{\mu}$. Following lemma describes boundary points and convergence property for lower- and upper-side beliefs.

Lemma 1. Fix $G(g, \lambda, \theta, \delta)$. Then, following holds:

(1) For any $k \geq 0$, $\underline{\mu} \leq \mu_{HL}^k \leq \mu_{HL}^{k+1} < \theta$ and $\theta < \mu_{LH}^{k+1} \leq \mu_{LH}^k \leq \overline{\mu}$,

(2) $\forall \varepsilon > 0$, $\exists T_{HL} < \infty$ such that $\forall k \geq T_{HL}$, $|\mu_{HL}^k - \theta| < \varepsilon$,

(3) $\forall \varepsilon > 0$, $\exists T_{LH} < \infty$ such that $\forall \tau \geq T_{LH}$, $|\mu_{LH}^k - \theta| < \varepsilon$

Proof. See the appendix. □

[Lemma 1] says that (1) lower- and upper-side belief is divided by a boundary point θ, and ((2), (3)) can arrive arbitrarily closely to the boundary point within finite periods of equilibrium play (L, L) from any side. From lemma 1, I define *the least arrival time* as following.

Definition 7. Fix $G(g, \lambda, \theta, \delta)$. $\overline{T}(\mu)$ is *the arrival time from* $\overline{\mu}$ *to* μ if $\mu_{LH}^{\overline{T}(\mu)+1} \leq \mu$, $\mu < \mu_{LH}^{\overline{T}(\mu)}$, and $\overline{T}(\mu) < \infty$. Similarly, $\underline{T}(\mu)$ is *the arrival time from* $\underline{\mu}$ *to* μ, if $\mu_{HL}^{\underline{T}(\mu)} \geq \mu$, $\mu > \mu_{HL}^{\underline{T}(\mu)-1}$, and $\underline{T}(\mu) < \infty$.

Moreover, I define $\underline{T}(\mu) = \infty$ if $\mu \geq \theta$ and $\overline{T}(\mu) = \infty$ if $\mu \leq \theta$ respectively.

The least arrival time $\underline{T}(\mu)$ counts the least number of periods such that μ_{HL}^k becomes higher than some belief level μ after that. So, for any $k \geq \underline{T}(\mu)$, $\mu_{HL}^k \geq \mu$ holds. Similarly, the arrival time $\overline{T}(\mu)$ counts the greatest number of periods such that μ_{LH}^k remains higher than μ until that time. So, for any $k \leq \overline{T}(\mu)$, $\mu \leq \mu_{LH}^k$ holds.

By using above definitions, we can find a partition on M according to the arrival times from $\underline{\mu}$ and $\overline{\mu}$ respectively. Formally,

Definition 8. Fix $G(g, \lambda, \theta, \delta)$. I define a *partition* on M $\mathbf{P_M} = \{\mu_{HL}^0 = \underline{\mu}, \mu_{HL}^1, \cdots, \theta, \cdots, \mu_{LH}^1, \mu_{LH}^0 = \overline{\mu}\}$ and a *Markovian state space* $\Omega \equiv P_M \times P_M$ which contains countably many states.

1.5 Equilibrium Characterization

In this section, I will characterize threshold equilibrium in terms of equilibrium path of plays. For this purpose, we will see that threshold equilibrium can be characterized by the level of threshold

beliefs. [Proposition 1] restricts forms of threshold equilibria with respect to their threshold belief pairs' least arrival time.

Proposition 1. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$ with a initial belief $\mu^0 = (\mu_1^0, \mu_2^0)$. Then, any threshold τ^* of the threshold equilibrium satisfies either

(1) $\tau_A^* \geq \theta$, $\tau_A^* \geq \tau_B^*$, and $\overline{T}(\tau_A^*) < \min\{\overline{T}(\tau_B^*) + 1, \underline{T}(\tau_B^*)\}$ or

(2) $\theta > \tau_A^* > \tau_B^* \geq \underline{\mu}$ and $\underline{T}(\tau_A^*) > \underline{T}(\tau_B^*)$

(3) $\theta > \tau_A^*$ and $\underline{T}(\tau_A^*) = \underline{T}(\tau_B^*)$ or

Proof. See the appendix.. □

[Proposition 1] restricts a pair of thresholds to three cases: (1) threshold for own belief is not only weakly higher than the threshold for the opponent's belief but also higher than the probability distribution of the type $N\ \theta$, (2) both are lower than θ, or (3) otherwise, at least they have the same level. The threshold pair $\tau = (\tau_A, \tau_B)$ where $\tau_A < \tau_B$ cannot support any threshold equilibrium. From the [proposition 1] that restricts the available formation of the threshold pairs, we can easily describe the formation of threshold equilibrium only by describing the the least arrival time of each threshold. Combining this result with the construction of Markovian state space allows us to find equivalence between two threshold equilibria in terms of the least arrival time of each threshold.

Corollary 1. Consider any two different thresholds (τ_A^*, τ_B^*) and $(\tau_A^\#, \tau_B^\#)$ that supports threshold equilibrium respectively. If

$$(\underline{T}(\tau_A^*), \overline{T}(\tau_A^*), \underline{T}(\tau_B^*), \overline{T}(\tau_B^*)) = (\underline{T}(\tau_A^\#), \overline{T}(\tau_A^\#), \underline{T}(\tau_B^\#), \overline{T}(\tau_B^\#)),$$

then induced threshold strategies α^* and $\alpha^\#$ follows exactly same on-the-equilibrium (and off-the-equilibrium paths) for any history h^t.

[Corollary 1] provides us a device we can exploit to characterize threshold equilibrium. I define a quartet $(\underline{T}(\tau_A^*), \overline{T}(\tau_A^*), \underline{T}(\tau_B^*), \overline{T}(\tau_B^*)) = (n, m, s, t)$ to characterize threshold equilibrium. In the

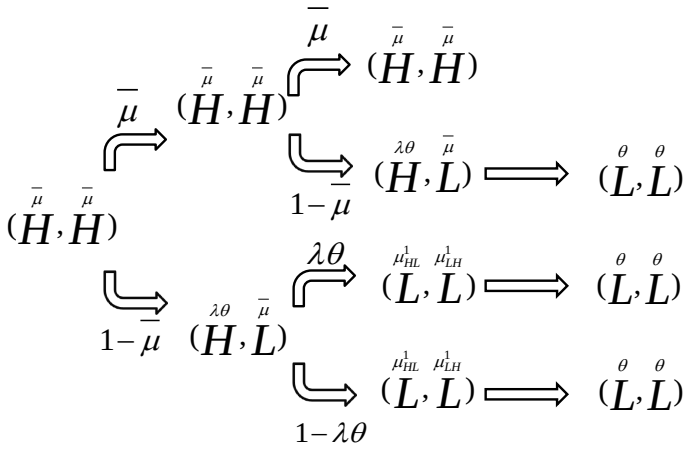


Figure 1.2: On-The-Equilirium path from $\alpha_{(\mu_i^*, \mu_{-i}^*)}$ at $(\mu_i, \mu_{-i}) = (\overline{\mu}, \overline{\mu})$

following sections, we will see that the quartet is enough to depict whole on-the-equilibrium paths.

1.5.1 Infinite Punishment Equilibrium

In this subsection, I describe infinite punishment equilibrium, the case corresponds to (1) of [Proposition 1]. [Theorem 1] formally describes behavioral pattern and threshold characterization of infinite punishment equilibrium.

Theorem 1. (Infinite Punishment Equilibrium)

Fix $G(g, \lambda, \theta, \delta)$. Suppose that there exists some threshold equilibrium (α^*, τ^*). Then, followings are equivalent;

(1) Threshold strategy α^* has two phases on the equilibrium path:

(Cooperation) Any type N players play H at t if $a^{t-1} = (H, H)$.

(Punishment) For any t, if there is any L played in h^t, then any player plays L at t .

(2) Threshold pair τ^* is characterized by the quartet (n, m, s, t) where $m < \infty$ and $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner < 1$.

Infinite punishment equilibrium depicts a case in which players set higher threshold for own threshold level than opponent's belief on oneself $(\tau_A^* \geq \tau_B^*)$, but players cannot set too high own threshold

level than opponent's threshold level ($\overline{T}(\tau_A^*) < \min\{\overline{T}(\tau_B^*)+1, \underline{T}(\tau_B^*)\}$). Consider the case with $(\mu_1, \mu_2) = (\overline{\mu}, \underline{\mu})$ where $m < \infty$ and $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner < 1$. As long as they follow equilibrium path, players will arrive to $(\mu_1, \mu_2) = (\mu_{LH}^m, \mu_{HL}^m)$ after m periods. Since $\mu_{HL}^m < \min\{\mu_{LH}^{t+1}, \mu_{HL}^s\}$ and $\mu_{HL}^m < \mu_{LH}^m$, both players would play L; even though player 1's belief level is high enough, player 2's belief level is not enough high so that she will not respond to cooperate action H. For further play of the equilibrium strategy, both players' belief will be depreciate to below μ_{LH}^m so that they would not play H at all. That is, (infinite) punishment phase describes in which (at least) one side of player cannot accumulate higher than threshold level before her opponent's belief depreciates below threshold level.

(1) of [Theorem 1] characterizes behavioral patterhn of infinite punishment strategy on the equilibrium path. That is, players play H only if when they observed (H, H) from the past history (cooperation phase) and plays L forever once they observe any L in any past history (punishment phase). (2) of [Theorem 1] characterizes symmetric thresholds that induce such infinite punishment strategy. Following proposition describes an incentive compatibility condition to support infinite punishment equilibrium.

Proposition 2. Suppose $G(g, \lambda, \theta, \delta)$ satisfies

$$V_{IC0}^{GT} \leq \left(z + \frac{\eta}{1-\eta} y\right) < min\{V_{IC1}^{GT}, V_{IC2}^{GT}\}$$

where $V_{IC0}^{GT} = \frac{\left(\overline{\mu} w + \frac{1-\overline{\mu}(1-\eta)}{1-\eta} y - \frac{x}{1-\eta\overline{\mu}}\right)}{1-\overline{\mu}}$, $V_{IC1}^{GT} = \frac{\left(\frac{1-\eta(1-\theta)}{1-\eta} y - z - \eta\theta w - \frac{\eta^2\theta\overline{\mu}}{1-\eta\overline{\mu}} x\right)}{\eta^2\theta\left((1-\overline{\mu}) + \overline{\mu}\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}\right)}$, $V_{IC2}^{GT} = \frac{\left(\frac{y}{1-\eta} - z - \frac{\mu_{LH}^1\eta}{1-\eta\overline{\mu}} x\right)}{\eta\left((1-\mu_{LH}^1) + \mu_{LH}^1\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}\right)}$.
Then, the game G has (at least one) threshold equilibrium characterized by (n, m, s, t) such that $m < \infty$ and $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner < 1$.

Condition of the [Proposition 2] is an incentive compatibility conditions under which the game has at least one threshold equilibrium characterized by (n, m, s, t) such that $m < \infty$ and $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner < 1$. $V_{IC0}^{GT} \leq z + \frac{\eta}{1-\eta} y$ is an incentive compatibility condition for a cooperation phase. That is, as long as the condition is satisfied, type N players observed (H, H) in the last period will contnitue

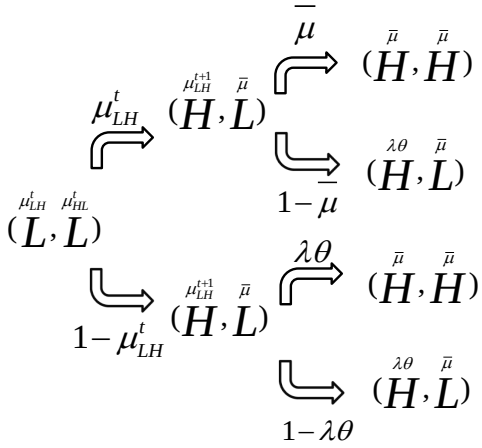


Figure 1.3: One-shot-deviation path from α^* at $(\mu_i, \mu_{-i}) = (\mu^t_{LH}, \mu^t_{HL})$

to play H to maintain the cooperation phase. $z + \frac{\eta}{1-\eta} y \leq V^{GT}_{IC1}$ is an incentive compatibility condition for a punishment phase which started from the opponent side's deviation. That is, once the condition is satisfied, the type N player would not deviate from playing L of punishment phase which started from the opponent side's play L. For such "innocent" side player, playing H will not be effective way of deviation. Even though her deviation "signals" her type and will induce her opponent to respond with following H, she will not be better off than keep playing L because of her cost of "signaling." $z + \frac{\eta}{1-\eta} y \leq V^{GT}_{IC2}$ is an incentive compatibility condition for punishment phase which started from her own side's deviation. That is, player would not deviate from playing L when herself or one of her predecessors is the first guilty player who broke the latest cooperation phase. For this reason, the deviation will be more effective signaling in that it is likely to responded by following H from both sides. V^{GT}_{IC2} restricts long-run benefit from deviation followed by possible cooperation phase. Once all these conditions are satisfied, we always have an infinite punishment equilibrium characterized by $(n, m, s, t) = (\infty, 0, \infty, 0)$. IC condition for the other infinite punishment equilibria will be always implied automatically from the [Proposition 3].

[Figure 1. 3] depicts a path of one-shot-deviation at (μ^t_{LH}, μ^t_{HL}) and $\tau^*_A = \mu^{t+1}_{LH}$ and $\theta > \tau^*_B > \mu^{t+1}_{HL}$. By exploiting one-shot-deviation principle, I compare expected payoffs from equilibrium strategy α^*_i and a one-shot-deviation strategy α_i'. Let α'_i to be one-shot-deviation strategy such that type N

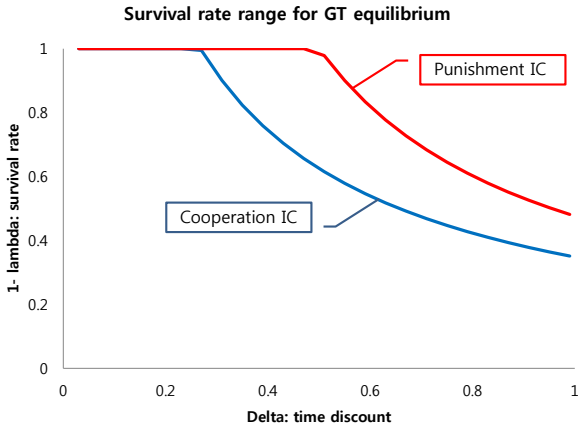
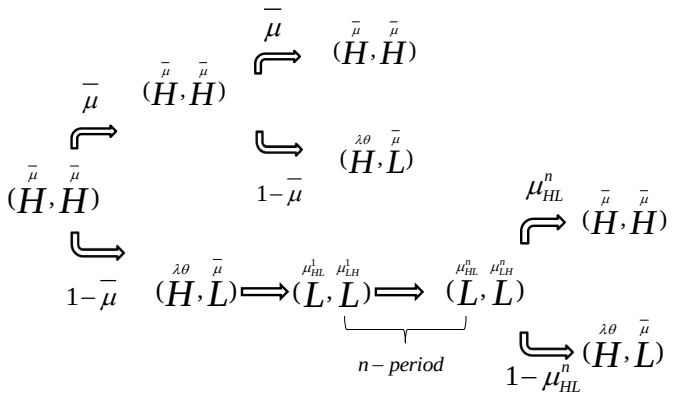


Figure 1.4: Ranges of IC conditions at the infinite punishmentequilibrium, where $\theta = 0.9$, $x = 2$, $y = 1$, $w = 2.5$, $z = -1$, y-axis: [1-Lambda] = 1-$\lambda \in (0, 1)$, x-axis : [Delta] = $\delta \in (0, 1)$

player i plays H only at period t and then return to α_i^*. Then playing H will immediately "signal" her type to her opponent (player -i) and μ_{-i}^{t+1} will jump to $\overline{\mu}$. As long as her opponent plays equilibrium strategy α_{-i}^*, the player i's belief μ_{LH}^{t+1} will remain on the threshold τ_A^*. As a result, player i has a belief pair $(\mu_{LH}^{t+1}, \overline{\mu}) \geq (\tau_A^*, \tau_B^*)$ and player -i has a belief pair $(\overline{\mu}, \mu_{LH}^{t+1}) \geq (\tau_A^*, \tau_B^*)$ which supposed to induce the cooperative play (H, H) at the next period as long as both players are remained as type N. That is, the player i considers a continuation payoff from deviation followed by cooperation phase. Interestingly, from some game $G(g, \lambda, \theta, \delta)$, we can find the case in which those conditions are all satisfied. See the [Figure 1-4].

[Figure 1. 4] depicts range of parameters δ and 1 - λ that satisfies [Proposition 3] at g ($x = 2$, $y = 1$, $w = 3$, $z = 0$) and $\theta = 0.9$. An area above blue line satisfies $V_{IC0} \leq z + \frac{\eta}{1-\eta} y$ and an area below red line satisfies condition $z + \frac{\eta}{1-\eta} y \leq \min\{V_{IC1}, V_{IC2}\}$. Especially, along the blue line, infinite punishment equilibrium is the only available threshold equilibrium. That is, within the boundary line of parameters, cooperation is supported as a part of equilibrium play and any deviation from cooperation phase will trigger an infinite play of (L, L).

Such infinite punishment equilibrium is supported because of uncertainty about opponent side player's type. That is, uncertainty about opponent's type restricts actual effectiveness of renegotiation to the case where opponent is the type N player. For this reason, the type N player cannot



(*Fig*.4) On-The-Equilibrium Path from $\alpha_{(\mu_i^*, \mu_{-i}^*)}$ at $(\overline{\mu}, \overline{\mu})$

Figure 1.5: Finite Punishment Equilibrium from $\alpha_{(\mu_i^*, \mu_{-i}^*)}$ at $(\mu_i, \mu_{-i}) = (\overline{\mu}, \overline{\mu})$

assure of renegotiation that agrees to play H during punishment phase unless they are "persuaded by action." However, low δ will not return high enough long-run payoff as much as the cost paid to show their actual type. As a result, even though both players are allowed to renegotiate to finish infinite punishment, they will not take short-run loss required to relaunch a new cooperation phase.

1.5.2 Finite Punishment Equilibrium

In this subsection, I characterize threshold equilibrium that allows cyclic interchange between cooperation phase and punishment phase. (3) of [Proposition 1] corresponds to this case. Similar to infinite punishment equilibrium, both players will play H during the cooperative phase as long as they observe (H, H) in the last period. Once either of side plays L, they will immediately turn to punishment phase and play (L, L) for finite periods. After finite periods of playing (L, L), type N players of both sides play cooperate action H. That is, type N players signal to each other irrelevant to causes to the latest punishment phase. After this signaling, both type N players will continue to play equilibrium strategy of cooperative phase or another punishment phase. We call it a finite punishment equilibrium.

Theorem 2. Fix $G(g, \lambda, \theta, \delta)$. Then, followings are equivalent;

(1) Threshold equilibrium (α^*, τ^*) consists of following phases:

(Cooperation) Any type N players play H at t if $a^{t-1} = (H, H)$.

(Punishment) Suppose some player (side) deviated from cooperation phase. Then, both players play L for n periods. Right after n periods of (L, L), any type N players plays H.

(2) Threshold τ^* of (α^*, τ^*) is characterized by (n, m, s, t) where $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner = \infty$ and $s = n < \infty$.

(3) There exists $n < \infty$ such that

(IC 1) $\mu_{HL}^n V_{IC_A}^{FN}(n) + (1 - \mu_{HL}^n) V_{IC_B}^{FN}(n) \geq \mu_{HL}^n V_{IC_C}^{FN}(n) + (1 - \mu_{HL}^n) V_{IC_D}^{FN}(n)$,

(IC 2) $z + \eta\left(\mu_{LH}^1 V_{IC_A}^{FN}(n) + (1-\mu_{LH}^1) V_{IC_B}^{FN}(n)\right) \leq \frac{1-\eta^n}{1-\eta} y + \eta^n\left(\mu_{LH}^n V_{IC_A}^{FN}(n) + (1-\mu_{LH}^n) V_{IC_B}^{FN}(n)\right)$,

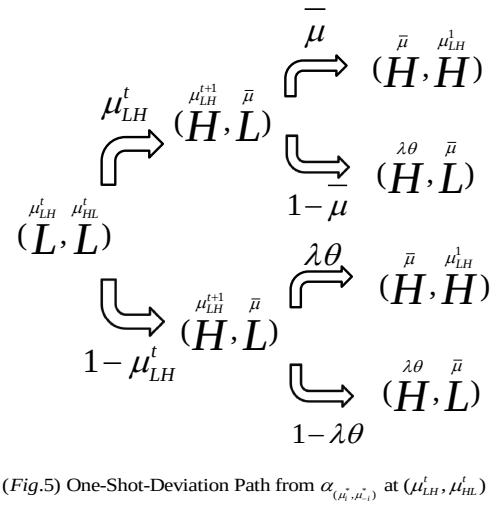
where $V_{IC_A}^{FN}(n) = \frac{x+\eta(1-\overline{\mu})V_{IC_B}^{FN}(n)}{1-\eta\overline{\mu}} = x + \eta V_i(\alpha^*|(\overline{\mu}, \overline{\mu}))$,

$$V_{IC_B}^{FN}(n) = \frac{\left(z+\eta\frac{1-\eta^n}{1-\eta}y+\eta^{n+1}\frac{\mu_{HL}^n \cdot x}{1-\eta\overline{\mu}}\right)}{1-\left(\mu_{HL}^n\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}+(1-\mu_{HL}^n)\right)\eta^{n+1}} = z + \eta V_i(\alpha^*|(\underline{\mu}, \overline{\mu}))$$

$$V_{IC_C}^{FN}(n) = w + \eta\frac{1-\eta^n}{1-\eta}y + \eta^{n+1}\left(\mu_{LH}^n \cdot V_{IC_A}^{IF}(n) + (1-\mu_{LH}^n)\cdot V_{IC_B}^{IF}(n)\right) = V_i(\alpha^*|(\overline{\mu},\underline{\mu})),$$

$$V_{IC_D}^{FN}(n) = \frac{1-\eta^{n+1}}{1-\eta}y + \eta^{n+1}\left(\mu_{HL}^n \cdot V_{IC_A}^{IF}(n) + (1-\mu_{HL}^n)\cdot V_{IC_B}^{IF}(n)\right) = V_i(\alpha^*|(\underline{\mu},\underline{\mu})).$$

[Theorem 2] characterizes threshold equilibrium that imposes finite n periods of punishment play L for one-sided deviation or for two-sided deviation. Each incentive compatibility condition characterizes a condition under which players will not deviate from equilibrium path of each phase. (IC 1) is incentive compatibility condition for cooperation phase. Left hand side of inequality depicts an expected continuation payoff from equilibrium path of coopeation phase (*i.e.*, type N player i plays H). Similarly, right hand side of inequality depicts expected payoff from using a one-shot deviation strategy (*i.e.*, type N player i plays L and then return to equilibrium strategy). For the case of two-sided deviation, the least level of belief for coopeation phase will be μ_{HL}^n. As long as (IC 1) holds at the least belief level μ_{HL}^n, any other belief above it will hold with strict inequality. [Figure 1-5] depicts on-the-equilibrium path starting from $(\overline{\mu}, \overline{\mu})$. (IC 2) is incentive compatibility condition for punishment phase started from own side deviation. Left hand side of (IC2) inequality depicts expected continuation payoff of the type N player i at $(\mu_i, \mu_{-i}) = (\overline{\mu}, \underline{\mu})$ when she plays the



(*Fig*.5) One-Shot-Deviation Path from $\alpha_{(\mu^*_t, \mu^*_{-t})}$ at (μ^t_{LH}, μ^t_{HL})

Figure 1.6: One-shot-deviation Path at (μ^t_{LH}, μ^t_{HL})

one-shot deviation action H during punishment phase. By sacrificing one period payoff as much as (y - z), player i can expect earlier return from current punishment phase at the next period. Right hand side of (IC 2) depicts expected continuation payoff of player i when she follows equilibrium strategy of punishment phase. Once (IC 2) holds for the worst case at which player has n remaining punishment period at higher belief level $\overline{\mu}$, the other cases at which she has shorter remaining punishment phase at μ^t_{LH} will automatically hold.

[Figure 1. 6] depicts one-shot-deviation path at punishment phase at belief level (μ^t_{LH}, μ^t_{HL}) and $\tau^*_A \leq \mu^{t+1}_{LH}$. In case that only one side deviated from the latest cooperation phase, cooperate action H from the guilty side will be effective to finish punishment phase early. After deviation, belief pair will be updated to (μ^{t+1}_{LH}, $\overline{\mu}$) and both players will agree to play H at the next period. On the other hand, deviation H from the innocent side will not be so effective to finish punishment phase. Action H from the innocent side signals only for her side's type, the guilty side may not be assured about whether the innocent side has enough belief about own her side. For this reason, players cannot agree to the innocent side's willingness for cooperation.

[Figure 1. 7] describes result of numerical example that satisfies IC conditions for finite punishment equilibrium in the plane of δ and 1 - λ. As in the example of infinite punishment equilibrium, we can find restriction on existence of finite punishment equilibrium. A lower bound for existence

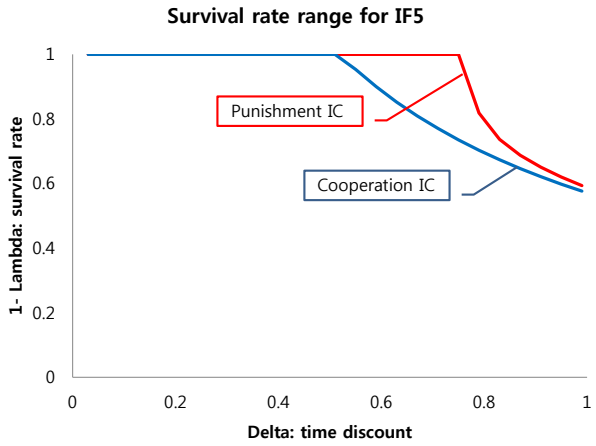
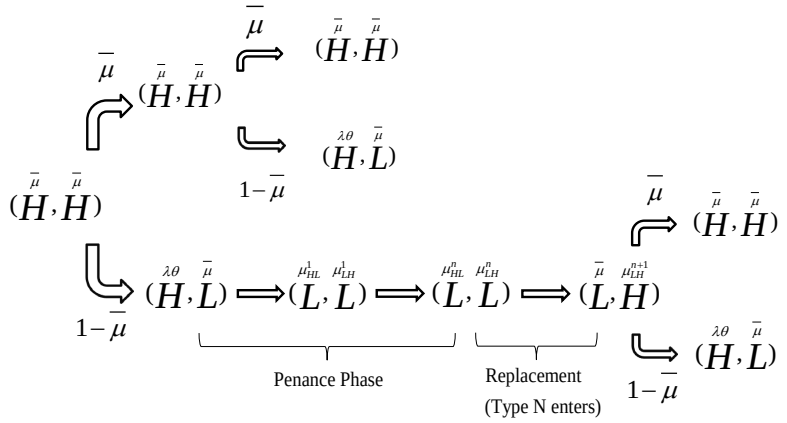


Figure 1.7: Ranges of IC conditions at the 5-period-punishment equilibrium at the game where θ = 0.9, x = 2, y = 1, w = 2.5, z = -1, y-axis: [1-Lambda] = 1-λ, x-axis : [Delta] = δ

of each finite punishment equilibrium (of certain n) corresponds to combination of parameters at which (IC 1) holds with equality. Similarly, a upper bound corresponds to (IC 2). Similar to infinite punishment equilibrium case, δ and λ play different roles at each IC condition. In (IC 1), δ and λ play a role of discount factor that encourages players to play cooperate action by enhancing continuation payoff of cooperation phase. That is, players are more willing to play cooperate action because they can can expect longer periods of cooperation phase with less future discount. On the other hand, in (IC 2), δ and λ discourages players to deviate from punishment phase. As higher δ and 1-λ induce higher expected payoff at cooperation phase, the guilty side player is more tempted to signal herself by playing H rather than keep following equilibrium path of punishment phase. In other words, players are more tempted to exploit her private information. For this reason, supporting threshold equilibrium with certain length of punishment phase requires parameters not to high.

1.5.3 Penance Equilibrium

In this subsection, I consider threshold equilibrium where players play different action depends on the past history of actions. Threshold formation (2) of [Proposition 1] corresponds to this



(*Fig*.7) On-The-Equilibrium Path from α^* at $(\bar{\mu}, \bar{\mu})$

Figure 1.8: Penance Equilibrium from α^* at $(\mu_i, \mu_{-i}) = (\overline{\mu}, \overline{\mu})$

case. Players play cooperate action H during cooperation phase. Different from previous threshold equilibria, behavioral pattern after observing first deviation from cooperation phase will be differed depends on history. First, suppose that only player i played defect action during cooperation phase. Then, players turn into a penance phase. During penance phase, type N player i (guilty side) immediately plays H and player -i (innocent side) will play L until she observes player i's action H. Once player i plays H, both players return to cooperation phase. Second, suppose that both players played defect action at the same time during cooperation phase. Then, they turn into punishment phase and play defect action L for finite periods. After punishment phase, both players return to cooperation phase. [Figure 1. 8] describes on-the-equilibrium path of penance equilibrium when player 1 is type N player.

Theorem 3. Fix $G(g, \lambda, \theta, \delta)$. Then, followings are equivalent;

(1) Threshold equilibrium (α^*, τ^*) consists of following phases:

(Cooperation) Any type N players play H at t if $a^{t-1} = (H, H)$.

(Punishment) Suppose both players play L at the same time during the cooperation phase. Then, both players play L for n periods and then return to cooperation phase.

(Penance) Suppose player i's side deviated during cooperation phase while -i did not. Then, any

type N player i plays H when she enters the game. Player -i plays L until she observes player i's action H. Once player i plays H for a period, then both players return to cooperation phase.

(2) Threshold equilibrium τ^* of (α^*, τ^*) is characterized by (n, m, s, t) where $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner = \infty$ and $0 = s < n < \infty$.

(3) There exists $n < \infty$ such that, for any i,

(IC 1) $\mu^n_{HL}(x + \eta V_i(\alpha^*|(\bar{\mu}, \bar{\mu}))) + (1 - \mu^n_{HL})\left(z + \eta V_i(\alpha|(\underline{\mu}, \bar{\mu}))\right) \geq$
$\mu^n_{HL}\left(w + \eta V_i(\alpha|(\bar{\mu}, \underline{\mu})\right) + (1 - \mu^n_{HL})\left(y + \eta V_i(\alpha|(\underline{\mu}, \underline{\mu}))\right)$,

(IC 2) $\frac{1-\eta^n}{1-\eta} y + \eta^n V_i(\alpha|\,(\mu^n_{HL}, \mu^n_{HL})) \geq z + \eta V_i(\alpha|\,(\mu^1_{HL}, \overline{\mu}))$,

(IC 3) $z + \eta V_i(\alpha|\,(\mu^{n+1}_{HL}, \overline{\mu})) \geq y + \eta\left[z + \eta V_i(\alpha|(\mu^{n+2}_{HL}, \underline{\mu}))\right]$.

Penance equilibrium allows players reacting differently to their opponent depends on the latest past deviation history. Finite punishment equilibrium or infinite punishment equilibrium served both players in the same way irrelevant to their past behavior. Such "equal" treatment was able to justify because of replacement that partially separates current player from the past deviation. However, penance equilibrium enforces any descendant player to take some responsibility for the past deviation so that this "unequal" treatment can be considered as unfair treat. By sacrificing such "equal" treatment, penance equilibrium allows players not to stay in punishment phase for unnecessarily long periods. That is, players from guilty side can legitimately avoid costly punishment as long as they are willing to pay for "signaling" themselves. As a result, existence of penance equilibrium is less likely to be restricted by high enough δ and 1 -λ.

(IC 1) and (IC 2) respectively corresponds to (IC 1) and (IC 2) condition from finite punishment equilibrium; that is, (IC 1) inequality characterizes the condition that any type N players play H as long as they observed (H, H) at the previous period. (IC 2) inequality characterizes the condition under which any type N player would not deviate from the punishment phase before they arrive at the threshold belief level.

(IC 3) inequality characterizes the condition under which players would not defer or avoid their current penance action (= H) to the next period even at the exact threshold belief. In (IC 3), defer-

ring their current penance to the next period incurs exchanges in the expected continuation payoff. When the player takes penance action immediately, she may have expected payoff from the cooperation phase and the punishment phase weighted with μ_{HL}^{n+1} and 1 - μ_{HL}^{n+1} respectively. By deferring one more period, players may avoid a cost for penance, but will to face the same problem with slightly higher belief level μ_{HL}^{n+2}. That is, the player compares the expected continuation payoff from the path after taking immediate penance action at the current belief to that from deferring penance action at the higher belief. Since μ_{HL}^{n+2} - μ_{HL}^{n+1} is nonnegative according to [Lemma 1], having a particular n that (IC 3) hold with (at least) equality is enough[5] to show that (IC 3) is satisfied at any $k > n$.

1.6 Discussion on Non-Trivial Threshold Equilibrium

1.6.1 Renegotiation-Proofness

In this subsection, I define renegotiation-proofness in the repeated game under the two-sided replacement environment. Basic notion of renegotiation-proofness may adopted from Pearce (1989). Consider a perfect Bayesian equilibrium (α, μ) and (α', μ') of $G(g, \lambda, \theta, \delta)$. We assume that there exists some h^t such that $\alpha^t \neq \alpha'^t$. We call (α, μ) (weakly) dominates (α', μ') if $\forall i$, $\forall h^t$, $V_i(\alpha', \mu'; h^t) \geq V_i(\alpha, \mu; h^t)$ and holds with strict inequality at least for one h^t and i.

Definition 9. (Pearce, 1989)

Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$. Suppose that G has a initial belief pair μ^0. A perfect Bayesian equilibrium (α, μ) is a renegotiation-proof equilibrium if (α, μ) is not dominated by any other perfect Bayesian equilibrium.

In case where $\lambda = 0$ and $\theta = 1$, renegotiation-proofness implies the weak renegotiation-proofness (WRP) of Farrell and Maskin (1989). In the notion of Farrell and Maskin, WRP requires the

[5] For more details, see the proof of [Proposition 5] (c) that shows existence of penance equilibrium.

equilibrium payoffs at any reachable state not to be dominated by any payoff at another reachable state. Without loss of generality, we can define the state as a function of history ω: $\mathscr{H} \to \Omega$, where Ω is the set of (countable) states and $\omega^t \equiv \omega(h^t)$. By replacing the history to the state, we can similarly define a (subgame) perfect equilibrium $(\alpha, \mu, \omega, \Omega)$.

Definition 10. Consider a repeated game under two-sided replacement $G(g, 0, 1, \delta)$[6]. Suppose that G has a initial belief pair μ^0. A perfect Bayesian equilibrium $(\alpha, \mu, \omega, \Omega)$ is weakly renegotiation-proof if for all w^t and w^s (such that w^t, $w^s \in \Omega$ respectively), $V_i(\alpha, \mu; w^t) > V_i(\alpha, \mu; w^s)$ for some i, then $V_{-i}(\alpha, \mu; w^s) \geq V_{-i}(\alpha, \mu; w^t)$.

Suppose that $(\alpha, \mu, \omega, \Omega)$ violates WRP at reachable state ω^t and ω^s respectively. Formally, for ω^t, $\omega^s \in \Omega$, $V_i(\alpha, \mu; w^t) > V_i(\alpha, \mu; w^s)$ and $V_{-i}(\alpha, \mu; w^t) > V_{-i}(\alpha, \mu; w^s)$. Then, consider an agreement from both players at state ω^s such that both players substitute their strategies to the truncated strategies from ω^t $\alpha|_{w^t}$ and induced truncated state $\omega|_{w^t}$. This substitution immediately improves both players' expected payoff to $V_i(\alpha, \mu; w^t)$ and $V_{-i}(\alpha, \mu; w^t)$ respectively, so that $(\alpha, \mu, \omega, \Omega)$ violates the renegotiation-proofness. However, we cannot assure that the notion of renegotiation-proofness is equivalent to WRP in general $G(g, \lambda, \theta, \delta)$ because there are cases in which WRP is violated while renegotiation-proofness is satisfied.

On the other hand, the notion of one-shot deviation (henceforth, OSD) principle that guarantees optimality of subgame perfect equilibrium strategy can be extended to show renegotiation-proofness. Different from the original OSD, renegotiation-proofness in our environment requires robustness to one-shot deviation from the both sides of players at any history.

Proposition 3. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$. Suppose that G has the initial belief pair μ^0. A pure strategy perfect Bayesian equilibrium (α, μ) is renegotiation-proof if and only if, for any h^t and μ^t, there is no pure strategy and corresponding belief system $(\hat{\alpha}, \hat{\mu})$ that agrees with (α, μ) except at the single stage t, and such that $(\hat{\alpha}, \hat{\mu})$

[6]Assuming $\lambda = 0$ and $\theta = 1$ as common knowledge fixes $\mu = (1, 1)$ for whole states and history. For that reason, in such environment, belief pair plays no role. We keep μ to be included to the definition of the perfect Bayesian equilibrium only for consistency of notation.

dominates (α, μ) conditional on history h^t and belief μ^t being reached.

Proof. ($\Rightarrow$) Renegotiation-proofness automatically satisfies robustness to any two-sided OSD strategies.

($\Leftarrow$) Suppose that a PBE (α, μ) is not renegotiation-proof in $G(g, \lambda, \theta, \delta)$. We will show that assuming robustness of (α, μ) to OSD for both sides contradicts with the non-renegotiation-proofness of (α, μ).

Suppose that (α, μ) is robust to OSD for both sides at any h^t and μ^t. Then, playing OSD strategy $\hat{\alpha}$ at any h^t will not be beneficial for both players. Let denote the following history from such deviation as $\hat{h}^{t+1}$ and induced belief $\hat{\mu}^{t+1}$. Then, for such history and induced belief, (α, μ) will be not be dominated by another OSD, say (α', μ'), by the assumption. By this way, concatenating any finite length of (pure strategy) deviation from (α, μ) cannot dominate (α, μ) itself at the last period of its deviation. By induction, any length of deviation from (α, μ) will not be better than (α, μ) at any history and belief. Since we assume that (α, μ) is robust to any OSD for both sides at any history and belief, (α, μ) is not dominated by any length of deviation starting from any history so that it is renegotiation-proof, which is contradictory to the assumption. □

Exploiting [Proposition 3] allows us to find renegotiation-proof equilibrium in the (nonempty) set of threshold equilibrium.

Proposition 4. Fix $G(g,\delta,\lambda,\theta)$. Suppose the set of threshold equilibria is nonempty. Then there exists a threshold equilibrium (α^*, τ^*) that is a (pure strategy) renegotiation-proof equilibrium.

Proof. See the Appendix. □

Any two-sided renegotiation, especially for deviation from the punishment phase, is based upon the actual "practice" of cooperate action, and such practice is also restricted by the fact that the player is the type N player. Threshold equilibrium considers this fact explicitly so that it assumes that players must confirm that not only herself has a high enough belief on her opponent but also she

knows that her opponenet also believes about her type with high enough belief. That is, the set of feasible non-trivial threshold equilibrium is already a restricted set of pure strategy equilibria that assumes actual practicibility of each action. From this fact, finding the best threshold equilibrium which is not dominated by any other feasible non-trivial threshold equilibria naturally allows the equilibrium to be also robust to any other pure strategy OSD.

To show renegotiation-proofness of a certain non-trivial threshold equilibrium (under the assumption the set of non-tirivial threshold equilibria is nonempty), I followed several steps in the proof. Following paragraph describes the procedural order of the proof in the appendix, and explains the purpose of each lemma. We begin with categorization of the Section 3. Each form of equilibrium is defined with corresponding threshold level and incentive compatibility conditions. [Lemma 3] and [Lemma 4] (of appendix) shows that each nonempty set of finite punishment and/or penance equilibria has a certain finite and/or penance equilibium, which is not dominated by any other threshold equilibria of the same form at each game respectively. Let denote such finite punishment and penance equilibrium as (α^{FN}, τ^{FN}) and (α^{PN}, τ^{PN}) respectively. [Lemma 5] compares those two threshold equilibria of different forms and shows that at least one of them will not be dominated by another one at any (belief) state. We may call them as the best feasible threshold equilibrium at the game. In [Lemma 6], I show that such best threshold equilibrium is also robust to any two-sided OSD at each game. Then, by applying the [Proposition 3], we can find that there exists a renegotiation-proof equilibrium in the nonempty set of threshold equilibria.

1.6.2 Existence and Uniqueness

In this subsection, I consider existence of threshold equilibrium on the range of δ and λ. From a proof of existence for each equilibrium, a uniqueness of penance equilibrium is implicitly derived. In the perspective that both parameters destermine individual player's expectation for future payoffs, they seems to have the similar role that controls the value of future payoffs. Especially when we consider weighted discount factor $\eta = \delta(1 - \lambda)$, this similarity of role is vivid. However,

two parameters play conflicting role related to a change of phase in threshold equilibrium, and this difference in their role results uniqueness of threshold equilibrium at the asymptotic environment where δ and 1 - λ are arbitrarily close to 1. [Proposition 5] describes existence of non-trivial threshold equilibrium. In a process of existence proof, we will see that the only threshold equilibrium that survives in the asymptotic environment is penance equilibrium. This result implicitly implies that we have only one threshold equilibrium in such environment.

Proposition 5. Fix g and θ. Then, there exists a nonempty set of parameters $(\delta_E, \lambda_E)(g, \theta) \subseteq (0, 1)^2$ such that a nonempty set of non-trivial threshold equilibria exists if $(\delta, \lambda) \in (\delta_E, \lambda_E)$ (g, θ). Also, there exists a nonempty set of parameters $(\delta_P, \lambda_P)(g, \theta) \subseteq (\delta_E, \lambda_E)$ (g, θ) such that the nonempty set of non-trivial equilibria only consists of penance equilibrium if $(\delta, \lambda) \in (\delta_P, \lambda_P)(g, \theta)$.

Proof. We will see that each threshold equilibrium's IC conditions do not hold at the same time.

(1) Infinite punishment equilibrium: Consider [Proposition 4]. In the formula of the proposition, we have

$$\left(z+\frac{\eta}{1-\eta}y\right) < min\{V_{IC1}^{GT}, V_{IC2}^{GT}\}$$

where $V_{IC1}^{GT} = \frac{\left(\frac{1-\eta(1-\theta)}{1-\eta}y-z-\eta\theta w-\frac{\eta^2\theta\overline{\mu}}{1-\eta\overline{\mu}}x\right)}{\eta^2\theta\left((1-\overline{\mu})+\overline{\mu}\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}\right)}$, $V_{IC2}^{GT} = \frac{\left(\frac{y}{1-\eta}-z-\frac{\mu_{LH}^1\eta}{1-\eta\overline{\mu}}x\right)}{\eta\left((1-\mu_{LH}^1)+\mu_{LH}^1\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}\right)}$. For simplicity of comparison, we turn them into the per-period average payoff by discounting them with (1-η). Then we have

$$((1-\eta)z+\eta y) < min\left\{(1-\eta)V_{IC1}^{GT}, (1-\eta)V_{IC2}^{GT}\right\}.$$

As $\eta\to1$, by exploiting L'Hopital's rule, we have (1-η)V_{IC1}^{GT} $\to$ θ(y - x) and (1-η)V_{IC2}^{GT} $\to$ (y - x) respectively. Since $x > y > 0$ and $\theta \in (0, 1)$, by the [Proposition 4], no infinite punishment equilibria exist as $\eta \to 1$ for any game.

(2) Finite punishment equilibrium: Consider (3) of [Theorem 2]. Without loss of generality, suppose that $\tau_i^* = \mu_{HL}^{n*}$, where $n^* \geq 2$. Discounting the expected payoffs to per-period average,

we have (1-η)$V^{FN}_{IC_A}(n) = \frac{(1-\eta)\left(a+\eta(1-\overline{\mu})V^{FN}_{IC_B}(n)\right)}{1-\eta\overline{\mu}}$ and (1-η)$V^{FN}_{IC_B}(n) = \frac{(1-\eta)\left(z+\eta\frac{1-\eta^n}{1-\eta}y+\eta^{n+1}\frac{\mu^n_{HL}\cdot x}{1-\eta\overline{\mu}}\right)}{1-\left(\mu^n_{HL}\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}+(1-\mu^n_{HL})\right)\eta^{n+1}}$ respectively. Then, (IC 1) is rearranged as following:

$$V_i(\alpha^*_\tau|(\tau^*_A,\tau^*_B)) \quad \geq \quad V_i((\alpha'_i,\alpha^*_{\tau,-i})|(\tau^*_A,\tau^*_B)) \qquad (IC\,1)$$

$$\Leftrightarrow \quad (1-\eta)\left\{\begin{array}{c}\left(\tau^*_A V^{FN}_{IC_A}(n*)+(1-\tau^*_A)V^{FN}_{IC_B}(n*)\right)+\\ -\eta^{n*+1}\left(\frac{V^{FN}_{IC_A}(n*)-V^{FN}_{IC_B}(n*)}{1-\eta}\right)\tau^*_A(\mu^{n*}_{LH}-\tau^*_A)\end{array}\right\} \quad \geq \frac{1-\eta^{n*}}{1-\eta}y+\tau^*_A(w-y).$$

When we fix n*, $\tau^*_A \downarrow 0$ as $\eta \to 1$. As a result, we have (1-η)$V^{FN}_{IC_B}(n) \to \frac{ny+z}{n+1}$ as $\eta \to 1$, so that we violate the IR condition. To avoid such an IR violation, we need to assume $\tau^*_A > v$, where $v \in [\underline{\mu}, \theta)$, as $\eta \to 1$. When we suppose that $\tau^*_A > v$ as $\eta \to 1$, $\left(\frac{\mu^{n*}_{LH}-\tau^*_A}{1-\eta}\right)\tau^*_A\left(V^{FN}_{IC_A}(n*)-V^{FN}_{IC_B}(n*)\right)$ should not diverge to infinity. This term represents a loss from "missed signaling" that the player would lose if the replaced opponent turned out to be a type S player. However, we have $\frac{V^{FN}_{IC_A}(n*)-V^{FN}_{IC_B}(n*)}{1-\eta}$ $\to \infty$ as $\eta \to 1$, letting $\tau^*_A(\mu^{n*}_{LH}-\tau^*_A) > 0$ (as $\eta \to 1$) will not prevent LHS of (IC 1) diverging to infinity. The only way we can avoid such divergence is allowing $\mu^{n*}_{LH} \downarrow \theta$ and $\tau^*_A \uparrow \theta$, so that $\left(\frac{\mu^{n*}_{LH}-\tau^*_A}{1-\eta}\right)\tau^*_A\left(V^{FN}_{IC_A}(n*)-V^{FN}_{IC_B}(n*)\right) < \infty$ for any $\eta \simeq 1$. However, such direction requires $n^* \uparrow \infty$ as $\eta \to 1$.

Now we consider (IC 2). Rearranging (IC 2), we have brings follwing inequality:

$$V_i(\alpha^*_\tau|(\overline{\mu},\underline{\mu})) \quad \geq \quad V_i((\alpha'_i,\alpha^*_{\tau,-i})|(\overline{\mu},\underline{\mu})) \qquad (IC\,2)$$

$$\Leftrightarrow \quad \frac{1-\eta^{n*}}{1-\eta}y+\left\{\begin{array}{c}\eta(\eta^{n*-1}-1)\left(z+\eta\underline{V}^{FP}(n*)\right)+\\ \underbrace{\eta\left(\eta^{n*-1}\mu^{n*}_{LH}-\mu^1_{LH}\right)\left((x+\eta\overline{V}^{FP}(n*))-(z+\eta\underline{V}^{FP}(n*))\right)}_{\to -(n*-1)x \text{ and } n*\uparrow\infty}\end{array}\right\} \quad \geq z.$$

LHS of (IC2) is consists of two items; the expected payoffs during the (n* periods of) punishment phase and the opprtunity cost incurred during the punishment phase. As $\eta \to 1$, LHS of (IC 2) converges to n^*y - (n* - 1)x so that (IC 2) is modified into following rearrangement;

$$n* \quad \leq \quad \frac{x-z}{x-y} \qquad (IC\,2')$$

For any fixed g, we may some $\eta \simeq 1$ such that corresponding $n*$ violates (IC 2'). As a result, for any fixed g and θ, we can find (λ, δ) such that (IC2) is violated.

(3) Penance equilibrium: Consider (3) of [Theorem 3]. Without loss of generality, suppose that τ_A^* = μ_{HL}^n, where $n > 1$. We can rearrange (IC 3):

$$(1-\eta)\left\{z+\eta\left[z+\eta\left(z+\eta V_B(n)\right)\right]+\eta\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta}\left(\left(x+\eta V_A(n)\right)-\left(z+\eta V_B(n)\right)\right)\right\} \quad \geq \quad y \quad (IC\,3)$$

where $V_A(n) = \frac{\overline{\mu}x+(1-\overline{\mu})z+\frac{(1-\overline{\mu})\eta\left(\underline{\mu}w+(1-\underline{\mu})y\right)}{1-(1-\underline{\mu})\eta}}{1-\eta\left(\overline{\mu}+(1-\overline{\mu})\frac{\underline{\mu}\eta}{1-(1-\underline{\mu})\eta}\right)}$, $V_B(n) = \frac{\underline{\mu}w+(1-\underline{\mu})y+\underline{\mu}\eta V_A(n)}{1-(1-\underline{\mu})\eta}$. From the fixed θ, $\eta \in [0, 1]$ we have $V_A(n) \geq V_B(n)$. We have limit payoff $\lim_{\lambda\to 0}\lim_{\delta\to 1}(1-\eta)V_B(n) = \frac{\theta x+(1-\theta)y}{2}$ and $\lim_{\lambda\to 0}\lim_{\delta\to 1}(1-\eta)V_A(n) = \frac{(1+\theta)x+(1-\theta)y}{2}$ respectively. Moreover, for fixed $\tau_A^* = \mu_{HL}^{n+1}$, $\lim_{\lambda\to 0}\lim_{\delta\to 1}\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta} = 2\mu_{HL}^{n+1} - \theta$. Then, we can simplify the LHS of (IC 3) as $\eta \to 1$ (with regarding to the order of limit):

$$\begin{aligned}(1-\eta)\left\{z+\eta\left[z+\eta\left(z+\eta V_B(n)\right)\right]+\eta\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta}\left(\left(x+\eta V_A(n)\right)-\left(z+\eta V_B(n)\right)\right)\right\}& \\ \to \frac{\theta x+(2-\theta)y}{2}+(2\tau_A^*-\theta)\left(\frac{x-y}{2}\right) = y+\tau_A^*(x-y) \qquad & > \quad y\end{aligned}$$

For another direction, we have $\lim_{\delta\to 1}\lim_{\lambda\to 0}(1-\eta)V_B(n) = y$, $\lim_{\delta\to 1}\lim_{\lambda\to 0}(1-\eta)V_A(n) = x$, and $\lim_{\delta\to 1}\lim_{\lambda\to 0}\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta}$ = $\mu_{HL}^{n+1} = \tau_A^*$ respectively. Similarly, the LHS of (IC 3) approaches to $y + \tau_A^*(x - y) > y$ (with regarding to the order of limit).

Now what we need to see is whether the limit around $\eta = 1$ from each direction does not fail (IC 3) condition. To confirm it, I consider the first derivative of the LHS for each parameter. I define $L(\delta, \lambda, n) = (1-\eta)\left\{z+\eta\left[z+\eta\left(z+\eta V_B(n)\right)\right]+\eta\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta}\left(\left(x+\eta V_A(n)\right)-\left(z+\eta V_B(n)\right)\right)\right\}$.

(Claim 1) Fix g and θ. For an arbitrarily small $\lambda > 0$, there exists $\delta_\lambda \in (0, 1)$ such that $\lim_{\delta\to 1}\lim_{\lambda\to 0} L(\delta_\lambda, \lambda, n) > y$.

Consider that $\lim_{\lambda\to 0}(1-\eta)V_A(n) = x$, $\lim_{\lambda\to 0}(1-\eta)V_B(n) = y$, and $\lim_{\lambda\to 0}\frac{\mu_{HL}^{n+1}-\eta\mu_{HL}^{n+2}}{1-\eta} = \mu_{HL}^{n+1} = \tau_A^*$. Then we have $\frac{\partial}{\partial\delta}\lim_{\lambda\to 0} L(\delta, \lambda, n) = 3(y\text{-}z) > 0$. Now we fix an arbitrarily small $\lambda* > 0$. For such $\lambda*$, we need

to find $\delta^* \in (0, 1)$ such that $d(L(\delta, \lambda^*, n), y + \tau_A^*(x - y)) < \tau_A^*(x - y)\ \forall\ \delta \in [\delta^*, 1)$. Recall that, at small enough λ^*, we have $\frac{\partial}{\partial \delta} L(\delta, \lambda, n) = 3(y - z) + o(\lambda^*)$, where $o(\lambda^*) \leq 0$. Then,

$$\begin{aligned} d\left(L(\delta, \lambda*, n), y + \tau_A^*(x-y)\right) &< \\ \left(3(y-z) + o(\lambda*)\right)(1-\delta*) &\leq \\ 3(y-z)(1-\delta*) &\leq \\ \tau_A^*(x-y) & \end{aligned}$$

Then, having $\delta^* > 1 - \frac{\tau_A^*(x-y)}{3(y-z)}$ will preserve $L(\delta, \lambda^*, n) > b\ \forall\ \delta \in [\delta^*, 1)$.

(Claim 2) Fix g and θ. For an arbitrarily small $\lambda > 0$, there exists $\delta_\lambda \in (0, 1)$ such that $\lim_{\lambda \to 0} \lim_{\delta \to 1} L(\delta_\lambda, \lambda, n) > y$.

Consider that $\lim_{\delta \to 1}(1-\eta)V_A(n) = \frac{\lambda\left((1-(1-\theta)\lambda)x + (1-\theta)\lambda z + \frac{(1-\theta)\lambda(1-\lambda)(\lambda\theta w + (1-\lambda\theta)y)}{\lambda(1+\theta-\lambda\theta)}\right)}{1-(1-\lambda)\left((1-(1-\theta)\lambda)(1-\theta)\lambda + \frac{\lambda\theta(1-\theta)}{\lambda(1+\theta-\lambda\theta)}\right)}$ and $\lim_{\lambda \to 0}[\frac{\partial}{\partial \lambda} \lim_{\delta \to 1}(1 - \eta)V_A(n)] = 0$ by L'hopital's rule. Similarly, since $V_B(n) = \frac{\underline{\mu} w + (1-\underline{\mu})y + \underline{\mu}\eta V_A(n)}{1-(1-\underline{\mu})\eta}$, we also have $\lim_{\lambda \downarrow 0}$ $[\frac{\partial}{\partial \lambda} \lim_{\delta \to 1}(1-\eta)V_B(n)] = -\frac{\theta(b + \theta\frac{(1+\theta)x + (1-\theta)y}{2}) + (1+\theta)\theta(w-y)}{(1+\theta)^2} < 0$.

I define $M_{HL}^n(\lambda, \delta) = \frac{\mu_{HL}^{n+1} - \eta\mu_{HL}^{n+2}}{1-\eta}$. Then $\lim_{\delta \to 1} M_{HL}^n(\lambda, \delta) = (2 - \lambda)\mu_{HL}^{n+1} - (1 - \lambda)\theta$. With fixed $\mu_{HL}^{n+1} = \tau_A^*$, we have $\lim_{\lambda \downarrow 0}\left[\frac{\partial}{\partial \lambda} \lim_{\delta \to 1} M_{HL}^n(\lambda, \delta)\right] = \mu_{HL}^{n+1} - \theta$. With $\frac{\theta x + (2-\theta)y}{2} + (2\tau_A^* - \theta)(x - z + \frac{x-y}{2}) > y$, we have

$$\begin{aligned} \lim_{\lambda \downarrow 0}\left(\frac{\partial}{\partial \lambda} \lim_{\delta \to 1} L(\delta, \lambda, n)\right) &= \\ & 3\left[-z + \tfrac{\theta x + (2-\theta)y}{2}\right] - (\theta + 2(n+1))\tfrac{x-y}{2} - \theta(x-z) - (1+\theta)\left[\tfrac{(1+\theta)\theta(w-y) + \theta(y - \frac{(1+\theta)x + (1-\theta)y}{2})}{(1+\theta)^2}\right] \\ &< 0. \end{aligned}$$

Since $\lim_{\lambda \to 0} \lim_{\delta \to 1} L(\delta, \lambda, n) > y$, showing the negative sign will be enough.

From the result of (Claim 1) and (Claim 2), we can see that the LHS of (IC 3) increases toward b in both direction, which implies the (IC 3) will not fail around $\eta = 1$. □

[Proposition 5] shows that IC conditions of infinite and finite punishment equilibrium will not hold as η approaches to 1 close enough. Such nonexistence of threshold equilibrium came from

different role of replacement. In infinite punishment or finite punishment equilibrium, as $\eta \to 1$, the type N player of guilty side is temped enough to play cooperate action to finish lengthy punishment phase. That is, high 1 - λ implies that current cooperation phase is more likely to continue in the next period. For this reason, as long as the game has higher 1 - λ, cooperation phase becomes easier to maintain. Similarly, as δ grows higher, players evaluate their future payoffs more highly, so cooperation phase becomes easier to be kept. However, high expected continuation payoff at cooperation phase also increases temptation to deviate from punishment phase. For finite punishment equilibrium, type N player of guilty side still can have high enough belief for her opponent's type because her belief depreciates slowly with low λ. As an example, consider the situation in which the guilty side player is replaced by the type N player at the first period of the punishment phase. In this case, the (newly replaced) guilty side player still have $\overline{\mu}$, and she expects that her belief about her opponent type will be depreciated to θ as the punishment phase proceeds. Definitely such depreciation of belief implies that of her expected continuation payoff. That is, the guilty side player confronts two ways of opportunity cost from her own replacement and from her opponent's replacement to the type S. These two pressures from the replacement forces the (newly replaced) guilty side player to take a fixed amount of cost to finish punishment phase earlier. Even though players are not in punishment phase, expecting such "internal inconsistency"[7] can be a reason to deny the equilibrium.

On the other hand, penance equilibrium is free from deviation at (finite or infinite periods of) punishment phase since players are always able to finish a series of defect actions as long as they have high enough belief. The problem that penance equilibrium confronts is another aspect of signaling; they may be tempted to defer penance action with consideration that their opponent's type will not changed in a short time. Low λ implies the type S player will stay at the game for a long enough time so that their belief about the opponent's type will grow slowly. For this reason,

[7]Pearce (1989) mentioned about the internal consistency of the renegotiation-proof equilibrium. His concept based on the fact that the renogotiation-proofness should be supported by the optimality of the equilibrium path at any state of the game. Even though current environment does involves replacement of players, with strictly positive probability that players can survive to another phase, the same logic of optimality of each path is required to support the renegotiation-proofness.

the type N player in the penance side whose belief is higher than threshold level but lower than θ may willingly wait more periods in penance phase to earn higher belief about her opponent's type. In (IC 3), the type N player of guilty side may compare her expected continuation payoff from immediate penance and deferred penance (OSD strategy) in the next period. In such situation, as $\eta \to 1$, the expected continuation payoff from an original equilibrium path and that from deferred penance become similar while immediate benefit from avoiding immediate penance becomes relatively small. At the same time, marginal increment of expected continuation payoff from deferring penance action becomes smaller as $\lambda \to 0$. As a result, penance equilibrium payoff will remain as a available one even at the asymptotic case while the other threshold equilibria fail in such case. This refinement of equilibrium was also considered by Hillas (1994) and McLennan (1985), but they considered similar notion of refinement only in the complete information environment.

[Figure 1. 9] shows that infinite and finite punishment equilibrium may not exist even with low enough λ (i.e., high enough survival rate) and high enough δ. In the figure, multiple threshold equilibria can exist between red and blue solid curves. In the [Figure 1. 9], the blue line represents combination of the least 1 -λ and δ where infinite punishment equilibrium's incentive compatibility at the cooperation phase holds with equality. The red line similarly represents combination of the highest 1 -λ and δ where finite punishment equilibrium's (with some $n > 1$) incentive compatibility hold with equality. In the region where $\delta \in$[0.3, 0.78], existence region monotonously increases as δ increases. However, at the level of $\delta > 0.78$ and high enough 1 -λ, the game does not have any infinite or finite punishment equilibria. Similar phenomenon is found around for 1 -λ ≈ 1.

For a matter of existence itself, there is a possibility of multiple equilibria in some region. Consider 1 -λ = 0.5 at the [Figure 1. 9]. At that point, some finite punishment equilibrium and infinite equilibrium are both supported as threshold equilibrium. On the other hand, at the point of 1 -λ = 0.7 and $\delta \simeq 1$, only finite (5-period) punishment equilibrium can be found as an available threshold equilibrium which will achieve the best Pareto efficiency. Similarly at a point of 1 -λ = 0.4, only infinite punishment equilibrium is supported as available threshold equilibrium.

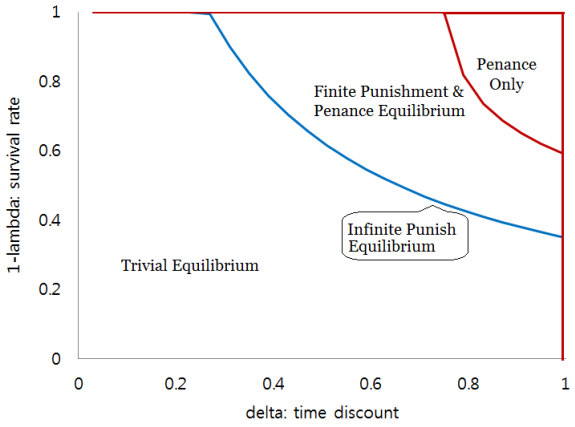


Figure 1.9: The map for the renegotiation-proof threshold equilibrium at the game where $\theta = 0.9$, $x = 2$, $y = 1$, $w = 2.5$, $z = -1$, y-axis: [1-Lambda] = 1-λ, x-axis : [Delta] = δ

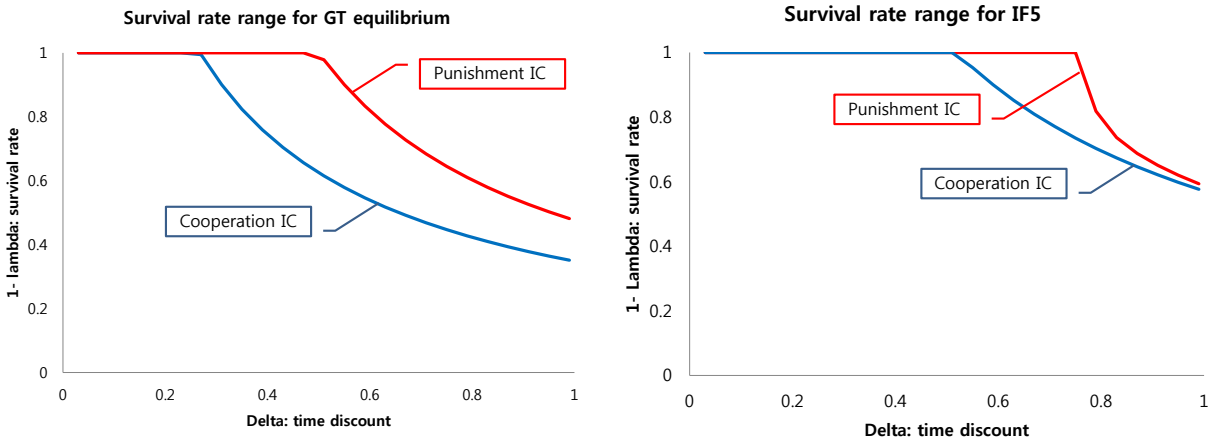


Figure 1.10: Left : ranges of IC conditions at infinite punishment equilibrium , Right : ranges of IC conditions at the finite ($n = 5$) punishment equilibrium at the game where $\theta = 0.9$,$a = 2$, $b = 1$, $c = 2.5$, $d = -1$, y-axis: [1-Lambda] = 1-λ, x-axis : [Delta] = δ

1.6.3 Non-monotonicity of the Equilibrium Payoffs

Conjecture 1. Fix g and θ. Suppose the set of the threshold equilibria is nonempty. Then, renegotiation-proof equilibrium has a nonempty set of (averaged per-period) equilibrium payoffs that is either a set of countable singleton payoffs or a subset of Pareto frontier of g.

In Ray (1994) and Pearce (1989), it is shown that the renegotiation-proof equilibrium payoff either forms singleton sets or locates at the Pareto frontier of the one-shot game. They call such a property as "non-monotonicity" of payoffs. Even in a two-sided player replaceable case, we can expect renegotiation-proof payoffs similarly preserve such non-monotonicity. Since we could always find a renegotiation-proof equilibrium in a nonempty set of non-trivial threshold equilibria, we can take advantage of this. For any fixed game, the expected continuation payoff of threshold equilibrium consists of $V_i(\alpha|\ (\overline{\mu}, \overline{\mu}))$, $V_i(\alpha|\ (\underline{\mu}, \underline{\mu}))$, $V_i(\alpha|\ (\underline{\mu}, \overline{\mu}))$, and $V_i(\alpha|\ (\overline{\mu}, \underline{\mu}))$. For infinite punishment equilibrium, we can equivalently put all the other equilibrium payoff except $V_i(\alpha|\ (\overline{\mu}, \overline{\mu}))$ to $\frac{b}{1-\eta}$, which greatly simplifies the formation of equilibrium payoffs. For finite-punishment equilibrium, we can put $V_i(\alpha|\ (\underline{\mu}, \underline{\mu}))$ and $V_i(\alpha|\ (\underline{\mu}, \overline{\mu}))$ to be equal. When we simplifies $x + \eta V_i(\alpha|\ (\overline{\mu}, \overline{\mu})) = V^{IF}_{IC_A}(n) = \frac{x+\eta(1-\overline{\mu})V^{IF}_{IC_B}(n)}{1-\eta\overline{\mu}}$ and $z + \eta V_i(\alpha|\ (\underline{\mu}, \overline{\mu})) = V^{IF}_{IC_B}(n) = \frac{\left(z+\eta\frac{1-\eta^n}{1-\eta}y+\eta^{n+1}\frac{\mu^n_{HL}\cdot x}{1-\eta\overline{\mu}}\right)}{1-\left(\mu^n_{HL}\frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}}+(1-\mu^n_{HL})\right)\eta^{n+1}}$ respectively, the equilibrium payoff at each state can be recursively calculated as weighted average of $V^{FN}_{IC_A}(n)$ and $V^{FN}_{IC_B}(n)$. Moreover, each history of the game is characterized by each discrete singleton belief state, and combination of the belief state and singleton payoffs will bring a singleton payoff. It implies that the expected continuation at each history will be characterized by a singleton point of the Pareto payoff set. For penance equilibrium, we can similarly consider that each expected continuation payoff of history will be characterized by a singleton point of the Pareto payoff set and each history will be characterized by discrete belief state.

1.7 Conclusion

This paper studied non-trivial threshold equilibrium in the repeated game with two-sided replacement of players. Characterization of threshold equilibrium resulted three behavioral patterns on the equilibrium path. Infinite punishment equilibrium does not allow any player to return from punishment phase after any deviation from cooperation phase. Finite punishment equilibrium consists of cooperation phase and finite periods of punishment phase. In finite punishment equilibrium, players will allows return from punishment phase once both sides equally stay in the punishment phase. Penance equilibrium distinguishes (finite) punishment phase and (indefinite) penance phase based upon initial deviation pattern. It enforces any guilty side player to pay the cost for the past deviation to return to cooperation phase.

Different from one-sided replaceable player environment or long-run players with type change, two-sided player replacement takes different way of sharing a risk of type uncertainty. In the intermediate level of the replacement rate, players share their risk equally in infinite or finite punishment equilibrium. However, as the replacement rate becomes small enough, players rather ask a one guilty side player to take all responsibility of the past deviation regardless of her actual guiltiness. This result implies uniqueness of penance equilibrium in the asymptotic case where the replacement rate and the time discount rate are small enough.

Moreover, a set of non-trivial threshold equilibria have at least one pure strategy renegotiation-proof equilibrium. Since players are randomly replaced, players cannot predict the timing that they will enter the game. For this reason, player who entered the game during the punishment or penance phase may be temped to renegotiate with her opponent to change current phase. As long as we can find a nonempty set of non-trivial threshold equilibria, it will always contain a pure strategy renegotiation-proof equilibrium. This result implies that players will find some threshold equilibrium as non-dominated choice for any other (pure strategy) alternative equilibrium.

For further research, we can consider several ways to make it more "realistic" setting. Having more than two players in a general non-cooperative game will be the most natural way we can extend

this study. As the number of players increases, players may find a different way to share risk of uncertainty when they confronts temptation of myopic deviation.

1.8 Appendix

1.8.1 Proof of Lemma 1

Lemma 1. Consider a repeated game under the two-sided replacement $G(g, \lambda, \theta, \delta)$. Then, following holds:

(1) for any $\tau \geq 0$, $\underline{\mu} < \mu_{HL}^{\tau} \leq \mu_{HL}^{\tau+1} \leq \theta$ and $\theta < \mu_{LH}^{\tau+1} \leq \mu_{LH}^{\tau} \leq \overline{\mu}$,

(2) $\forall \varepsilon > 0$, $\exists T_{HL} < \infty$ such that $\forall \tau \geq T_{HL}$, $|\mu_{HL}^{\tau} - \theta| < \varepsilon$,

(3) $\forall \varepsilon > 0$, $\exists T_{LH} < \infty$ such that $\forall \tau \geq T_{LH}$, $|\mu_{LH}^{\tau} - \theta| < \varepsilon$

Proof. For μ_{HL}^{τ}, showing $\underline{\mu} < \mu_{HL}^{\tau} < \mu_{HL}^{\tau+1} \leq \theta$ for any $\tau \geq 0$ automatically satisfies (2).

(i) By using induction, we can show that $\underline{\mu} < \mu_{HL}^{\tau} < \mu_{HL}^{\tau+1}$ for any $\tau \geq 0$. Since $\mu_{HL}^{0} = \underline{\mu}$ and $\mu_{HL}^{1} = \overline{\mu}\cdot\underline{\mu} + (1-\underline{\mu})\cdot\underline{\mu}$, $\mu_{HL}^{0} < \mu_{HL}^{1}$ holds. Now assume that $\mu_{HL}^{\tau-1} < \mu_{HL}^{\tau}$ for some $\tau \geq 2$. Without loss of generality, denote $\begin{bmatrix} \underline{\mu} & 1-\underline{\mu} \end{bmatrix} \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \\ \underline{\mu} & 1-\underline{\mu} \end{bmatrix}^{\tau-2} = \begin{bmatrix} A_{HL}^{\tau-1} & 1-A_{HL}^{\tau-1} \end{bmatrix}$ and $\begin{bmatrix} \underline{\mu} & 1-\underline{\mu} \end{bmatrix} \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \\ \underline{\mu} & 1-\underline{\mu} \end{bmatrix}^{\tau-1} = \begin{bmatrix} A_{HL}^{\tau} & 1-A_{HL}^{\tau} \end{bmatrix}$ respectively. From the formulation of $\mu_{HL}^{\tau-1}$ and μ_{HL}^{τ}, $\mu_{HL}^{\tau-1} < \mu_{HL}^{\tau}$ holds if and only if $A_{HL}^{\tau-1} < A_{HL}^{\tau}$. Moreover, $A_{HL}^{\tau-1} < A_{HL}^{\tau} = A_{HL}^{\tau-1}\cdot\overline{\mu} + (1 - A_{HL}^{\tau-1})\cdot\underline{\mu}$ implies $A_{HL}^{\tau-1} < \overline{\mu}$. Similarly, we have $\mu_{HL}^{\tau} = \begin{bmatrix} A_{HL}^{\tau} & 1-A_{HL}^{\tau} \end{bmatrix} \begin{bmatrix} \overline{\mu} & 1-\overline{\mu} \\ \underline{\mu} & 1-\underline{\mu} \end{bmatrix} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix}$ $\equiv \begin{bmatrix} A_{HL}^{\tau+1} & 1-A_{HL}^{\tau+1} \end{bmatrix} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix}$. Then, the assumption $A_{HL}^{\tau-1} < A_{HL}^{\tau}$ implies $A_{HL}^{\tau} = A_{HL}^{\tau-1}\cdot\overline{\mu} + (1 - A_{HL}^{\tau-1})\cdot\underline{\mu} < A_{HL}^{\tau+1} = A_{HL}^{\tau}\cdot\overline{\mu} + (1 - A_{HL}^{\tau})\cdot\underline{\mu}$. So, we have $\mu_{HL}^{\tau} < \mu_{HL}^{\tau+1}$.

(ii) By using induction, we can show that $\underline{\mu} < \mu_{HL}^{\tau} < \theta$ for any $\tau \geq 0$. For $\tau = 0$, $\mu_{HL}^{0} = \underline{\mu} < \theta$. Now assume that $\mu_{HL}^{\tau-1} < \theta$ holds. As in the (i), we can decompose $\mu_{HL}^{\tau-1} = \begin{bmatrix} A_{HL}^{\tau-1} & 1-A_{HL}^{\tau-1} \end{bmatrix} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix}$

$= A_{HL}^{\tau-1}\cdot\overline{\mu} + (1 - A_{HL}^{\tau-1})\cdot\underline{\mu}$ so that $A_{HL}^{\tau-1} < \theta = \theta\cdot\overline{\mu} + (1 - \theta)\cdot\underline{\mu}$. Since $A_{HL}^{\tau} = A_{HL}^{\tau-1}\cdot\overline{\mu} + (1 - A_{HL}^{\tau-1})\cdot\underline{\mu} < \theta$, $\mu_{HL}^{\tau} = \begin{bmatrix} A_{HL}^{\tau} & 1 - A_{HL}^{\tau} \end{bmatrix} \begin{bmatrix} \overline{\mu} \\ \underline{\mu} \end{bmatrix} = A_{HL}^{\tau}\cdot\overline{\mu} + (1 - A_{HL}^{\tau})\cdot\underline{\mu} < \theta$.

For μ_{LH}^{τ} case, we can exploit similar proof. □

1.8.2 Proof of Proposition 1

Proposition 1. Consider a repeated game under two-sided replacement $G(g, \lambda, \theta, \delta)$ with a initial belief $\Gamma^0 = (\mu_i^0, \mu_{-i}^0)$. Then, any threshold (τ_i^*, τ_{-i}^*) of the TEQ satisfies either

(1) $\tau_i^* \geq \theta$, $\tau_i^* \geq \tau_{-i}^*$, and $\overline{T}(\tau_i^*) < \min\{\overline{T}(\tau_{-i}^*) + 1, \underline{T}(\tau_{-i}^*)\}$ or

(2) $\theta > \tau_i^* > \tau_{-i}^* \geq \underline{\mu}$ and $\underline{T}(\tau_i^*) > \underline{T}(\tau_{-i}^*)$

(3) $\theta > \mu_i^*$ and $\underline{T}(\tau_i^*) = \underline{T}(\tau_{-i}^*)$ or

Proof.

(A) Consider the case $\tau_{-i}^* > \tau_i^* \geq \theta$ where $\overline{T}(\tau_i^*) < \overline{T}(\tau_{-i}^*)$. Define $\eta \equiv \delta(1\text{-}\lambda)$. Without loss of generality, I assume that $\tau_i^* = \mu_{LH}^t$ and $\tau_{-i}^* = \mu_{LH}^n$ where $t > n$. To have (α^*, τ^*) as a threshold equilibrium, we need optimality condition such that $V_i(\alpha_{\mu^*} | (\tau_i^*, \tau_{-i}^*)) \geq V_i((\alpha_i', \alpha_{-i,\mu^*}) | (\tau_i^*, \tau_{-i}^*))$ where α_i' is an OSD strategy such that player i deviates from H to L only at the one period and then return to equilibrium strategy α_{i,μ^*}after that. At $(\mu_1, \mu_2) = (\mu_{LH}^t, \mu_{LH}^n)$, the type N player 1 plays H and then plays L while (type N) player 2 plays L because $(\mu_{LH}^n, \mu_{LH}^t) \not\geq (\tau_i^*, \tau_{-i}^*) = (\mu_{LH}^t, \mu_{LH}^n)$. As a result, we have $V_1(\alpha^* | (\mu_1, \mu_2)) = d + \frac{\eta}{1-\eta}b < V_1((\alpha_1', \alpha_2^*) | (\mu_1, \mu_2)) = \frac{b}{1-\eta}$, which violates optimality condition of PBE.

(B) Consider the case $\tau_i^* < \theta$ and $\tau_{-i}^* > \theta$. Without loss of generality, I assume that $\tau_i^* = \mu_{HL}^t$ and $\tau_{-i}^* = \mu_{LH}^n$, where $n, t < \infty$. By using the similar logic in (A), I define an OSD strategy α_i'. Suppose that $(\mu_1, \mu_2) = (\mu_{HL}^t, \mu_{LH}^n)$. Then, $V_1(\alpha^* | (\mu_1, \mu_2)) = \frac{d}{1-\eta} < V_1((\alpha_1', \alpha_2^*) | (\mu_1, \mu_2)) = \frac{b}{1-\eta}$ which violates the optimality condition.

(C) Consider the case $\tau_i^* < \tau_{-i}^* < \theta$. Without loss of generality, I assume that $\tau_i^* = \mu_{HL}^t$ and $\tau_{-i}^* = \mu_{LH}^n$, where $n < t < \infty$. I similarly define an OSD strategy $\alpha_i^{'}$. At $(\mu_1, \mu_2) = (\mu_{HL}^t, \mu_{LH}^n)$, $V_1(\ \alpha^*|\ (\mu_1, \mu_2)) = \frac{1-\eta^{n-t}}{1-\eta} d + \eta^{n-t} V_1(\ \alpha^*|\ (\mu_{HL}^n, \overline{\mu}))$ $\leq V_i(\ (\alpha_1{}', \alpha_2^*)|\ (\mu_1, \mu_2)) = \frac{1-\eta^t}{1-\eta} b + \eta^t \frac{1-\eta^{n-t}}{1-\eta} c + \eta^n V_1(\ \alpha^*|\ (\overline{\mu}, \mu_{HL}^n))$, which violates optimality condition.

(D) Consider the case $\tau_i^* > \theta > \tau_{-i}^*$ and $\infty > \overline{T}(\tau_i^*) \geq \underline{T}(\tau_{-i}^*)$. Without loss of generality, let suppose $\tau_i^* = \mu_{LH}^t$ and $\tau_{-i}^* = \mu_{HL}^n$ where $t > n$. Consider that $(\mu_1, \mu_2) = (\overline{\mu}, \underline{\mu})$. At the state, a player 1 has $V_1(\alpha^*|(\overline{\mu}, \underline{\mu})) = \frac{1-\eta^n}{1-\eta} b + \eta^n V_1(\alpha^*|(\mu_{LH}^n, \mu_{HL}^n))$. We can decompose $V_1(\alpha^*|(\mu_{LH}^n, \mu_{HL}^n)) = d + \eta V_1(\alpha^*|(\mu_{LH}^{n+1}, \overline{\mu})) = d + \eta \left[\mu_{LH}^{n+1} \left(a + \eta V_1(\alpha_{\mu^*}|(\overline{\mu}, \overline{\mu})) \right) + (1 - \mu_{LH}^{n+1}) \left(a + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})) \right) \right]$. On the other hand, using an OSD strategy $\alpha_1^{'}$ gives $V_1((\alpha_1^{'}, \alpha_2^*)|\ (\overline{\mu}, \underline{\mu})) = d + \eta V_1(\alpha^*|(\mu_{LH}^1, \overline{\mu}))$. Note that we can also decompose $V_1(\alpha^*|(\mu_{LH}^1, \overline{\mu})) = \mu_{LH}^1 (a + \eta V_1(\alpha^*|(\overline{\mu}, \overline{\mu}))) + (1\text{-}\mu_{LH}^t) \left(d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})) \right)$. Optimaility condition of PBE requires

$$\frac{1-\eta^n}{1-\eta} b + \eta^n V_1(\alpha^*|(\mu_{LH}^n, \mu_{HL}^n)) \geq d + \eta V_1(\alpha^*|(\mu_{LH}^1, \overline{\mu})). \tag{1}$$

Unless we have

$$a + \eta V_1(\alpha^*|(\overline{\mu}, \overline{\mu})) \leq d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})), \tag{2}$$

(1) does not hold. Now we consider whether (2) holds.

Suppose that (2) holds. Since $V_1(\alpha^*|(\overline{\mu}, \overline{\mu})) = \overline{\mu} (a + \eta V_1(\alpha^*|(\overline{\mu}, \overline{\mu}))) + (1\text{-}\overline{\mu}) \left(d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})) \right)$, we have

$$V_1(\alpha^*|(\overline{\mu}, \overline{\mu})) \leq d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})), \tag{3}$$

and

$$V_1(\alpha^*|(\overline{\mu}, \overline{\mu})) = \frac{a + \eta(1-\overline{\mu}) \left(d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})) \right)}{1 - \eta \overline{\mu}}. \tag{4}$$

Moreover, from the assumption $a \geq (c + d)/2$, we have $d + \eta V_1(\alpha^*|(\underline{\mu}, \overline{\mu})) \leq \frac{a}{1-\eta}$. Then, we have

$$V_1(\alpha^*|(\overline{\mu},\overline{\mu})) \geq d + \eta V_1(\alpha^*|(\underline{\mu},\overline{\mu})), \qquad (5)$$

so that we have $V_1(\alpha^*|(\overline{\mu},\ \overline{\mu})) = d + \eta V_1(\alpha^*|(\underline{\mu},\ \overline{\mu}))$. Then, we finally have $a + \eta V_1(\alpha^*|(\overline{\mu},\ \overline{\mu})) = d + \eta V_1(\alpha^*|(\underline{\mu},\ \overline{\mu})) = \frac{a}{1-\eta}$, which implies $V_1((\alpha_1^{'},\alpha_2^*)|\ (\overline{\mu},\ \underline{\mu})) = V_1(\alpha^*|(\mu_{LH}^n,\ \mu_{HL}^n))$. To have $V_1(\alpha^*|(\overline{\mu},\ \overline{\mu})) \geq V_1((\alpha_1^{'},\alpha_2^*)|\ (\overline{\mu},\ \underline{\mu}))$, we need $\frac{b}{1-\eta} \geq V_1((\alpha_1^{'},\alpha_2^*)|\ (\overline{\mu},\ \underline{\mu})) = V_1(\alpha^*|(\mu_{LH}^n,\ \mu_{HL}^n))$. On the while, the individual rationality condition requires $V_1(\alpha^*|(\mu_{LH}^n,\ \mu_{HL}^n)) \geq \frac{b}{1-\eta}$ so we have $V_1(\alpha^*|(\mu_{LH}^n,\ \mu_{HL}^n)) = \frac{b}{1-\eta}$.

Now we consider $V_1(\alpha^*|(\mu_{HL}^n,\ \mu_{HL}^n))$ and $V_1((\alpha_1^{"},\alpha_2^*)|(\mu_{HL}^n,\ \mu_{HL}^n))$ where $\alpha_1^{"}$ is an OSD strategy at $(\mu_1,\ \mu_2) = (\mu_{HL}^n,\ \mu_{HL}^n)$. From $V_1(\alpha^*|(\mu_{HL}^n,\ \mu_{HL}^n)) = \frac{b}{1-\eta}$ and $V_1((\alpha_1^{"},\alpha_2^*)|(\mu_{HL}^n,\ \mu_{HL}^n)) = d + \eta V_1(\alpha^*|(\mu_{HL}^{n+1},\ \overline{\mu}))$, optimality condition requires

$$\frac{b}{1-\eta} \geq d + \eta V_1(\alpha^*|(\mu_{HL}^{n+1},\overline{\mu})). \qquad (6)$$

$V_1(\alpha^*|(\mu_{HL}^n,\ \mu_{HL}^n)) = \frac{b}{1-\eta} \geq V_1((\alpha_1^{"},\alpha_2^*)|(\mu_{HL}^n,\ \mu_{HL}^n)) = d + \eta V_1(\alpha^*|(\mu_{HL}^{n+1},\ \overline{\mu}))$. By using the similar decomposition, we have $V_1(\alpha^*|(\mu_{HL}^{n+1},\ \overline{\mu})) \geq V_1(\alpha^*|(\underline{\mu},\ \overline{\mu}))$, which implies $V_1((\alpha_1^{"},\alpha_2^*)|(\mu_{HL}^n,\ \mu_{HL}^n)) \geq \frac{a}{1-\eta} > \frac{b}{1-\eta}$. This inequality is contradictory to (6).

Now we showed that (2) will not hold, which implies that

$$a + \eta V_1(\alpha^*|(\overline{\mu},\overline{\mu})) > d + \eta V_1(\alpha^*|(\underline{\mu},\overline{\mu})). \qquad (2')$$

However, (2') implies

$$\frac{1-\eta^n}{1-\eta} b + \eta^n V_1(\alpha^*|(\mu_{LH}^n,\mu_{HL}^n)) < d + \eta V_1(\alpha^*|(\mu_{LH}^1,\overline{\mu})), \qquad (1')$$

which violates the optimality condition at $(\mu_1,\ \mu_2) = (\overline{\mu},\ \underline{\mu})$. □

1.8.3 Proof of Proposition 4

Lemma 2. Fix $G(g,\delta,\lambda,\theta)$. Suppose the set of the finite punishment equilibria is nonempty and there exists a finite-punishment equilibrium (α^*, τ^*) such that $V_i(\alpha^*|\,\overline{\mu}) \geq V_i(\alpha'|\,\overline{\mu})$ for any feasible finite-punishment equilibrium (α', τ') at G and for any i. Then, (α^*, τ^*) is not dominated by any feasible finite-punishment equilibria at G for any $\mu \in \Omega$.

Proof. Suppose that a finite punishment equilibrium (FN TE) (α^*, τ^*) satisfies $V_i(\alpha^*|\,\overline{\mu}) \geq V_i(\alpha'|\,\overline{\mu})$ for any feasible finite punishment equilibrium (α', τ') at G and for any i. This assumption also implies $V_i(\alpha^*|\,(\underline{\mu}, \overline{\mu})) \geq V_i(\alpha'|\,(\underline{\mu}, \overline{\mu}))$ for any feasible finite punishment equilibrium (α', τ') and for any i.

Now we assume that there exists some $\mu < \tau^*$ such that $V_i(\alpha^*|\,\mu) \leq V_i(\alpha^*_{OSD}|\,\mu)$ for any i and $V_i(\alpha^*|\,\mu) < V_i(\alpha^*_{OSD}|\,\mu)$ for some j where α^*_{OSD} is the two-sided one-shot deviation strategy at μ. With out loss of generality, assume that $V_1(\alpha^*|\,\mu) < V_1(\alpha^*_{OSD}|\,\mu)$, $\mu = \mu^{n'}_{HL}$ and $\tau^* = \mu^{n}_{HL}$ where $n' < n$ [8]. Then, if there is another feasible FN TE (α', τ') such that $\tau' = \mu^{n'}_{HL}$,

$$
\begin{aligned}
V_1(\alpha^*_{OSD}|\mu) &= \mu\left(a+\eta V_1(\alpha^*|\bar{\mu})\right) + (1-\mu)\left(d+\eta V_1(\alpha^*|(\underline{\mu},\bar{\mu})\right) \\
&< \mu\left(a+\eta V_1(\alpha^*|\bar{\mu})\right) + (1-\mu)\left(d+\eta V_1(\alpha^*_{OSD}|(\underline{\mu},\bar{\mu})\right) \\
&= \mu\left(a+\eta V_1(\alpha^*|\bar{\mu})\right) + (1-\mu)\left(d+\eta\left[\tfrac{1-\eta^{n'}}{1-\eta}b+\eta^{n'}V_1(\alpha^*_{OSD}|\mu)\right]\right) \\
&\;\;\vdots \\
&< V_1(\alpha'|\mu') = V_1(\alpha'|\mu).
\end{aligned}
$$

This relation implies $V_1(\alpha^*|\,\mu) < V_1(\alpha^{'}|\,\mu)$. However, this relation also violates the assumption $V_1(\alpha_{\mu^*}|\,(\underline{\mu}, \overline{\mu})) \geq V_1(\alpha_{\mu'}|\,(\underline{\mu}, \overline{\mu}))$ because

[8] For other cases where $n' \geq n$, $V_i(\alpha_{\mu^*}|\,\overline{\mu}) \geq V_i(\alpha_{\mu'}|\,\overline{\mu})$ trivially implies $V_i(\alpha_{\mu^*}|\,\mu) \geq V_i(\alpha_{\mu'}|\,\mu)$ so that we only consider the case $n < n'$.

$$\begin{aligned} V_1(\alpha_{\mu^*}|(\underline{\mu},\overline{\mu})) &= \tfrac{1-\eta^{n'}}{1-\eta}b + \eta^{n'}V_1(\alpha^*|\mu) \\ < \; V_1(\alpha^{'}|(\underline{\mu},\overline{\mu})) &= \tfrac{1-\eta^{n'}}{1-\eta}b + \eta^{n'}V_1(\alpha^{'}|\mu). \end{aligned}$$

So, for a FN TE (α^*, τ^*) that achieves the highest $V_i(\alpha^*|\,\overline{\mu})$ among all feasible FN TEs, (α^*, τ^*) cannot have any OSD strategy at any $\mu < \tau^*$. Moreover, we already have $V_i(\alpha^*|\,\overline{\mu}) \geq V_i(\alpha'|\,\overline{\mu})$ and $V_i(\alpha^*|\,(\underline{\mu},\,\overline{\mu})) \geq V_i(\alpha'|\,(\underline{\mu},\,\overline{\mu}))$ for any feasible FN TE (α', τ') and for any i . As a result, we can say that (α^*, τ^*) is not dominated by any other feasible FN TE at G. □

Lemma 3. Fix $G(g,\delta,\lambda,\theta)$. Suppose the set of the penance punishment equilibria is nonempty and there exists a penance equilibrium ($\alpha^{\#}$, $\tau^{\#}$) such that $V_i(\alpha^{\#}|\,\tau^{\#}) \geq V_i(\alpha'|\,\tau^{'})$ for any feasible penance equilibrium (α', τ') at G and for any i. Then, ($\alpha^{\#}$, $\tau^{\#}$) is not dominated by any penance equilibrium (α', τ') at G for any $\mu \in \Omega$.

Proof. We consider the game G such that all feasible penance equilibria (PN TEs) have $\tau_i' = \mu_{HL}^{n}$ and $\tau_{-i}' = \mu_{HL}^{n'}$ where $n > n' = 0$. At such game, $V_i(\alpha^{'}\,|\,(\bar{\mu},\bar{\mu}))$ and $V_i(\alpha^{'},|\,(\underline{\mu},\bar{\mu}))$ is the same for all feasible PN TEs. Since they all share the same equilibrium expected payoff at ($\underline{\mu}$,$\bar{\mu}$) and ($\bar{\mu}$,$\bar{\mu}$), we are allowed to only consider the dominance among the feasible PN TEs by comparing the expected payoff at ($\underline{\mu}$, $\underline{\mu}$).

Now suppose that a feasible PN TE ($\alpha^{\#}$, $\tau^{\#}$) where $\tau_i^{\#} = \mu_{HL}^{s}$ has the highest expected payoff at ($\underline{\mu}$, $\underline{\mu}$) among all feasible PN TEs. Then for any feasible PN TE ($\alpha^{'}$, $\tau^{'}$) we have

$$\begin{aligned} V_i(\alpha^*|\underline{\mu}) &= \tfrac{1-\eta^{s}}{1-\eta}b + \eta^{s}V_i(\alpha^{\#}|\tau^{\#}) \\ \geq \; V_i(\alpha^{'}|\underline{\mu}) &= \tfrac{1-\eta^{n'}}{1-\eta}b + \eta^{n'}V_i(\alpha^{'}|\tau'). \end{aligned}$$

So, for any $\mu < \tau^{\#}$, playing any other PN TE cannot dominate ($\alpha^{\#}$, $\tau^{\#}$). □

Lemma 4. Fix $G(g,\delta,\lambda,\theta)$. Suppose the set of the threshold equilibrium is nonempty. Then, for a there hold equilibrium (α^*, τ^*), two following properties are equivalent;

(1) $V_i(\alpha^*|\,\overline{\mu}) \geq V_i(\alpha'|\,\overline{\mu})$ for any feasible threshold equilibrium (α', τ'),

(2) $V_i(\alpha^*|\,\mu) \geq V_i(\alpha'|\,\mu)$ for any feasible threshold equilibrium (α', τ') at any reachable state μ.

Proof. Suppose that (α^{FN}, μ^{FN}) is the finite punishment equilibrium with the highest expected payoff at $\overline{\mu}$ and (α^{PN}, μ^{PN}) is the penance equilibrium with the highest expected payoff at μ^{PN}.

(1) Suppose $V_i(\alpha^{FN}|\,\overline{\mu}) \geq V_i(\alpha^{PN}|\,\overline{\mu})$ for any i. Then we have $V_i(\alpha^{FN}|\,(\underline{\mu}, \overline{\mu})) \geq V_i(\alpha^{PN}|\,(\underline{\mu}, \overline{\mu}))$ for all i. From (IC 3) of PN TE and (IC 2) of FN TE, we have

$$\begin{aligned} V_i(\alpha^{PN}|(\overline{\mu}, \underline{\mu})) &= d + \eta V_i(\alpha^{PN}|(\mu^1_{LH}, \overline{\mu})) \\ &\leq d + \eta V_i(\alpha^{FN}|(\mu^1_{LH}, \overline{\mu})) \\ &\leq V_i(\alpha^{FN}|(\overline{\mu}, \underline{\mu})). \end{aligned}$$

Now we assume that $\mu^{FN} = \mu^n_{HL}$ and $\mu^{PN} = \mu^{n'}_{HL}$. For case $n \geq n^{'}$, we have

$$\begin{aligned} V_i(\alpha^{PN}|\underline{\mu}) &= \tfrac{1-\eta^{n'}}{1-\eta} b + \eta^{n'} V_i(\alpha^{PN}|\mu^{n'}_{HL}) \\ &\leq \tfrac{1-\eta^{n'}}{1-\eta} b + \eta^{n'} V_i(\alpha^{FN}_{OSD}|\mu^{n'}_{HL}) \\ &= V_i(\alpha^{FN}_{OSD}|\underline{\mu}), \end{aligned}$$

where α^{FN}_{OSD} is a two-sided OSD strategy that plays H at $\mu^{n'}_{HL}$ and then return to α^{FN} after that. In the previous lemma, we showed that any feasible PN TE at any $\mu < \mu^{FN}$ cannot dominate $V_i(\alpha^{FN}|\,\mu)$ so that $V(\alpha^{PN}|\,\mu)$ cannot dominate $V(\alpha^{FN}|\,\mu)$ for any $\mu < \mu^{FN}$.

For case $n < n$', let suppose (α^{PN}, μ^{PN}) dominates (α^{FN}, μ^{FN}) at $(\underline{\mu}, \underline{\mu})$. Then, there exists some s where $n \leq s < n$' such that $V_i(\alpha^{FN}|\,(\mu^s_{HL}, \mu^s_{HL})) < V_i(\alpha^{PN}|\,(\mu^s_{HL}, \mu^s_{HL}))$ and $V_i(\alpha^{FN}|\,(\mu^{s+1}_{HL}, \mu^{s+1}_{HL}))$ $\geq V_i(\alpha^{PN}|\,(\mu^{s+1}_{HL}, \mu^{s+1}_{HL}))$. So

$$\begin{aligned} V_i(\alpha^{FN}|\underline{\mu}) &= \tfrac{1-\eta^n}{1-\eta} b + \eta^n V_i(\alpha^{FN}|\mu^n_{HL}) \\ &< \tfrac{1-\eta^{s+1}}{1-\eta} b + \eta^s V_i(\alpha^{PN}|\mu^s_{HL}) \\ &\leq \tfrac{1-\eta^{s+1}}{1-\eta} b + \eta^s V_i(\alpha^{FN}|\mu^s_{HL}). \end{aligned}$$

Then let define another FN TE ($\alpha^{FN'}$, $\mu^{FN'}$) where $\mu^{FN'} = \mu_{HL}^{s+1}$. As long as ($\alpha^{FN'}$, $\mu^{FN'}$) is feasible, $V_i(\alpha^{FN'}|\underline{\mu}\ ) > V_i(\alpha^{PN}|\ \underline{\mu})$ implies $V_i(\alpha^{FN'}|\overline{\mu}) > V_i(\alpha^{PN}|\ \overline{\mu})$ because of the facts that the expected payoff at ($\underline{\mu}$, $\overline{\mu}$) is equal to that at ($\underline{\mu}$, $\underline{\mu}$) and $V_i(\alpha^{FN'}|\overline{\mu}) > V_i(\alpha^{PN}|\ \overline{\mu})$ if and only if $V_i(\alpha^{FN'}|(\underline{\mu}, \overline{\mu})) > V_i(\alpha^{PN}|\ (\underline{\mu}, \overline{\mu}))$. This result contradicts to the assumption that no other feasible FN TE dominates (α^{FN}, μ^{FN}). Consequently, (α^{PN}, μ^{PN}) cannot dominate (α^{FN}, μ^{FN}) at ($\underline{\mu}$, $\underline{\mu}$).

(2) Suppose $V_i(\alpha^{PN}|\ \overline{\mu}) \geq V_i(\alpha^{FN}|\ \overline{\mu})$ for all i. Similarly we have $V_i(\alpha^{PN}|\ (\underline{\mu}, \overline{\mu})) \geq V_i(\alpha^{FN}|\ (\underline{\mu}, \overline{\mu}))$ for all i. For ($\overline{\mu}$, $\underline{\mu}$), the optimality condition (1) of PBE satisfies

$$\begin{aligned} V_i(\alpha^{PN}|(\underline{\mu},\bar{\mu})) &= \underline{\mu}\left(c+\eta V_i(\alpha^{PN}|(\bar{\mu},\mu_{LH}^1))\right) + (1-\underline{\mu})\left(b+\eta V_i(\alpha^{PN}|(\underline{\mu},\mu_{LH}^1)\right) \\ &\geq \underline{\mu}\left(a+\eta V_i(\alpha^{PN}|\bar{\mu})\right) + (1-\underline{\mu})\left(d+\eta V_i(\alpha^{PN}|(\underline{\mu},\bar{\mu})\right) \\ &= V_i(\alpha_{OSD}^{PN}|(\underline{\mu},\bar{\mu})). \end{aligned}$$

for all i. That is, a two-sided OSD strategy α_{OSD}^{PN} at ($\overline{\mu}$, $\underline{\mu}$) is worse than α^{PN} for player i with $\mu_i = \underline{\mu}$ so that player j with $\mu_j = \overline{\mu}$ cannot assume that his/her opponent i will follow α_{OSD}^{PN}. For this reason, having the optimality condition as a one-sided OSD robustness is enough for showing a two-sided OSD robustness.

Now we assume that $\mu^{FN} = \mu_{HL}^{n}$ and $\mu^{PN} = \mu_{HL}^{n'}$. For a case $n \leq n$', let suppose (α^{FN}, μ^{FN}) dominates (α^{PN}, μ^{PN}) at ($\underline{\mu}$, $\underline{\mu}$). Then, assume that another PN TE ($\alpha^{PN'}$, $\mu^{PN'}$) where $\mu^{PN'} = \mu_{HL}^{n'}$ is feasible at G. Since all PN TEs have the equivalent expected payoff at ($\overline{\mu}$, $\overline{\mu}$) and ($\underline{\mu}$, $\overline{\mu}$), as long as it is feasible,

$$\begin{aligned} V_i(\alpha^{PN}|\underline{\mu}) &= \tfrac{1-\eta^n}{1-\eta}b + \eta^n V_i(\alpha^{PN}|\mu_{HL}^n) \\ &< \tfrac{1-\eta^{n'}}{1-\eta}b + \eta^{n'} V_i(\alpha^{FN}|\mu_{HL}^{n'}) \\ &\leq \tfrac{1-\eta^{n'}}{1-\eta}b + \eta^{n'} V_i(\alpha^{PN}|\mu_{HL}^{n'}) \\ &= V_i(\alpha^{PN'}|\underline{\mu}). \end{aligned}$$

This result contradicts to the assumption that (α^{PN}, μ^{PN}) is not dominated by any other feasible PN TE. To avoid such contradiction, any other PN TE that has a threshold pair higher than μ^{PN} should not be feasible at G. However, having threshold pair lower than μ^{PN} implies that those feasible PN TEs will not dominate (α^{PN}, μ^{PN}) because of the fact that $V_i(\alpha^{PN}|\ (\overline{\mu}, \overline{\mu})) \geq V_i(\alpha^{PN}|\ (\underline{\mu}, \overline{\mu}))$ which is implied from the individual rationality.

The case $n > n$' also can be similarly proved. □

Lemma 5. Fix $G(g,\delta,\lambda,\theta)$. Suppose the set of the threshold equilibria is nonempty and there exists a threshold equilibrium (α^*, τ^*) such that $V_i(\alpha^*|\ \overline{\mu}) \geq V_i(\alpha'|\ \overline{\mu})$ for any feasible threshold equilibrium (α', τ') at G. Then, (α^*, τ^*) is robust to any pure action OSD strategy.

Proof. Suppose that the set of the threshold equilibria is nonempty and there exists a threshold equilibrium (α^*, τ^*) such that $V_i(\alpha^*|\ \overline{\mu}) \geq V_i(\alpha'|\ \overline{\mu})$ for any feasible threshold equilibrium (α', τ'). We only consider the robustness for the two-sided OSD strategy since the robustness to the one-sided OSD is achieved by the optimality condition of PBE. We also assume that non-trivial threshold equilibrium (α^*, τ^*) allows players to play H at ($\overline{\mu}$, $\overline{\mu}$) and play L at at ($\underline{\mu}$, $\underline{\mu}$) as the equilibrium strategy respectively.

(1) At $\mu \geq \tau^*$, playing an OSD strategy (= L) rather than the equilibrium strategy (= H) will not be beneficial for both players. Let denote the OSD strategy (α^*_{OSD}, τ^*_{OSD}) that deviates from (α^*, τ^*) at ($\overline{\mu}$, $\overline{\mu}$). From the individual rationality, we have

$$\begin{aligned}
V_i(\alpha^*_{OSD}|\overline{\mu}) &= b + \eta V_i(\alpha^*|(\mu^1_{LH}, \mu^1_{LH})) \\
&= b + \eta\left\{\mu^1_{LH}\left(a + \eta V_i(\alpha^*|\bar{\mu})\right) + (1-\mu^1_{LH})\left(d + \eta V_i(\alpha^*|(\underline{\mu}, \bar{\mu})\right)\right\} \\
&\leq \overline{\mu}\left(a + \eta V_i(\alpha^*|\bar{\mu})\right) + (1-\overline{\mu})\left(d + \eta V_i(\alpha^*|(\underline{\mu}, \bar{\mu})\right) \\
&= V_i(\alpha^*|\overline{\mu})
\end{aligned}$$

for any i. Given that playing (α^*, τ^*) is weakly better than the OSD strategy, (α^*, τ^*) will be robust to the two-sided OSD strategy at any other states $\mu \in \Omega$ where $\mu \geq \tau^*$ by the similar extension.

(2) At $\mu < \tau^*$, playing an OSD starategy (= H) rather than the equilibrium strategy (= L) will not be beneficial for both players. Let denote the OSD strategy (α^*_{OSD}, τ^*_{OSD}) that deviates at such μ. Knowing that the opponent, say -i, plays H, player i has

$$\begin{aligned} V_i(\alpha^*_{OSD}|\mu) &= \mu\left(a+\eta V_i(\alpha^*|\bar{\mu})\right) &+ (1-\mu)\left(d+\eta V_i(\alpha^*|(\underline{\mu},\bar{\mu})\right) \\ &\leq \mu\left(c+\eta V_i(\alpha^*|(\bar{\mu},\underline{\mu}))\right) &+ (1-\mu)\left(d+\eta V_i(\alpha^*|\underline{\mu}\right) \end{aligned}$$

by the (IC 1) condition of each TEQ. Given that playing (α_{μ^*}, μ^*) is weakly better than the OSD strategy, (α^*, τ^*) will be robust to the two-sided OSD strategy at any other states $\mu \in \Omega$ where $\mu < \tau^*$ by the similar extension.

Consider the case that players agree to play OSD with the expectation of playing another threshold equilibrium, say (α', τ'), which allows players to play H at $\mu < \tau^*$. That is, players agree to play L as if they play another TEQ's strategy but renegotiate again to return to (α^*, τ^*) once they arrive to the next state. In such case, given the assumption that (α^*, τ^*) provides the weakly highest expected payoff at any states, players may realize that the expected payoff at such state will not be feasible or not dominate the expected payoff from (α^*, τ^*) from the result of the previous lemma. As a result, (α^*, τ^*) will be robust to the two-sided OSD.

(3) Consider the state μ where $\mu_1 \geq \tau^*_i$ and $\mu_2 < \tau^*_{-i}$ without loss of generality. For player 2, as in (2), playing the equilibrium action (= L) will bring weakly better expected payoffs than the OSD strategy (= H) even if the player 1 has agreed to play the OSD strategy (because of her low belief about the player 1's type). As a result, the player 2 may not have an incentive to follow the OSD strategy. For the player 1, with consideration that the player 2 would not follow the OSD strategy, her expected payoff from the two-sided OSD strategy will not be actually realized. As a result, the one-sided OSD strategy is the only available OSD strategy she can consider and IC conditions of each TEQ blocks such OSD strategy as the weakly worse strategy than (α^*, τ^*). □

1.8.4 Proof of Theorem 1

Theorem 1. Fix $G(g, \lambda, \theta, \delta)$. Suppose that there exists some threshold equilibrium (α^*, τ^*). Then, followings are equivalent;

(1) Threshold strategy α^* have two phases on the equilibrium path:

(a : Cooperation) Any type N players play H at t if $a^{t-1} = (H, H)$.

(b : Punishment) For any t, if there is any L played in h^t, any types of players play L at t .

(2) Threshold pair τ^* is characterized by the quartet (n, m, s, t) where $m < \infty$ and $\llcorner \frac{m}{min\{s, t+1\}} \lrcorner < 1$.

Proof.

To show equivance of two considtions, I will show that (1) $\Rightarrow$(2) and then (2) $\Rightarrow$(1) holds.

((1) $\Rightarrow$ (2)) Suppose the TEQ (α^*, τ^*) follows on- and off-the-equilibrium path described in (1) repectively. Then, to satisfy the equilibrium path of (a) and (b), $m < \min\{s, t+1\}$ needs to be satisfied. Suppose not. First consider that $t + 1 < s$. Then, the case $m \geq t + 1$ corresponds to the proof (A) case of the propotion 1, which does not support the TEQ. So, we exclude this case. Second, consider the case $s < t + 1$[9] and $m \geq s$. Then we can put the equilibrium threshold pair $(\tau_i^*, \tau_{-i}^*) = (\mu_{LH}^m, \mu_{HL}^s)$. From this threshold, consider the player i triggered the punishment phase at the τth calender time period. Then, the equilibrium path requires both players to play L for s periods. After s periods of punishment play (L, L), the player i would arrive to the state $(\mu_{LH}^s, \mu_{HL}^s) \geq (\tau_i^*, \tau_{-i}^*) = (\mu_{LH}^m, \mu_{HL}^s)$. Then, at the $\tau+s+1$ th calender period, which is the same to the $s+1$ th period of the punishment phase, the type N player i must play the equilibrium play H which is followed by the equilibrium play H if the player -i is the type N. As a result, having $s < t+1$ and $m \geq s$ will not have the on-the-equilibrium path of (a) and (b).

((2) $\Rightarrow$(1)) From the above part, it is shown that having a threshold characterized by the quartet (n, m, s, t) which satisfies $m < \min\{s, t+1\}$ automatically implements the equilibrium path of (a) and (b). □

[9] Since $s < \infty$ automatically implements $t = \infty$, having strict inequality for $t + 1 < s$ and $s < t + 1$ would not be affect the result of the proof.

1.8.5 Proof of Theorem 2

Theorem 2. Suppose a threshold (τ_i^*, τ_{-i}^*) supports TEQ. Then followings are equivalent;

(1) TEQ consists of following phases :

(a : Cooperation) Any type N players play H at t if $a^{t-1} = (H, H)$.

(b : Punishment) Suppose any side(s) deviates from the cooperation phase. Then, both sides' players play punishment L for n (=s) periods. Right after n periods of (L, L), any type N players plays H.

(2) TEQ is characterized by (n, m, s, t) such that $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner = \infty$ and $n = s < \infty$

(3) for given $(g, \lambda, \theta, \delta)$, there exists $n < \infty$ such that

(IC 1) $\mu_{HL}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{HL}^n) \cdot V_{IC_B}^{IF}(n) > \mu_{HL}^n \cdot V_{IC_C}^{IF}(n) + (1 - \mu_{HL}^n) \cdot V_{IC_D}^{IF}(n)$,

(IC 2) $d + \eta \cdot \left(\mu_{LH}^1 \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{LH}^1) \cdot V_{IC_B}^{IF}(n)\right) \leq \frac{1-\eta^n}{1-\eta} b + \eta^n \left(\mu_{LH}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{LH}^n) \cdot V_{IC_B}^{IF}(n)\right)$,

where $V_{IC_A}^{IF}(n) = \frac{a + \eta(1-\overline{\mu}) V_{IC_B}^{IF}(n)}{1 - \eta\overline{\mu}}$, $V_{IC_B}^{IF}(n) = \frac{\left(d + \eta \frac{1-\eta^n}{1-\eta} b + \eta^{n+1} \frac{\mu_{HL}^n \cdot a}{1-\eta\overline{\mu}}\right)}{1 - \left(\mu_{HL}^n \frac{\eta(1-\overline{\mu})}{1-\eta\overline{\mu}} + (1 - \mu_{HL}^n)\right) \eta^{n+1}}$,

$$V_{IC_C}^{IF}(n) = c + \eta \frac{1-\eta^n}{1-\eta} b + \eta^{n+1} \left(\mu_{LH}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{LH}^n) \cdot V_{IC_B}^{IF}(n)\right),$$

$$V_{IC_D}^{IF}(n) = \frac{1-\eta^{n+1}}{1-\eta} b + \eta^{n+1} \left(\mu_{HL}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{HL}^n) \cdot V_{IC_B}^{IF}(n)\right).$$

$$V_{IC_C}^{IF}(n) = c + \eta \frac{1-\eta^n}{1-\eta} b + \eta^{n+1} \left(\mu_{LH}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{LH}^n) \cdot V_{IC_B}^{IF}(n)\right),$$

$$V_{IC_D}^{IF}(n) = \frac{1-\eta^{n+1}}{1-\eta} b + \eta^{n+1} \left(\mu_{HL}^n \cdot V_{IC_A}^{IF}(n) + (1 - \mu_{HL}^n) \cdot V_{IC_B}^{IF}(n)\right).$$

Proof. In this proof of the theorem 2, I will show that (1) $\Leftrightarrow$ (2) holds. Since I explained how the IC conditions of (3) are implemented from the threshold pair $(\tau_i^*, \tau_{-i}^*) = (\mu_{HL}^n, \mu_{HL}^n)$ in the original paper, we will omit for the part (3) $\Leftrightarrow$ (1).

((1) $\Rightarrow$ (2)) Suppose that TEQ (α^*, τ^*) that follows the on- and off-the-equilibrium path in (1). Suppose that both sides triggered the punishement phase at the same time. Then, to return to the cooperation phase at n (= s) th period of the punishment phase, it is required to have $(\tau_i^*, \tau_{-i}^*) = (\mu_{HL}^n, \mu_{HL}^s)$. so we have $m = \infty$ which implies $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner = \infty$.

((2) $\Rightarrow$ (1)) From the above part, it is true that having the quartet satisfies $\llcorner \frac{m}{min\{s,t+1\}} \lrcorner = \infty$ and $n = s < \infty$ automatically implement (a) and (b) part of (1). In this part, I will show that the quartet also implements off-the-equilibrium path of (c) and (d). First, suppose that the player i's side triggered the punishement phase at the calendar time τ. If the player i deviates to play H at v ($< n$-1)th period of the punishment phase, then the player i and j arrives at the state $(\mu_{LH}^{v+1}, \bar{\mu})$ and $(\bar{\mu}, \mu_{LH}^{v+1})$ respectively. Since $\bar{\mu} \geq \mu_{LH}^{v+1} > \theta > \mu_{HL}^{n} = \tau_i^* = \tau_{-j}^*$, any type N players of both sides would play H. If the player j deviates to play H v ($< n$-1) th period of the punishment phase, then the player i and j arrives at the state $(\bar{\mu}, \mu_{HL}^{v+1})$ and $(\mu_{HL}^{v+1}, \bar{\mu})$ respectively. Since $\mu_{HL}^{v+1} < \mu_{HL}^{n}$, both player needs to play L for remaining n - (v+1) periods to arrive to μ_{HL}^{n}. Second, suppose that both sides deviates from the cooperation phase at the same calendar time τ. If the player i solely deviate to play H at the $v(<n-1)$th period of the punishment phase, then he player i and j arrives at the state $(\mu_{HL}^{v+1}, \bar{\mu})$ and $(\bar{\mu}, \mu_{HL}^{v+1})$ respectively so that they have to wait for another n - (v+1) periods to arrive to the threshold level μ_{HL}^{n}. If both players deviate at the same time at any $v(<n)$, then both player immediately jumps to the state $(\bar{\mu}, \bar{\mu})$ and any type N players of the both sides plays H at the next period. □

2 Chapter 2: Sequential Choice with Preference Learning

2.1 Introduction

Most of decision theory frameworks assumes a decision maker's preference to be fixed over time. However, such standard setting is vulnerable to some real world observations, especially to a "*choice reversal*." A choice reversal depicts the following situation: consider two (available) alternatives a and b. The decision maker chooses a at the first chance. At the second chance of her choice she chooses b even though a is still available. Reily (1982) and Grether and Plott (1979) found experimental evidences for such an choice reversal which cannot be explained when they assume a fixed preference. A fixed preference restricts a decision maker to have the same choice outcome at the same choice set, so that such a choice reversal should not happen.

To explain a choice reversal, some frameworks separated actual choice set and nominal choice set. They assumed that, even though a decision maker faces some choice set, the actual choice set within which she considers to choose alternatives can be different from that. They called it as a "*consideration set*." In particular, Masatlioglu, Nakajima, and Ozbay (2012) (henceforth, MNO), Masatlioglu and Nakajima (2012) (henceforth, MN), and Caplin and Dean (2011) (henceforth, CD) tried to explain choice reversal in such a perspective.

On the other hand, Manzini and Mariotti (2007, 2011) assumed a decision maker who considers

multiple attributes of alternatives. Such a decision maker assumed to have a hierarchy on attributes and filter out alternatives sequentially according to such a hierarchy.

Different from such works, I consider a situation in which a decision maker's underlying (complete) preference is partially reflected on a decision making while the choice set is fully considered. That is, a decision maker has a temporal preference that consists of subset of a complete preference. A sequence of temporal preferences nondecreasingly converges to her complete preference. To explain such a situation, I construct a *choice sequence with binary learning* (henceforth, CBL) framework.

CBL framework assumes a sequential choice environment in which decision maker faces arbitrary choice sets and chooses her optimal alternative over time. Decision maker has her fixed underlying preference which is complete, transitive and acyclic preference, say a "well-defined" preference, over all the available alternatives. However, her underlying preference may not fully known to herself. We can think of several reasons for such a status: insufficient information about specifications of commodities, inexperience of commodity feature, etc.. Then, decision maker may uncover her underlying preference by sequential experience of choices. In this study, I focus on a specific learning rule called "*binary learning rule*."

Binary learning rule basically resembles binary information search process from computer science. Based on her underlying preference, decision maker compares some "reachable" alternatives to her current choice. If she find some alternative is better than current choice in her underlying preference, she learns a binary relation and use it for her next choice. I also allowed decision maker to learn her underlying preference not only via direct comparace but also via indirect comparance. For example, once she learned that "a is bettern than b" and "b is better than c" separatedly, then she will automatically learn "a is bettern than c." I axiomitized binary learning rule into two properties called "$elimination$" and "$transitivity$." However, different from computer, human decision maker may not reach to all the alternatives in the process of learning. When she compares her current choice with another alternative, her comparance may depend on her horizon of experience. For example, when she considers a laptop computer to buy, she would remind her current experience.

When she felt that her current one is heavy, which weights 4.2 lb, she will consider candidates that weight less than 4.2 lb. However, if she did not felt any inconvenience so far from a screen definition, she may not care of it. Likewise, her learning process itself is restricted by her current horizon of experience. I depicted such a restriction as a "*range of learning*." That is, when decision maker compares her current choices with other alternatives, she is restricted consider alternatives only in her current range of learning.

And then, I characterize a choice sequence that comes from a behavior of a decision maker who follows binary learning rule with certain range of learning. I formalize this characterization as *sequential weak axiom of revealed preference* (WARP - S). Once we find a sequence that satisfies WARP - S, then we can uncover some well-defined underlying preference.

Moreover, we can explain some extent of multiple-time choice reversal using CBL framework[1]. Assuming fully known underlying preference is not able to explain multiple-time/repeated choice reversals that is observed in the field experiments like Chu and Chu (1990) and Cox and Grether (1996). In these works, they exhibited that frequency of such event is reducing over repetition of conducts. That is, the more subjects repeat choices between two alternatives, the less subjects turn over from their previous change. Most of subjects stopped to turn over between the second or the third time repetition. In psychology, *frequency − based* (*probability*) *learning theory* explains such a situation[2]. The model explans well a particular event with a lottery, however it cannot embrace a decision making situation which requires standard preference order. On the other hand, CBL framework embraces a standard preference order and multiple-time choice reversal within the same framework.

The remaining part of this paper has the following structure. In the section 2, I will introduce the basic setting of the model and describe the learning rule. In the section 3, I will characterize choice sequences that satisfy CL framework according to the decision maker's range of learning.

[1]Multiple-time/repeated choice reversal depicts the situation in which the choice reversal occurs more than or equal to two times. That is, once a is chosen over b, and in the later time b is chosen over a. Moreover, in the next time, a is chosen over b again, and so on. Flipping between two alternatives more than one time is depicted as multiple-time/repeated choice reversal.

[2]For more detailed explanation on this model, see the Humprey (2006).

And then, in the section 4, I will discuss about properties and relationship among differently characterized choice sequences. Section 5 will compare the implication and the prediction result to those from previous works, especially CD (2011) and MN (2012). Section 6 will conclude this paper and suggest the path for future research.

2.2 Model

2.2.1 Basic Environment

I consider a decision maker who faces a sequence of choice sets. Let X be a finite set of alternatives and $\mathbb{X} = 2^{|X|}\backslash\{\emptyset\}$ be the set of all nonempty subsets of X. $t \in \mathbb{T} = \{\ 1, 2, \cdots\ \}$ is a discrete time index. $S_t \in \mathbb{X}$ is a choice set at time t given as subset of X. $c_t \in S_t$ is a *choice outcome* at time t. A *choice sequence* is a sequence of pairs $h^T = (S_t, c_t)_{t=0}^{T}$ for some finite $T \in \mathbb{T}$. $\mathscr{H}$ is the set of all finite choice sequences. $\mathbb{P} = 2^{|X\times X|}$ is the set of all nonempty binary relations on X. A binary relation on X is a set of binary pairs of alternatives in X.

I define the binary relation over X at time t $\succsim_t \in \mathbb{P}$. I call the set $\succsim_t$ as *temporal preference* at t and for any $x, y \in X$ such that $(x, y) \in \succsim_t$ holds, I say "x is *preferred to* y at time t" and denote $x \succsim_t y$. I also define a set $\succ_t$

$$\succ_t = \{(x, y) \in X \times X \mid (x, y) \in \succsim_t \textit{ and } (y, x) \notin \succsim_t \}$$

and for any $x, y \in X$ such that $(x, y) \in \succ_t$ holds, I say "x is *strictly preferred to* y at time t" and denote $x \succ_t y$. Similarly, I define a set $\sim_t$

$$\sim_t = \{(x, y) \in X \times X \mid (x, y) \in \succsim_t \textit{ and } (y, x) \in \succsim_t \}$$

and for any $x, y \in X$ such that $(x, y) \in \succ_t$ holds, I say "x is *indifferent to* y at time t" and denote $x \sim_t y$.

I assume that the *initial preference* $\succsim_0 = X \times X$ so that the decision maker starts from the state that she is indifferent among all the alternatives. I also assume the existence of the *underlying preference* $\succ$ which is complete, transitive, and asymmetric relation[3] over X. The set of the underlying preference is defined as $\mathbb{P}_{\succ}$.

2.2.2 Learning Rules

Learning rule governs change of the decision maker's preference over time. The decision maker's preference change is irreversible: decision maker's preference is only allowed to changed from the indifference relation to the strict relation. We can consider such preference change as the "learning" process. For this concept of preference learning, I characterize the learning process with two features. First, the learning process should be restricted to change the temporal preference to fixed underlying preference. That is, the learning process will stop after some finite steps of learning. Second, the learning process should be restricted to some range of learning. That is, when decision maker learns of his/her underlying preference, he/she would not learn all the preference relation between all alternatives. With those two feature in mind, I formalize the *learning rule* as a function Γ: $\mathscr{H} \times \mathbb{X} \times \mathbb{P}_{\succ} \to \mathbb{P}$. At some t, $\Gamma(h^t, B_t, \succ)$ is a temporal preference with respect to a history h^t, a range of learning $B_t \subseteq X$, and underlying preference $\succ \in \mathbb{P}_{\succ}$. Defining a learning rule requires to specify corresponding underlying preference and the (sequence of) range of learning $(\succ, \{B_t\}_{t=0}^{\infty})$. For this reason, when it is required, we will explicitly denote $(\succ, \{B_t\}_{t=0}^{\infty})$ for the specification of the learning rule.

From $h^0 \in \mathscr{H}$, the learning rule *produces* a sequence of temporal preferences up to T $\{\Gamma(\xi_t)\}_{t=0}^{T}$, where $\Gamma(\xi_t)$ is the binary relation over X. For simplicity, I define a set of all triplets $\Xi \equiv \mathscr{H} \times \mathbb{X} \times \mathbb{P}_{\succ}$ and triplet $\xi_t \equiv (h^t, B_t, \succ)$.

Now we consider two properties on the learning rules: for all t, for any $x, y \in X$,

[3]For simplicity, I omitted the explanation of completeness, transitiveness, and asymetricity of the binary relation. I define the conditions as following: (1) Completeness: for any $x, y \in X$, either $x \succ y$ (or $(x, y) \in \succ$) or $y \succ x$ (or $(y, x) \in \succ$), (2)Transitiveness: if $x \succ y$ (or $(x, y) \in \succ$) and $y \succ x$ (or $(y, z) \in \succ$), then $x \succ z$ (or $(x, z) \in \succ$), (3) Asymmetricity: if $x \succ y$ (or $(x, y) \in \succ$), then not $y \succ x$ (or $(y, x) \notin \succ$).

(*Elimination*) For any $x = c_t$, $y \in B_t$, either $[x \succ y \Rightarrow (y, x) \notin \Gamma(\xi_t)]$ or $[y \succ x \Rightarrow (x, y) \notin \Gamma(\xi_t)]$

(*Transitivity*) For any $x = c_t$, $\{y, y'\} \subseteq B_t$, $x \succ y$ and $y' \succ x \Longrightarrow (y', y) \in \Gamma(\xi_t)$.

(Elimination) property formally defines how the decision maker's temporal preference between current choice and alternative in her range of learning comes to converge to her underlying preference. Given c_t and B_t, she compares her current choice and alternatives in her range of learning. Then, she realizes her (binary) underlying preference between them. This process is formalized as elimination of temporal preferences that is not matched to her underlying preference. Through elimination, she only left temporal preference that coincides with her underlying preference from the next periods. From the (Delimitation) property, any sequence of preferences must be non-increasing sequence, formally $\Gamma(\xi_t) \subseteq \Gamma(\xi_{t+1})$ for all t.

(Transitivity) property formally imposes indirect way of learning branches from (Elimination). It assumes that decision maker is discernible between two alternatives that are better and worse respectively than her current choice. If some alternative in range of learning is considered strictly better than the current choice and the other one is worse than that, I assume that she can compare those two even though she didn't directly experienced them this period.

These properties can be thought as the least bound of learning process. Even though the decision maker cannot figure out the whole ranking of given alternatives, she can at least learn whether her current choice is better and/or worse than what she compares to. And also, from her learning, she never fails to learn the transitive relation between the current choice and the other alternatives. For this reason, this study restricts its focus on the learning rules that equips (Elimination) and (Transitivity) as the least properties of learning process. This restriction brings the formal definition of the *binary learning rule*.

Definition 1. Assume Γ_b is a binary learning rule if and only if , for all t and for any $x, y \in X$,

$$(x, y) \in \Gamma_b(\xi_{t+1}) \Leftrightarrow$$

$$(x, y) \in \Gamma_b(\xi_t) \textit{ and } (x, y) \textit{ is neither excluded by } (\textit{Elimination}) \textit{ nor by } (\textit{Transitivity}) \textit{ at } t.$$

2.3 Characterization

I characterize a choice function that rationalizes binary learning rule. Different from the static choice function that only considers current choice sets, we restricts the choice function to be consistent with sequential choice histories. Formally, choice function C: $\mathscr{H} \times \mathbb{X} \to X$ induces an alternative from the current choice set with respect to choice history so far. With consideration of availability of chosen alternative at certain time, we denote $C_t \equiv C(h^t, S_t)$. [4]

I define a choice function that depicts the behavior of decision maker who adopts binary learning rule.

Definition 2 (Choice with Binary Learning)**.** A choice function C is a *choice with binary learning* (CBL) if and only if there exists a binary learning rule Γ_b with respect to $(\succ, \{B_t\}_{t=0}^{\infty})$ such that, for any h^t, C_t is $\succsim_t$ - best alternative and $\Gamma_b(\zeta_{t-1}) = \succsim_t$. *i.e.*, for any t

$$C_t \in x \in S_t \,|\, \textit{there exists no } y \neq x \in S_t \textit{ such that } y \succ_t x \textit{ where } \Gamma_b(\zeta_{t-1}) = \succsim_t.$$

Then I call a choice function C is*generated* by binary learning rule Γ_b with respect to $(\succ, \{B_t\}_{t=0}^{\infty})$, or simply say C is *generated* by $(\succ, \{B_t\}_{t=0}^{\infty})$.

The definition of CBL allows the decision maker to have indifferent preference among some alternatives. For such case, we adopt the tie-breaking rule that the decision maker chooses any of indifferent alternatives with the same probability. Since the number of alternatives are restricted to be finite, the probability that is assigned to indifferent alternatives will be strictly positive.

Before we get into the deeper discussion, we briefly recall the static choice function. In static choice situation, the decision maker's preference is assumed to be fixed over time. That is, se-

[4] For $t = 0$, I assume $\mathscr{H}_{t-1} = (\{\emptyset\}, \{\emptyset\}, \{\emptyset\})$.

quential consistency of choice is automatically induced from static consistency. A notion of static consistency is formalized by standard weak axiom of revealed preference (WARP).

Definition 3 (Standard WARP)**.** A choice function C satisfies the *standard Weak Axiom of Revealed Preference* (WARP) if the following property holds: for any $x, y \in X$

$$x = C_t \, and \, x, y \in S_t \Longrightarrow \forall t' \neq t \, such \, that \, x, y \in S_{t'}, y \neq C_{t'}.$$

Example 1. Let $X = \{a, b, c\}$, underlying preference $\succ : a \succ b \succ c$, and $B_t = X$ for all t. Assume that $(S_0, C_0) = (\{a, b, c\}, c)$, $(S_1, C_1) = (\{a, b, c\}, b)$, $(S_t, C_t) = (\{a, b\}, a)$ for any $t \geq 2$. This choice sequence violates standard Weak Axiom of Revealed Preference (WARP) since (S_0, C_0) and (S_1, C_1) show different choice behaviors even b and c are both available at $t = 1$ and $t = 2$.

On the other hand, consider a choice function C generated by $(\succ, X)$. $(\mathbf{0})$ We start from the initial preference $\succsim_0 : a \sim_0 b \sim_0 c$. $(\mathbf{1})$ $\Gamma(\succ, \succsim_0, c, X) = \{(a, b), (b, a), (a, c), (b, c)\} = \succsim_1$. Since $\succ$ includes (a, c) and (b, c), (c, a) and (c, b) are excluded from $\succsim_0$ by (Elimination). $(\mathbf{2})$ From $t = 1$ to 2, $\Gamma(\succ, \succsim_1, b, X) = \{(a, b), (a, c), (b, c)\} = \succsim_2$. As in the first step, (b, a) is excluded from $\succsim_1$ by (Elimination). As a result, from the next time, $C_t = a$ given $S_t = \{a, b\}$.

In this example, we can see the case that is unable to be explained by standard fixed preference but able to by choice with binary learning. For each period, each choice outcome C_t is the optimal outcome according to $\succsim_t$. c is not rejected as an optimal choice given her initial preference. At $t = 1$, b is not rejected as her optimal choice given her preference $a \sim_1 b \succ_1 c$. Excluding c from her choice outcome is rational enough behavior at this step. From $t = 2$, a is always chosen. Since a is the best outcome according to her preference $a \succ_2 b \succ_2 c$, it is also said to as rational behavior.

On the other hand, from the above example, we can observe the features that characterize choice with binary learning: there is no choice reversal between already chosen items. Once her learning of preference between alternatives occurs, the decision maker does not revert choice between them. For this reason, if she chooses another alternative in the later time (even though the past choice is

still available), it reveals that her underlying preference strictly prefers the later alternative to the past one. For the same reason, if she chooses the same item in the later time, it also reveals that the item is the most preferred one among her choice set.

I generalize these observations to characterize revealed preference for sequential choice with learning. To begin with, we need to define the notion of revealed preference on sequential choice environment. Since the binary learning rule produces temporal preferences depends on its range of learning, we need to specify it.

Definition 4 (Revealed Preference)**.** Assume C is a choice function. Define the *revealed preference at t* with respect to $B = \{B_t\}_{t=0}^{\infty}$ $\mathbf{P_B^t} \in \mathbb{P}$.

For any $x, y \in X$, $(x, y) \in \mathbf{P_B^t}$ if

(i) $C_s = y$, $C_r = x$ and $x, y \in B_s, S_r$, where $s < r \leq t$, or

(ii) $C_s = C_r = x$ and $x, y \in B_s, S_r$, where $s < r \leq t$, or

(iii) $\exists\, w \in X$ such that $x\, \mathbf{P_B^r}\, w$ and $w\, \mathbf{P_B^s}\, y$ where $r \neq s \leq t$, and $\exists t' \leq t$ such that $C_{t'} \in \{x, y, w\}$ and $x, y \in B_{t'}$.

Then, I denote $x\, \mathbf{P_B^t}\, y$ and call x is *revealed preferred to y at t*.

Using the notion of revealed preference, I define sequential weak axiom of revealed preference with respect to $B = \{B_t\}_{t=0}^{\infty}$.

Definition 5 (WARP - S)**.** A choice sequence C satisfies the *sequential weak axiom of revealed preference with respect to $B = B_{t_{t=0}}^{\infty}$* (WARP-S(B)) if the following property holds:

for any $x, y \in X$ and for any t and h^t,

$$x\, \mathbf{P_B^t}\, y \Longrightarrow y \neq C_{t'}\, \forall t' > t\, such\, that\, x, y \in S_{t'}.$$

From definition of WARP-S(B), I characterize the choice function C generated by $(\succ, B)$ in terms of WARP-S(B).

Proposition 1. For some $B = \{B_t\}_{t=0}^{\infty}$ C satisfies WARP - S(B) if and only if C is a CBL.

This result also allows us to find some underlying preference $\succ$ that generates C as a result of her binary learning. Following result dictates this finding formally.

Corollary 1. C satisfies WARP - S(B) if and only if there exists some $\succ \in \mathbb{P}_{\succ}$such that C is generated by ($\succ$, B).

We will briefly see how the above result can be used. Consider the example 1 again.

Example. 1 (Revisited) Let $X = \{a, b, c\}$. Assume that $(S_0, C_0) = (\{a, b, c\}, c)$, $(S_1, C_1) = (\{a, b, c\}, b)$, $(S_t, C_t) = (\{a, b\}, a)$ for any $t \geq 2$.

First, we will see WARP - S(X) is satisfied for this choice sequence. From $t = 0$ to 1, we have $b\ \mathbf{P_X^1}\ c$ since a, b are both available at $S_1 = \{a, b, c\}$ and $C_0 = c$ and $C_1 = b$. By the same way, we have $a\ \mathbf{P_X^2}\ b$ since $S_1 = \{a, b\}$ and $C_1 = b$ and $C_2 = a$. Also we have $b\ \mathbf{P_X^2}\ c$ that is inherited from $b\ \mathbf{P_X^1}\ c$. From the next period, we have $\mathbf{P_X^t} = \{(a, b), (b, c)\}$ and decision maker stays at a. So it does not violate WARP - S(X).

Now we see how theorem 1 works. Put $\mathbf{P_X} = \lim_{t\to\infty} \mathbf{P_X^t}$.[5] Consider some underlying preference that includes $\mathbf{P_X}$. In this example, we have $\mathbf{P_X} = \mathbf{P_X^2}$. Since the underlying preference requires to be complete, transitive, and asymmetric, the only candidate for underlying preference including $\mathbf{P_X}$ is $\succ = \{(a, b), (b, c), (a, c)\}$. As we can see in the above, C is generated by ($\succ$, X).

Theorem guarantees existence of underlying preference $\succ$ that explains given choice function C as a result of learning. In other words, by showing satisfaction of WARP - S(B), we can find at least one complete, transitive, and asymmetric underlying preference $\succ$ such that C is generated by ($\succ$, B).

This result points out the connection between temporal consistency and well-defined[6] (underlying) preference. WARP - S(B) requires the consistency based on sequential choice: revealed preference

[5]For existence of this limit, we need additional proof. This issue will be considered in the section 4. In this example, it exists and $\mathbf{P_X} = \mathbf{P_X^2}$.

[6]To indicate complete, transitive, and asymmetric underlying preference, I abuse the word 'well-defined.'

at each time should not conflict with upcoming choice outcomes. This sequential consistency does not require well-defined preference at given time. It only requires the partial preference that does not conflict with upcoming choice outcomes. However, WARP - S(B) suffices such sequential consistency to be continued until the end of choices, and this condition is necessary and sufficient condition to guarantee existence of well-defined preference.

Furthermore, such sequential consistency largely depends on choice of B. The same choice sequence can be valid for WARP - S with respect to some B, but not for another one. This finding make us consider the connection between the structure of range of learning and requirement for existence of well-defined preference. Wider range of learning more effectively narrows down the candidate of well-defined preference that is compatible to given sequential choice. On the other hand, such wider range of leaning restricts the possible choices that can be justified as not violating sequential consistency. That is, as we have more effective prediction in the sequential choice, requirement for sequential consistency comes to be more restrictive. We will discuss about this link more deeply at the section 5.

2.3.1 Proof of Proposition

In this subsection, I provide a proof for Proposition 1. The proof consists of a series of lemmas which will be also useful for the next section's result.

Lemma 1. Assume a choice function C satisfies WARP-S(B) and $\mathbf{P_B^t}$ is a revealed preference at t with respect to B. Then, there exists a *limit revealed preference* with respect to B $\mathbf{P_B} \equiv \lim_{t\to\infty} \mathbf{P_B^t}$.

Proof. Suppose that a choice function C satisfies WARP-S(B) and $\mathbf{P_B^t}$ is a revealed preference at t with respect to B. For the proof of this lemma, we will first show that for any x, $y \in X$ and $t \geq 0$ such that $x\mathbf{P_B^t}y$, $x\mathbf{P_B^j}y$ for any $j \geq t$ and $\neg(y\mathbf{P_B^k}x)$ for any $k \geq 0$. From this fact, the assumption that C satisfies WARP-S(B) implies $\mathbf{P_B^t}$ weakly increases over time. As a result, we may have $\lim_{t\to\infty} \mathbf{P_B^t} = \bigcup_{t\to\infty} \mathbf{P_B^t}$.

(WTS) Suppose that $x\mathbf{P}_{\mathbf{B}}^{\mathbf{t}}y$ where $x, y \in X$ and $t \geq 0$. Then, for any k, $\neg(y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x)$. Moreover, for any $j \geq t$, $x\mathbf{P}_{\mathbf{B}}^{\mathbf{j}}y$.

(Proof) By the definition of revealed preference, $x\mathbf{P}_{\mathbf{B}}^{\mathbf{t}}y$ requires some (i) $r < s \leq t$ such that $\{x, y\} \in B_r$,$C_r \in \{x, y\}$, $C_s = x$ or (ii) some $w \neq x, y \in X$ such that $x\,\mathbf{P}_{\mathbf{B}}^{\mathbf{r}}\,w$ and $w\,\mathbf{P}_{\mathbf{B}}^{\mathbf{s}}\,y$ where $r\,, s \leq t$, and some $t' \leq t$ such that $C_{t'} \in \{x, y, w\}$ and $\{x, y\} \in B_{t'}$.

For case (i), suppose that s is the least time such that x is revealed preferred to y. Then, by definition, $x\mathbf{P}_{\mathbf{B}}^{\mathbf{j}}y$ for any $j \geq s$. Since we assume that s is the least time that x is revealed preferred to y, we may not have $y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x$ for any $k \leq s$. If we have, then it violates WARP-S(B) at s. By the similar way, for any $k > s$, having $C_k = y$ allows to have $y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x$ but it also violates WARP-S(B) at k. Consequently we have $\neg(y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x)$ for any k.

For case (ii), suppose that $v \equiv \max\{r, s, t'\}$ is the least time such that x is revealed preferred to y. By definition, $x\mathbf{P}_{\mathbf{B}}^{\mathbf{j}}y$ for any $j \geq v$. To show that $\neg(y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x)$ for any $k \geq v$, suppose that there exists some $t" \geq v$ such that $C_{t"} = y$ and $\{x, y\} \subseteq S_{t"}$. Then, it violates WARP-S(B) at $t"$ so that it is contradictory. To show that $\neg(y\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}x)$ for any $k < v$, suppose that there exists some $t" < v$ such that $y\,\mathbf{P}_{\mathbf{B}}^{\mathbf{t"}}x$. Without loss of generality, suppose that $r, t', t" < s$. Then, at s - 1, at least $y\,\mathbf{P}_{\mathbf{B}}^{\mathbf{s-1}}w$ so that we have $\neg(w\mathbf{P}_{\mathbf{B}}^{\mathbf{k}}y)$ which contradictory to assumption. For the other cases, we have similar contradiction. □

From the above existence of limit preference, we find the following proposition.

Proposition 2. Assume a choice function C satisfies WARP-S(B) and $\mathbf{P_B}$ be a limit revealed preference with respect to B. Then, for any $x, y \in X$ and any $\succ \in \mathbb{P}_{\succ}$ such that C is generated by ($\succ$, B),

$$(x, y) \in \mathbf{P_B} \Rightarrow x \succ y.$$

Proof. Assume that C satisfies WARP-S(B). From the above lemma, we have $\mathbf{P_B} = \bigcup_{t=0}^{\infty}\mathbf{P_B}^t$. Now we suppose $x\,\mathbf{P_B}y$ and $t \geq 0$ is the least time such that $x\mathbf{P_B}^t y$. Then, by WARP-S(B), there is no s

≥ 0 such that $(x, y) \in S_{t+s}$ and $y = C_{t+s}$. So that $\neg(y\mathbf{P}_{\mathbf{B}}^{\mathbf{s}}x)\ \forall s \geq 0$. Moreover, since $t \geq 0$ is the least time such that $(x, y) \in \mathbf{P_B}^t$, there is no $t' \leq t$ such that $(y, x) \in \mathbf{P_B}^{t'}$. Then $(y, x) \notin \bigcup_{t=0}^{\infty}\mathbf{P_B}^t$. $x\mathbf{P_B}^t y$ implies there exists (i) $r < s = t$ such that $\{x, y\} \in B_r$,$C_r \in \{x, y\}$, $C_s = x$ or (ii) some $w \neq x, y \in X$ such that $x\ \mathbf{P}_{\mathbf{B}}^{\mathbf{r}}\ w$ and $w\ \mathbf{P}_{\mathbf{B}}^{\mathbf{s}}\ y$ where $r\ , s \leq t$, and some $t' \leq t$ such that $C_{t'} \in \{x, y, w\}$ and $\{x, y\} \in B_{t'}$.

Now we assume that C is generated by $(\succ, B)$ for some arbitrary $\succ \in \mathbb{P}_{\succ}$ such that $(y, x) \in \succ$. For (i), by (Elimination), $C_r \in \{x, y\}$ and $\{x, y\} \subseteq B_r$ implies $(x, y) \notin \Gamma_b(\xi_{r+1})$ so that $(x, y) \notin \Gamma_b(\xi_{t-1}) = \succsim_t$. Then, at t, $x = C_t$ is not $\succsim_t$- best alternative since $y \in S_t$ while $(y, x) \in \Gamma_b(\xi_{t-1})$. For (ii), without loss of generality, assume that $x\ \mathbf{P}_{\mathbf{B}}^{\mathbf{r}}\ w$ and $w\ \mathbf{P}_{\mathbf{B}}^{\mathbf{s}}\ y$ is earned by the case (i) respectively and $r < s$. Now we check each case for relative order for t' and $C_{t'}$. First, suppose $t' < r < s$ and $C_{t'} = x$ or y. Then, by (Elimination) at t', $(x, y) \notin \Gamma_b(\xi_{t'+1})$ so that $(x, y) \notin \Gamma_b(\xi_{r-1}) = \succsim_r$. Then, at r, having $C_r = x$ violates $\succsim_r$ since y is available at S_r while $(y, x) \in \succsim_r$. Second, suppose that suppose $t' < r < s$ and $C_{t'} = w$. Then, by (Transitivity) at t', $(x, y) \notin \Gamma_b(\xi_{t'+1})$ so that $(x, y) \notin \Gamma_b(\xi_{r-1}) = \succsim_r$. Similar to the first case, it violates $\succsim_r$ so that it is contradictory. By exploiting these tricks to the other cases, we have that assuming $(y, x) \in \succ$ always violates the definition of CBL. □

This proposition allows us to have weak acyclic of revealed preference.

Proposition 3. Assume a choice function C satisfies WARP-S(B) and $\mathbf{P_B}$ is a revealed preference with respect to B. Then, a *transitive closure* of $\mathbf{P_B}$ $tc(\mathbf{P_B})$ such that

$$\begin{aligned} tc(\mathbf{P_B}) \quad = \quad & \{(x, y) \in X \times X \,|\, \exists K \in \mathbb{N}\, and\, x^0, x^1, \cdots, x^K \in X\, such\, that \\ & x = x^0, (x^{k-1}, x^k) \in \mathbf{P_B}\, for\, all\, k \in \{1, \cdots, K\}\, and\, x^K = y\} \end{aligned}$$

is weakly acyclic. That is, for any $x, y \in X$ such that $(x, y) \in tc(\mathbf{P_B})$, $(y, x) \notin \mathbf{P_B}$.

Proof. For this proposition, I only consider the minimal transitive closure such that $(x, z) \in tc(\mathbf{P_B})$ from $(x, y) \in \mathbf{P_B}$ and $(y, z) \in \mathbf{P_B}$ where $x, y, z \in X$. Proof of longer length of transitive closure can be inductively induced from the extension of this minimal case. Suppose $(x, y) \in \mathbf{P_B}$ and $(y, z) \in$

$\mathbf{P_B}$. Then, we may have the least r, $s \geq 0$ such that $(x, y) \in \mathbf{P_B^r}$ and $(y, z) \in \mathbf{P_B^s}$. (1) Suppose that $(x, z) \in \mathbf{P_B}$. Then, from the above proposition, $(x, z) \notin \mathbf{P_B}$ is automatically induced. (2) Suppose that $(x, z) \notin \mathbf{P_B}$. Then, there is no $t \geq 0$ such that $(x, z) \in \mathbf{P_B^t}$. Now assume that $(z, x) \in \mathbf{P_B^{t'}}$ for some $t' \geq 0$. Let $v = \max\{r, s, t'\}$. Without loss of generality, let $v = t'$. To have $(z, x) \in \mathbf{P_B^{t'}}$, it needs some t" $< t$' such that $C_{t"} \in \{x, z\} \subseteq B_{t"}$. However, having such t" $< t$' automatically induces $(x, z) \in \mathbf{P_B^{max\{r,s,t"\}}}$ which implies contradiction. □

As a result of such propositions and lemma, we have Proposition 4.

Proposition 4. *C* satisfies WARP - S with respect to some range of learning $B = \{B_t\}_{t=0}^{\infty}$ if and only if *C* is a CBL.

Proof. The case that CBL implies WARP-S(B) is immediate from the construction of CBL. So, we only consider the case that WARP-S(B) implies CBL.

Suppose *C* satisfies WARP-S(B). This assumption returns that, for any $x, y \in X$, if $(x, y) \in \mathbf{P_B}^t$ for some t, then $(y, x) \notin \mathbf{P_B}^s$ for any s. So, for any $x, y \in X$ such that $(x, y) \in \mathbf{P_B}^t$ for some t, we have $(x, y) \in \mathbf{P_B}$ and $(y, x) \notin \mathbf{P_B}$. Then, we can find some set of underlying preferences $\succ_B \equiv \{ \succ \in \mathbb{P}_{\succ} | \mathbf{P_B} \subseteq \succ \}$. Since $tc(\mathbf{P_B})$ is weakly acyclic according to Proposition 3, $\succ_B$ is should be nonempty.

(WTS) Pick any $\succ \in \succ_B$. Then, for any Γ_b that equips $(\succ, B)$, $C_t \in \{x \in S_t |$ there exists no $y \neq x \in S_t$ such that $y \succ_t x$ where $\Gamma_b(\zeta_{t-1}) = \succsim_t \}$ for any t.

(Proof) Suppose not. That is, at some t, $C_t = x$, $y \succ_t x$, and $y \neq x \in S_t$. Then, we have some $s \leq t$ - 1 such that $(x, y) \in \Gamma_b(\zeta_{s-1})$ and $(x, y) \notin \Gamma_b(\zeta_s)$. Consequently $(x, y) \notin \Gamma_b(\zeta_s)$ for some $s \leq t$ - 1 implies $(x, y) \notin \succ$. On the other hand, according to Proposition 2, WARP-S(B) implies $(x, y) \in \mathbf{P_B}^t$ so that $(x, y) \in \mathbf{P_B} \subseteq \succ$, which is contradictory. □

2.4 Range of Learning and Sequential Consistency

In this section, I discuss about the way to restrict the range of learning that can recover well-defined underlying preference from choice sequence.

2.4.1 Range of Learning and Requirement for Sequential Consistency

First, larger range of learning implies the decision maker learns more about pairwise underlying preference at each time. As a result, the revealed preferences will weakly increase at each period. In other words, the monotonicity in the range of learning is continued to the weak monotonicity of revealed preference. The following lemma formally depicts this implication.

Lemma 2. Assume a choice function C generated by $(\succ^0, B^0)$ and $(\succ^1, B^1)$ where $B^0 = \{B^0_t\}^{\infty}_{t=0}$ and $B^1 = \{B^1_t\}^{\infty}_{t=0}$ respectively. $\mathbf{P^t_{B^0}}$ and $\mathbf{P^t_{B^1}}$ is revealed preference at t from each pair respectively. Then,

$$B^0_t \subseteq B^1_t \; for all t \Rightarrow \mathbf{P^t_{B^0}} \subseteq \mathbf{P^t_{B^1}} \; for all t.$$

Proof. Suppose that $(x, y) \in \mathbf{P^t_{B^0}}$ where $x, y \in X$ at some t and $B^0_t \subseteq B^1_t$ for all t. We consider three cases of revealed preference realizations respectively.

(1) Suppose there are $r \leq s \leq t$ such that $y = C_r$, $x = C_s$, $x, y \in B^0_r$, and $x, y \in S_s$. Then we have $x, y \in B^1_r$ so that, by definition, $(x, y) \in \mathbf{P^t_{B^1}}$.

(2) Suppose there are $r \leq s \leq t$ such that $x = C_r$, $x = C_s$, $x, y \in B^0_r$, and $x, y \in S_s$. Then we have $x, y \in B^1_r$ so that, by definition, $(x, y) \in \mathbf{P^t_{B^1}}$.

(3) Suppose there is an alternative $w \in X$ such that $(x, w) \in \mathbf{P^r_{B^0}}$ and $(w, y) \in \mathbf{P^s_{B^0}}$ where $r, s \leq t$, and there is $t' \leq t$ such that $C_{t'} \in \{x, y, w\}$ and $x, y \in B^0_{t'}$. Without loss of generality, $(x, w) \in \mathbf{P^r_{B^0}}$ and $(w, y) \in \mathbf{P^s_{B^0}}$ is realized according to (i) or (ii) of Definition 4 respectively. Then, by exploiting

proof in (1) or (2) above, we have $(x, w) \in \mathbf{P^r_{B^1}}$ and $(w, y) \in \mathbf{P^s_{B^1}}$. By the assumption, $x, y \in B^0_{t'}$ implies $x, y \in B^1_{t'}$. Then, by the (iii) of Definition 4, $(x, y) \in \mathbf{P^t_{B^1}}$. □

With this lemma, I discuss about the connection between the revealed preference and underlying preference. Main idea is this: each revealed preference is constructed to capture the sequential choice behavior that explicitly reveals the underlying preference. That is, at least each revealed preference can be confirmed as the reflection of underlying preference of the decision maker if she truly adopted the range of learning. WARP - S enforces such revealed preference to be consistent with her upcoming choice sequence. This enforcement assures the existence of well-defined underlying preference that allows revealed preferences as its partial reflection.

From those two results, we have the following result that governs the connection between consideration sets and underlying preferences. Wider range of learning allows us to catch more revealed preferences. And proposition above justifies revealed preferences as a proper reflection of the underlying preference if they are consistent to upcoming choice outcomes. These two results can be combined to have result that shapes the least amount of underlying preference according to relation between the ranges of learning. Following corollary formalizes this explanation.

Corollary 2. Assume a choice sequence C generated by $(\succ^0, B^0)$ and $(\succ^1, B^1)$. Then, for all $x, y \in X$,

$$B^0_t \subseteq B^1_t \; for all \; t \Rightarrow \left[x \mathbf{P_{B^0}} y \Rightarrow x \succ^1 y \right].$$

This result suggests the basic criterion we can exploit to confirm the minimum level of prediction for underlying preference. When we have enough (a priori or exogenous) information to determine the decision maker's range of learning, this finding allows us to exploit that information to set up the boundary for prediction level. Suppose that we have some exogenous reason to confirm that some reference set I is always included to her range of learning. Then, forming her range of learning to include I every time will guarantee the revealed preference relations $\mathbf{P_I}$ generated by $B_t = I$ as her underlying preference (as long as choice function satisfies WARP - S with respect to I.)

These results give us another way to finding proper range of learning. In this paper, I assume that the range of learning to be general function of sets generated from history of choice environment. For this generality, there is difficulty in guessing a proper form of function. Proposition (x)] allows us to try arbitrary sequences of sets as range of learning. If we find some sequence of sets to satisfy WARP - S, then we can narrow down proper candidates for range of learning to include such sequence of sets at each time.

Moreover, this connection exhibits continuous prediction level depends on given choice data's level of sequential consistency. As I briefly mentioned in section 3.2, larger range of learning needs tighter requirement for existence of well-defined underlying preference that generates given choice function as CBL. Since wider range of learning confirms more pairs as revealed preference than smaller one, revealed preference from larger one has higher possibility to conflict with upcoming choice outcomes. On the other hand, more pairs in revealed preference implies higher predictability for underlying preference. This results allows us to make a choice between predictability and consistency. For some choice data, even it is not fully fit to some high level of sequential consistency, we can find another (lower) level of sequential consistency that also provides another level of predictability. For this matter, CBL framework is very flexible to different level of sequential consistency while maintaining proper level of predictability.

The following corollary depicts this finding.

Corollary 3. Fix the choice function C and ranges of learning B^0, B^1. $\mathbf{P_{B^0}}$ and $\mathbf{P_{B^1}}$ is revealed preference from B^0 and B^1 respectively. Suppose that $B^0_t \subseteq B^1_t$ for all t. Then,

$$[C \textit{ satisfies } WARP - S(B^0) \Rightarrow C \textit{ satisfies } WARP - S(B^1)] \Leftrightarrow [\mathbf{P_{B^0}} \subseteq \mathbf{P_{B^1}}].$$

By combining [the above proposition] and [the lemma] we have a [theorem] that determines whether the certain range of learning can generate the CBL choice function.

Theorem 1. Fix $B^0 = \{B^0_t\}^\infty_{t=0}$ and $B^1 = \{B^t_t\}^\infty_{t=0}$ such that $B^0_t \subseteq B^1_t$ for all t. Suppose $tc(\mathbf{P_{B^0}})$ is not weakly acyclic. Then, there is no $\succ \in \mathbb{P}_\succ$ such that C is generated by $(\succ, B^1)$.

Proof. By the Proposition 3, if $tc(\mathbf{P_{B^0}})$ is not weakly acyclic, then C is not satisfying WARP - S(B^0). Then there is some s, s' where $s < s$' and alternatives $x, y \in X$ such that $(x, y) \in \mathbf{P^s_{B^0}}$ while $y = C_{s'}$ and $x, y \in S_{s'}$. From the [lemma 2] above, we have $(x, y) \in \mathbf{P^s_{B^1}}$. So C also violates WARP - S(B^1) which implies that there are no $\succ \in \mathbb{P}_{\succ}$ such that C is generated by $(\succ, B^1)$. □

This theorem restricts the range of learning by the inclusion relation. That is, once a revealed preference from a certain range of learning violates the sequential consistency (or WARP-S), any range of learning that includes such range of learning also violates the sequential consistency. This result makes us to save an effort to find a 'reasonable' range of leaning. As long as we have some non-empty range of learning that its limit revealed preference violates weak acyclicity, any other non-empty range of leaning that includes such range of learning can be excluded from our consideration.

2.5 Comparison to Previous Choice Set Based Frameworks

In this section, I will compare choice with learning framework (henceforth, CL) to previous choice set-based frameworks, especially Masatlioglu and Nakajima (2012) (henceforth M&N) and Caplin and Dean (2011) (henceforth C&D)[7]. To help reader's understanding, I will briefly summarize frameworks in two previous papers. Then, I will compare three frameworks with CBL framework.

Set-based frameworks focus on characterization of the choice sequences that explains choice reversal occurs within the same or fixed choice set. Those frameworks consider that the decision maker's changing consideration set or range of focus causes choice reversal over time. Assuming that the decision maker has fully known preference over all alternatives, they separate actual choice set and nominal choice set the decision maker faces. That is, even the decision maker faces

[7]For the comparison of model-based frameworks and model-free (or standard) frameworks, see Masatlioglu et al (2012).

superficially the same choice set, underlying choice set can be differ at each time.

Assuming such choice environment, set-based frameworks focuses on recover the decision maker's underlying preference by infer actual choice set. Depending on the generation process of (sequences of) actual choice set, their recovery can be different prediction from standard model-free framework[8].

However, set-based frameworks does not consider the order of choice sequences those occur on the difference choice sets. That is, they focus on recovery of underlying preferences that is revealed only from the same or fixed choice sets. However, if the choice sequence is combined of choice sequences from different choice sets and such whole sequential information is available, we can extract more information on the decision maker's preference.

Masatlioglu and Nakajima (**2012**)

Masatlioglu and Nakajima (2012) introduce a *choice by iterative search* (CIS) to capture the underlying preference in the sequential choice situation. They assume that the decision maker chooses the best alternative in her current consideration set at each time. Given that she knows her complete and transitive underlying preference without indifference, consideration set restricts her choice set itself. Even though the decision maker nominally faces choice set S_t, actual choice set is consideration set $B_t \subseteq S_t$ is different from S_t. This notion of consideration set is main difference from CL framework. I assume the consideration set as the decision maker's range of learning, not the range of actual choice. CL framework allows all the alternatives in the choice set to be considered for choice.

Another difference is placed on the cause of choice reversal. In their setting, it is assumed that the decision maker's consideration set changes over time depends on her current consideration set. In other words, next period's consideration set is supposed to be generated from current consideration

[8]To see more detailed discussion on this difference, see Masatioglu et al (2012).

set. They call such consideration set generation process as choice by iterative search (CIS) process. Because of CIS process, even though she faces the same choice set for a few periods, changes in her consideration sets can cause changes in choice outcomes. CIS process consequently enforces path of consideration set to be fixed according to her initial choice outcome and choice set. It means, if the decision maker starts from a particular initial choice outcome and a choice set, choice sequence should be exactly the same irrelevant to her choice history on another choice set.

Authors use the expression $C(S, x) = y$ to denote choice outcome from choice set S that started from x and end up with y. With my notation, it can be written in $(S, C_t)_{t=0}^{T}$, where $T < \infty$, with $C_0 = x$ and $C_T = y$. The final choice (C_T) is assumed as the most preferred alternatives in $\bigcup_{t=0}^{T} S_k$. That is, the decision maker stops her searching when she arrived at her best choice. This observation allows them to define revealed preference $\succ_c$ such that $y \succ_c x$ if $C(S,x) = y$.

To characterize the CIS process, they introduced a *dominating anchor axiom* defined by the following[9].

Definition 6 (Dominating Anchor)**.** For any finite set $S \in \mathbb{X}$, there exists some alternative $x^* \in S$ such that $C(\mathrm{T}, x^*) \notin S\backslash\{x^*\}$ for all T including x^*.

From the dominating anchor axiom, they characterize the choice function as the *choice by iterative search* (CIS). Moreover, axiom imposes acyclicity on $\succ_c$.

Theorem 2 (Choice by Iterative Search)**.** A choice function $C(\cdot, \cdot)$ obeys the dominating anchor axiom *if and only if* $C(\cdot, \cdot)$ is a CIS. Moreover, underlying preference $\succ_C$ generating C *if and only if* $\succ$ includes a transitive closure of $\succ_c$ $tc(\succ_c)$.

Since the CIS process covers very broad range of choice sequences, they attach additional structure on the generation process of consideration sets. They assume the next period's consideration set to be union of unique set of alternatives that contains current period's choice outcome and current consideration set itself. Actually, this structure depicts the (unique) expansion of consideration set depending on current choice outcome and consideration set. This process is called as

[9] I'm exploiting original paper's expression to maintain its conceptual philosophy.

the *Markovian choice by iterative search* (MCIS)[10]. MCIS identifies two additional cases as the revealed preference: $x \succ_c y$ if (i) $x = C(S\cup\{y\}, z) \neq C(S, z)$, or (ii) $x = C(S\cup\{y\}, z)$ and $y = C(\{y, z\}, z)$. The first case captures the bifurcation of path of consideration sets. Since MCIS process depends on choice set and current consideration set, the change of choice set itself can cause changes in consideration set generating process. For this reason, consequential change in path of consideration sets can cause different final choice outcome. The second case captures part of transitivity among choice outcomes. If some alternative, say y, is revealed to prefer another one, say z, and x is revealed to preferred to z even y was available, then x is also revealed to preferred to y.

Caplin and Dean (2011)

Caplin and Dean (2011) introduce an *alternative based search* (ABS) to explain choice reversal as the consequence of searching behavior. They assume given choice sequence to have infinite length and the decision maker to stay at her best alternative once she found it.

They define sequence of consideration sets to be non-decreasing and subset of choice set. Formally, $B_t \subseteq A$ and $B_t \subseteq B_{t+1}$ for all t, where A is a choice set. It is the same to the setting in MCIS of M&N. While MCIS bases on fixed consideration set generation process, they assume those consideration sets to be simply non-decreasing. For this reason, they call such consideration set as ever-searched set. They define the set of sequences of the non-decreasing sets as $\mathscr{Z}^{ND}$ and $\mathscr{Z}_A \in \mathscr{Z}^{ND}$ to comprise of (consideration) sets B_t selected from $A \in X$,

$$\mathscr{Z}_A = \{Z \in \mathscr{Z}^{ND} | B_t \subset A \, for all \, t \geq 0\}.$$

Also they defined the t th choice outcome from $\mathscr{Z}_A$ $C_A(t) \in A$. Without harming the philosophy of original notion, I introduce the alternative based search (ABS) representation as the following :

Definition 7 (ABS)**.** Choice sequence $(A, C_t)_{t=0}^{\infty}$ has an ABS representation $(\succ, S)$ if there exists a

[10]Authors characterized MCIS process with additional properties including strong version of dominating anchor axiom.

underlying preference $\succ$ and a search correspondence $S : \mathbb{X} \to \mathscr{Z}^{ND}$, with $S_A \in A$ for all $A \in \mathbb{X}$ such that

$$C_A(t) = \{x \in S_A(t) \,|\, x \succ y \, for \, all \, y \in Z_t\}.$$

ABS representation explicitly assumes the choice outcome to be the best alternative among the searched alternatives. In other words, as in M&N, C&D assume fully known underlying preference over consideration set. For this reason, ABS does not allow choice reversal between ever-searched alternatives. The only possible case that choice reversal can happen is when newly considered or added item to consideration set is better than current choice outcome. In other words, if the choice reversal happens between two alternatives, say x and y, then no choice reversal is allowed to happen again between them in the upcoming choice sequence.

And the assumption allows them to consider every choice reversal as the evidence of revealed preference. That is, once choice outcome is moved from x at time t to y at $t + s$ for any t, $s \geq 1$, the decision maker is considered to reveal her underlying preference $x \succ y$. This finding allows revealed preference relation $\succ_{ABS}$ on X to be defined by $x \succ_{ABS} y$ if there exists $A \in X$ and s,$t \geq 1$ such that $y \in C_A(s)$ and $x \in C_A(s + t)$, but $y \notin C_A(s + t)$.

2.5.1 Application of Sequential Order of Choices

In this subsection, I will show how can we use whole choice sequence to extract the decision maker's underlying preference. Consider that following example.

Example 2. Let $X = \{a, b, c, d, e\}$. Assume that $(S_0, C_0) = (X, e)$, $(S_1, C_1) = (\{b, c, d\}, e)$, $(S_2, C_2) = (X, c)$, $(S_3, C_3) = (\{b, c, d\}, b)$, $(S_4, C_4) = (X, a)$. From this whole choice sequence C, I define subsequences for different choice set X and $\{b, c, d\}$ and denote them C^X and C^{bcd} respectively. That is, $(S_0^X, C_0^X) = (S_0, C_0) = (X, e)$, $(S_1^X, C_1^X) = (S_2, C_2) = (X, c)$, and so on. (S^{bcd}, C^{bcd}) is defined similarly.

According to ABS, we can find two transitive closures $a \succ_{ABS} c \succ_{ABS} e$, and $b \succ_{ABS} d$. Given choice sequence C^X, the sequence ends up with a through c and e. Assuming that this sequence is the only available choice sequence within X, a pair $(X, \succ_{ABS_1})$, where $\succ_{ABS_1}$ is an arbitrary underlying preference order that includes $a \succ_{ABS} c$ and $a \succ_{ABS} e$, is qualified as ABS representation that generates C^X. As the same reason, a pair $(\{b, c, d\}, \succ_{ABS_2})$, where $\succ_{ABS_2}$ is an arbitrary underlying preference order that includes $b \succ_{ABS} d$, is qualified as CIS that generates C^{bcd}. Finally, a pair $(X, \succ^*_{ABS})$, where $\succ^*_{ABS}$ is an arbitrary preference order includes $tc(\succ_{ABS})$, is qualified as ABS representation for entire sequence C. MCIS recovers exactly same two transitive closures. However, ABS and MCIS both can't tell anything about relation between a and b or c and d. According to CL, we can have candidates for entire underlying preference. Assume $B_t = X$ for all t. C satisfies WARP - S(X) and ($\succ = a \succ b \succ c \succ d \succ e$, X) uniquely generates C as CBL. For the case $B_t = S_t$ for all t, C satisfies WARP - S(S) and ($\succ = a \succ b \succ c \succ d \succ e$, S_t) also generates C as CBL.

Example 2 exhibits how CL framework extracts additional information about the underlying preference from whole choice sequences. Why such difference comes out in the same choice sequence? The reason comes from the conceptual difference between set-based frameworks and CL framework. Set-based frameworks defines the preference relations on the base of (actual) choice set. They define each choice outcome as the most preferable alternative within some (actual) choice set. In other words, each choice outcome is considered as the reflection of fully known underlying preference at each (actual) choice set. For this reason, sequential order between different choice sets does not affect choice outcomes.

On the while, CL framework considers that the decision maker's temporal preference is affected by her past choice history. Even I assume that her underlying preference is fixed, her temporal preferences which partially reflect her underlying preference is shaped by her past choice sequence. Depending on what she had faced and experienced before her current preference is formed and affects her current decision. On this ground, I can define the decision maker's sequential consistency in the form of WARP - S and extract the additional information on her underlying preference. For this reason, additional information on underlying preference is differed by whole choice sequence's

consistency. Consider the following example.

Example 3. Let $X = \{a, b, c, d, e\}$. A choice sequence C, C^X, and C^{bcd} is defined as in Example 2. Now I define concatenation of C^X and C^{bcd} and denote $C^{'}$. For $C^{'}$, $(C^{'}_0, S^{'}_0) = (C^X_0, S^X_0) = (X, e)$, $(C^{'}_1, S^{'}_1) = (C^{bcd}_1, S^{bcd}_1) = (X, c)$, $(C^{'}_2, S^{'}_2) = (C^X_2, S^X_2) = (X, a)$, $(C^{'}_3, S^{'}_3) = (C^{bcd}_0, S^{bcd}_0) = (\{b, c, d\}, c)$, $(C^{'}_4, S^{'}_4) = (C^{bcd}_1, S^{bcd}_1) = (\{b, c, d\}, b)$.

ABS and MCIS can recover the same underlying preference on both C and C'. Since they consider the sequential consistency on the set-wise ground, unless the recovered underlying preference from each choice set does not conflict to another, both choice sequences are considered as consistent one.

On the while, CL framework defines different consistency for each choice sequence. For C, WARP - S(X) and WARP- S(S) are satisfied. However, for C', WARP - S(X) is not satisfied. WARP - S(S) is still satisfied for C', but recovered underlying preference is quite different from that from C. According to $B_t = S_t$ for all t, an arbitrary underlying preference $\succ^{'}$ that includes $a \succ c \succ e$ and $b \succ d \succ c$ can be paired with S_t to generate $C^{'}$. That is, depends on whole sequence's sequential order, extent of additional information can be differed and her range of learning can be specified.

This example briefly exhibits how the whole sequence's sequential order affect defining sequential consistency and how such consistency affect the recovery of underlying preference. According to CBL, sequential consistency is defined on whole sequence's sequential order. This relation of sequential order and consistency is captured by WARP - S. WARP - S explicitly restricts sequential consistency in terms of range of learning. By the range of learning, the recovery of underlying preference is defined in terms of revealed preference[11].

On the other hand, set-based frameworks provide the same recovery of underlying preference regardless of whole choice sequence's order. They define the whole choice sequence's consistency based on each set-wise sequence's consistency. In other words, they define whole choice sequence consistent if subsequences that are defined on each fixed choice set exhibits consistency and preference order derived from each choice sets is not in conflict with that from the other choice sets.

[11] Detailed discussion on this relation between range of learning and revealed preference will be provided in the section 5.

This difference raises question about the usage of whole choice sequence's order: whole sequential data can be used to make up insufficient set-wise sequential choice data. Set-based frameworks not requires the whole sequence's order. That is, they can define consistency that are robust to whole choice sequence's order. On the other hand, as a result of such context-free property, they hesitate to determine some relations as the revealed preference. If set-wise data on decision maker's preference is rich enough, such lost can be compensated by sharing recovery from different choice sequences. However, if set-wise choice data is not enough for such make-up, we can use whole sequential choice's order to make up insufficient information about individual's underlying preference.

2.5.2 Repeated Choice Reversal and Procedural Information

Compared to set-based frameworks, CBL framework can explain repeated choice reversal. Choice reversal is the situation at which the decision maker change her choice from one to another alternatives when two alternatives are both available. Repeated choice reversal especially indicates the case at which such choice reversal between a pair of alternatives occurs more than or equal to two times. Set-based frameworks do not allow such repeated choice reversal : since they assume that the decision maker's underlying preference is fully known and her next period's actual choice set always includes the current choice outcome, such repeated choice reversal cannot happen. However, CL frameworks restrictively allows repeated choice reversal depends on its range of learning. Consider the following example.

Example 4. Let $X = \{a, b\}$ and underlying preference $a \succ b$. Suppose that choice sequence C consists of $(S_0, C_0) = (\{a, b\}, a)$, $(S_1, C_1) = (\{a, b\}, b)$, $(S_t, C_t) = (\{a, b\}, a)$ for all $t \geq 2$. Since C exhibits repeated choice reversals, neither ABS nor MCIS consider it as consistent. From $X = S_t = \{a, b\}$ for $t = 1, 2$, WARP - S(X) and WARP - S(S) are both violated.

However, this inconsistency is explained by assuming $B_t = \bigcup_{k=0}^{t-1} \{C_k\}$ for all $t \geq 1$ and $B_0 = \{\emptyset\}$. This case assumes that the decision maker only learn the underlying preference between past experi-

enced alternatives and current choice outcome. In other words, the decision maker only learns from her direct experience. I denote WARP - S that uses this range of learning WARP - S(C).

$(\mathbf{0})\Gamma(\succ, \succsim_0, a, \{\emptyset\}) = \succsim_1 = \succsim_0$. At $t = 0$, no learning occurs. Since a and b are not yet experienced, decision maker cannot confirmed her preference. $(\mathbf{1})$ $\Gamma(\succ, \succsim_1, b, \{a\}) = \succsim_2 = \{(a, b)\}$. After her experience of both a and b, she can confirm which one is better. $(\mathbf{2})$ $\Gamma(\succ, \succsim_t, a, \{a, b\}) = \succsim_{t+1} = \{(a, b)\}$ for all $t \geq 3$. From her learning at $t = 1$, she chooses her best choice a according to her preference.

This case distinguishes the difference of CL framework from set-based frameworks. This particular example sheds some light on theoretical justification of the empirical findings from Chu and Chu (1990) and Cox and Grether(1994). In those empirical works, when subjects face two different lotteries (with different combination of award and probability), they occasionally exchanges their current choice with another lottery even it is available before and worse in numerically expected amount. They found that such choice reversal is remained even after the subjects' first experience of choice reversal. However, choice reversal is largely reduced after their experience of (realization of) both lotteries. Especially, Chu and Chu found that only a small proportion of subjects exhibit choice reversal after third time repetition [12]. CL framework provides theoretical ground to explain such repeated choice reversal as a result from the decision maker's learning process.

Moreover, CL framework captures some detailed relationship among alternatives by expoliting the procedural observations.

Example 5. Let $X = \{a, b, c, d\}$. $S_0 = S_1 = \{a, b\}$, $S_2 = S_3 = \{c, d\}$, $S_4 = S_5 = \{a, c\}$, $S_t = \{b, c\}$ for all $t \geq 6$ and $C_0 = C_1 = a$, $C_2 = C_3 = c$, $C_4 = c$, $C_5 = a$, $C_6 = c$, $C_t = b$ for all $t \geq 7$. This choice sequence compatible with ABS, and WARP - S(S). From ABS , we can find $\succ_{ABS} = \{(a, c), (b, c)\}$. In this case, we cannot define relation between a and b or c and d. However, from CL framework, we can find $a \ \mathbf{P_S} \ b \ \mathbf{P_S} \ c \ \mathbf{P_S} \ d$.

[12] This finding was formalized as so called *frequence – based probability learning* theory. To see detailed model, see the Humprey (2006).

In this example, changing sequential choices (S_4, S_5) and (S_6, S_7) can give a different interpretation. In example 5, changing choice from $C_6 = c$ to $C_7 = b$ is accepted as learning a relation between b and c since they are not learned even after learning the relation between a and c in (S_4, S_5). However, when we assume that $S_4, S_5 = \{b, c\}$, $S_6, S_7 = \{a, c\}$, and $(C_4, C_5) = (c, b)$, showing $(C_6, C_7) = (c, a)$ is not accepted as learning. According to learning process, chaging $C_4 = c$ to $C_5 = b$ also updates the relation between a and c by the transitivity, so that C_6 should be a rather than c. On the other hand, ABS accepts all such cases as relevant choice sequences and consider such changes in choices as the search process. In other words, ABS process ignores the procedural information between choice sequences which can reflect decision maker's information state that can be differed depends on information gained in her previous choices.

2.5.3 Brief Discussion on the Frameworks

From the examples in the above, it seems there is no exact relationship between set-based framework and general CL framework. For set-based framework, it requires each choice sequence to be complete that has ends up with some final choice. Moreover, when choice sequence only contains incomplete choice data that does not have final choice, they ignores information from such incomplete choice data. With exchange of such conservatism for data acceptance, they suggests more reliable recovery of underlying preference. On the other hand, CL framework accepts such incomplete data if they contains enough procedural information. However, CL framework's recovery of underlying preference is more vulnerable to small change in procedural information or context. Moreover, current CL framework does not allow bifurcation of choice path and final choice that is accepted in the MCIS model since it assumed non-decreasing update for temporal preference. Even with such limitness of current study, I think we can explain such bifrucation of final choice when we assume "forgetable" memory on the individual preference in the future study.

2.6 Conclusion

In this study, I constructed *choice with binary learning* (CBL) framework that captures sequential consistency that is governed by binary learning rule. I assumed that the decision maker's temporal preference (partially) reflects her underlying preference at that time but separated from that. Her sequential behavior is assumed to shape her temporal preference that weakly converges into her underlying preference. Sequential consistency requires her to choose the best and available alternative(s) according to her temporal preference at each time.

I focused on a simple learning rule with the least amount of rationality, and then placed all the other causes of non-standard behavior on the idiosyncrasy of range of learning. By doing this, we could narrow down our focus on the having the proper range of learning that justifies given choice function as CBL. According to its range of learning, we have the corresponding level of predictability for underlying preference without losing the decision maker's rationality.

Consequently, flexibility of sequential consistency has the two opposite side of implication. It provides us the ground to justify non-standard behavior as a result of sequential learning over time. It attributes reason of such non-standard behavior to the decision maker's temporal indifference, which mainly comes from "insufficient experience/learning." However, it also justify somewhat "totally irrational" behavior as result of rational behavior. It only blames such irrationality to the decision maker's idiosyncrasy. This issue is deeply connected to intrinsic question of our field of decision theory : what draws the rationality of human behavior?

Besides such philosophical question, I was able to find the explicit relationship between the predictability and sequential consistency. For this purpose, I characterize sequential weak axiom of revealed preference. Moreover, by using this identification, I could partially infer the underlying preference that the decision maker would had behind of her learning.

Understanding such sequential consistency gives us two implications. First, CBL framework allows us to explain the multiple-time choice reversal without losing well-defined underlying preferences' properties : transitivity and completeness. Even though there are philosophical issues,

we could theoretically rationalize repeated choice reversals as the result of the trial-and-error derived from learning process. This multiple-time choice reversal was not explained in the previous frameworks.

Second, it can improve recovery of underlying preferences by tightening sequential consistency. I suggested detailed characterization of sequential consistency according to the range of learning on which alternatives are compared to each other. As a result of characterization, we can capture weakly larger set of relation compare to previous frameworks.

Even though I found improvement in those aspects, there are limitations of this work. First, CBL framework itself is still capturing too wide range of choice sequences. Rationality of sequential choice behavior completely depends on how we are assuming/setting the decision maker's consideration set. For this reason, any choice sequence that can be explained by some consideration set also can be rejected by another consideration set. However, even with this concern, CBL framework itself provides a way to escape from those concern. Since CBL framework requires to explicitly exhibit the consideration set with its underlying preference, if there are two conflicting underlying preference that explains given choice sequence, we can focus only on whether each consideration set is appropriate as the decision maker's range of learning. This appropriateness may be confirmed via empirical test.

Second, I only focused on the three observable candidates of consideration sets. I was able to capture the sequential consistencies observed from the choice sequences generated by those consideration sets. Definitely they are not the exhaustive set of (general) axioms that will generate the class of choice sequences generated by general learning rule. Finding this set of axioms will make us better in understanding the whole aspect of sequential consistency.

3 Chapter 3: Mixing Propensity and Strategic Decision Making (with Duk-Gyoo Kim)

3.1 Introduction

A growing number of studies in economics and political science consider bounded rationality both in decision making and in strategic behavior. In decision making, rationality of individuals could be limited by a cognitive limitations of their minds. Individuals' strategic behaviors are also away from theoretical predictions with full rationality assumption, not only because their rationality is limited but also because their *belief about other individuals' bounded rationality* varies.

Our primary goal is to examine how individuals' non-strategic decision making (against probabilstic events) pattern is related to their strategic decision making (against actual anonyous opponent) pattern. In repetitive decision making under uncertainty, experimental observations suggest that significant amount (more than 40%) of subjects do not make decisions to maximize their expected payoff, but match their decisions to the probability of events. (Rubinstein (2002), Neimark and Shuford (1959)) We call such individual tendency in repetitive decision making as a **mixing propensity**. We claim that without considering individuals' heterogeneous mixing propensity, it is challenging to map individuals' strategic behaviors to their underlying belief.

We build upon two leading theories formalizing bounded rationality in strategic thinking: the Level-k (Lk) model (Costa-Gomes, Crawford and Broseta (2001), Costa-Gomes and Crawford (2006)) and the Cognitive Hierarchy (CH) model (Camerer, Ho, and Chung (2002, 2004)). Both models assume that individuals use only finite (=k) steps of iterative dominance, and such k varies by individual. One distinctive difference is that the Lk model assumes that individuals believe others' cognition level is homogeneous, while the CH model assumes that they believe it is mixed. To analyze experimental observations, they implicitly share an assumption that *every subject does not have the mixing propensity*, which may create sizable misinterpretation: An individual who has a certain type of mixing propensity may show homogeneous choice patterns even when she has a heterogeneous belief, while an individual who has another mixing propensity may make heterogeneous choice patterns that fully reflect her heterogeneous belief when the best response to the belief is a probabilistic mixture of many choices.

To address our question, we conducted two separate laboratory experiments: Odd-Ratio Decision Making (ODM) experiment and a modified beauty contest game, within the same subjects. In a nutshell, from ODM we can identify individuals' mixing propensities and categorize them into three types. With such mixing propensity types, observations from the beauty contest game can help us describe the belief distribution better.

ODM experiment can be understood as a repetition of matching penny games with unknown events. See Table 3.1 for illustration. Subject's options are on the first column; U, M and D in this example. The first row shows events and probabilities; (L=3/4, R=1/4) means the event L will be realized with probability 3/4, or R otherwise. The matrix shows the subject's payoff. For example, if she chooses M and an event L is randomly drawn, she earns $(3-v)/4$ points, where v is for adjusting certainty equivalent. v will be separately measured.

The ODM experiment consist of four separate games and each game consists of four sets respectively. Each set also consists of four rounds. A new event is drawn from the probability distribution at the beginning of each set. Subjects know that the event is realized, and the event will not be changed within a set, but do not know which event is realized. That is, subjects face the same

Game 1	L = 3/4	R = 1/4
U	1	0
M	(3-v)/4	(1-v)/4
D	0	1

Table 3.1: An example of ODM experiment

unknown event for four rounds, and after a new event is drawn, they face another unknown event for another four rounds, and so on. Based on subjects' choice patterns from four different games, we can categorize their mixing propensities into four types. Rational Optimizer (RO) will play an optimal action that maximizes the expected payoff for all rounds at all sets. Probability Matcher (PM) will mix his/her action to match the given probability within each set and this mixing proportion will be equal across the sets. Hedging Matcher (HM) will play intermediate action for all rounds at all sets. We will use the Maximum Likelihood estimation for categorization.

A modified beauty contest game will be conducted with the same subjects who participate in the ODM experiment. In the beauty contest game in Costa-Gomes and Crawford (2006), subjects earn more when they guess the match's action more accurately. This idea continues in our game. Both player 1 (P1) and player 2 (P2) know the choice intervals and target parameters of P1 and P2. P1's goal is to submit a number within P1's choice interval that is closest to P2's number times P1's target parameter. Three distinctive differences are as follows: (1) Subjects play eight rounds of beauty contest games in five sets. At each set, they play with a new match. This setup allows us to fully utilize the individuals' mixing propensity type. (2) The payoff function is deliberately designed in a way to distinguish a player's deterministic choice from a naive random choice within an interval. (3) We introduce a calculation panel which tracks subjects' exact thought process.

Estimation and inference about the underlying belief structure would be similar to those in Camere, Ho, and Chung (2004), but we believe the PM type subjects can represent the entire underlying belief structure more accurately. That is, if PM type subjects are likely to show the similar probability matching behavior in a SDM experiment, their individual distribution of choices in a SDM experi-

ment may represent their underlying belief about their opponent more accurately than other types. Interestingly, in the actual experiment, PM type subjects are more likely to diversify their choices to different levels of cognition compare to other types. Individual PM type subjects showed 1.06 for average variances of cognition level in individual choice distribution, while RO and HM types showed 0.61 and 0.65 respectively.

This study will proceed aby following order. In the section 2, we will describe details of experimental design. In the section 3, we will show results of experiment and discuss of its implication. Section 4 will conclude the study.

3.2 Related Literature

This study is developed on the empirical and theoretical findings those assume that individual subjects play strategies of different level of iterative dominance. Among many previous behavioral models, we mainly consider two models: Level-k model from Costa-Gomes and Crawford (2006) and Cognitive Hierarchy model from Camerer, Ho, and Chung (2002). Both models share the same assumptions that (1) individuals optimal chooses the best response to their underlying belief about their opponent's action and (2) every individual expects their opponent's play (at most) certain level of iterative dominant strategy. On the other hand, both model differs in the assumption that subjects adopt the uniform belief or heterogenous belief structure. Level-k model assumes that individuals have the uniform belief such that all their opponents play the same level of iterative dominant strategy. For example, $L2$ subject assumes that all their opponents play a one-time iterative dominant (or $L1$) strategy. From that assumption, Level-k subjects are supposed to play a certain strategy that best response to their uniform belief. In the Costa-Gomes and Crawford (2006), about 55% of subjects are identified as showing a certain level of play that can be interpreted as adopting the Level-k model. On the other hand, some subjects explicitly mixed two or more different strategies that represents different level of iterative dominance. Such systematic pattern is not coincides to the uniform belief assumption. In Costa-Gomes and Crawford (2006), authors found source of

such deviation from the learning. That is, even individuals started from the initial uniform belief, the experience leads subjects to shift to the higher level of iterative dominance while keeping the uniform belief structure. However, for some subjects, such mixing occured irrelevant to the time horizon. These observations suggested us to consider of alternative model that can explain such behavioral pattern. Cognitive Hierarchy model allows individuals to have heterogeneous belief structure. For example, $L2$ subject assumes their opponent plays not only $L1$ strategy but also $L0$, which is uniform random, strategy. Depends on the individual belief about the proportion of users of two different strategies, each subject may find a different best response. In Camerer, Ho, and Chung (2002), authors explicitly estimated the structure of belief by using observations from previous studies and their original experimental observation. However, even CH model allows the heterogeneous belief structure, they cannot fully explain the observations with mixed choices. That is, as the best response to the heterogenous belief, choosing the interim choices consistently can be strictly better than mixing the different choices those correspond to different strategies respectively.

This study attempt to explain such puzzle by relaxing the assumption that all individuals have the same (rational) response to the same belief. That is, we consider that individuals may show different response to the same belief and this difference in "individual optimization" may lead to the apparent puzzle that mixes different strategies. Especially, in Rubinstein (2002) subjects frequently showed matching their responses to the probability distribution of possible events. The author asked undergraduate students to solve five modified but similar sequential choice problems which have a unique stochastic dominant (or "Rational") solution each. Here is 'Catch the messenger' problem in Rubinstein (2002).

('Catch the messenger'). Imagine you are a detective at a shop- ping center. You know that every day at noon, a messenger arrives with an envelope. The identity of the messenger is unknown; he is one of dozens of messengers who work for a delivery company. The shopping center has four gates and you have only one video camera, which you have to install each morning in one of the four gates. Your aim is to take photos of the maximum number of messengers as they enter the shopping center. You have to choose a plan determining where to install the camera every morning. You have in hand the results of a reliable statistics on the entry of messengers according to gate: 36% use the Green gate, 25% the Blue gate, 22% the Yellow gate and 17% the Brown gate.

Day : Sun Mon Tue Wed Thu

Plan : ____ ____ ____ ____ ____

However, only a small portion of students always played stochastic dominant action (= "Green gate"). Likewise, many psychology literature found significant propensity for mixing different strategies. In psychology, such behavioral pattern is called as "*probability matching*" behavior. Neimark and Shuford (1959) and Vulkan (2002) also provide lab-experiment observations from psychology that support the existence of probability matching behavior. When we consider that similar probability matching behavior can occur at the strategic decision making process, the mixing strategies of different levels could be better reflection for the underlying belief structure.

However, only a little experiment studies explictly considered such behavioral pattern in the optimization process in the identification of underlying belief structure in the strategic decision making environment. Healy et al. (2015) considered whether individuals show the similar level of iterative dominance in the different form of the game. Healy et al. (2015) conducted several different (strategic) games to the same individuals. Specifically, subjects practiced four different non-strategic tests and played a strategic decision making session. In the strategic decision making session, subjects played the 'undercutting game' and 'beauty-contest game' for four and five times respectively. While the undercutting game only allowed the discrete choices, the beauty contest al-

lowed some interim choices that does not represent any level of iterative dominance. In the result, even two games shared the similar structure that requires players to exploit iterative dominant strategy, individuals show almost no correlation between the level of iterative dominance. Moreover, there was no significant connection was found between the individual trait, like IQ, and the level of iterative dominance. Healy et al. (2015) attempted to find a consistency of the strategic process in different environment, but did not considered it in terms of individual optimization pattern.

3.3 Experimental Design

3.3.1 Odd-Ratio Decision Making Experiment

We design an odd-ratio decision making experiment (henceforth, ODM experiment) to identify an individual's specific mixing propensity. The entire ODM experiment consists of four different Matching Pennies games, and subjects play each game repetitively. Each game consists of four sets and each set consists of four rounds. That is, each Matching Pennies game is repeated for 16 rounds. Subjects are told that a new state is randomly drawn from the known probability distribution per each set. Subjects face the same state for four rounds within a set and as the set changes they will face another state for another four rounds. Since there are four different games, each subject plays 64 rounds (4 games $\times$ 4 sets $\times$ 4 rounds) during the entire experiment.

To prevent subjects' learning about the state from previous outcomes, the outcome of the game will not be notified to the subjects during the experiment. They will be informed of the realized outcome at the end of the experiment and will get paid privately according to the outcome. Moreover, the game with the states (of computer player) allows us to prevent them from concerning the others' payoff (e.g., the inequality aversion of Fehr and Schmidt (1999)).

Table 3.2 describes four Matching Pennies games respectively. In every round, subjects choose one among the first column of rows. The state is drawn from the first row of columns with the probability associated to each state. For example, subjects can choose one among U, M and D in Game 1 and a state is either L with probability 3/4 or R with probability 1/4. Subject's payoff is described in the payoff matrices.

We varied structures of each Matching Pennies game by (1) existence of dominant actions, (2) the number of selectable actions and (3) the highest expected payoffs subjects can earn. See Table 3.3.1.

v_i is the discount of payoffs for the hedging action (M in Game 1, 2 and B in Game 3, 4) for subject i. This discount is adopted to prevent the subject's bias toward the hedging action due to risk

Game 1	L = $\frac{3}{4}$	R = $\frac{1}{4}$
U	1	0
M	$\frac{3-v_i}{4}$	$\frac{1-v_i}{4}$
D	0	1

Game 3	L = $\frac{1}{2}$	R = $\frac{1}{2}$
U	1	0
M	$\frac{1-v_i}{2}$	$\frac{1-v_i}{2}$
D	0	1

Game 2	L = $\frac{1}{2}$	C = $\frac{1}{4}$	R = $\frac{1}{4}$
U	1	0	0
M	0	1	0
D	0	0	1
B	$\frac{1-v_i}{2}$	$\frac{1-v_i}{4}$	$\frac{1-v_i}{4}$

Game 4	L = $\frac{1}{4}$	LC = $\frac{1}{4}$	RC = $\frac{1}{4}$	R = $\frac{1}{4}$
U	1	0	0	0
MU	0	1	0	0
MD	0	0	1	0
D	0	0	0	1
B	$\frac{1-v_i}{4}$	$\frac{1-v_i}{4}$	$\frac{1-v_i}{4}$	$\frac{1-v_i}{4}$

Table 3.2: Matching Pennies games in the ODM

Each subject may play all four games with random order; Discount for a hedging behavior v_i (Game 1 and 2: M, Game 3 and 4: B) varies by individuals.

	Existence of dom. actions?	The number of states	The highest expected payoff
Game 1	Y	2	$\frac{3}{4}$
Game 2	N	2	$\frac{1}{2}$
Game 3	Y	3	$\frac{1}{2}$
Game 4	N	4	$\frac{1}{4}$

Table 3.3: Comparison of Four Matching Pennies Games

aversion. Since the hedging action gives exactly the same expected payoff from each single action choice and always guarantees a positive amount of payoff, risk-averse subjects may consider the hedging action as the dominant choice. To exclude this concern, we measured their risk-averseness before the beginning of the ODM experiment, and discounted their payoff of the hedging action accordingly.[1]

From this formation, we categorize four possible types of individual mixing propensity: Rational Optimizer (RO), Probability Matcher (PM), Uniform Matcher (UM), and Hedging Matcher (HM). Those four types are distinguished by their behavioral pattern.

Observation 1. (Mixing Propensity) Individuals with a different mixing propensity are to show different decision-making patterns;

[1] detailed footnote goes here to address (1) how to measure the risk-averseness, (2) how to make subjects not to think about the relationship between this pretest and actual experiment.

Game 1	Set 1	Set 2	Set 3	Set 4
RO	U4	U4	U4	U4
PM	U3D1	U3D1	U3D1	U3D1
UM	U4	D4	U4	U4
HM	M4	M4	M4	M4

Table 3.4: Model Behavior of Four Types in Game 1

1. A RO always plays the action that maximizes expected payoff.
2. A PM mixes different actions within each set and the proportion of mixing follows the probability distribution of the states.
3. A UM plays a single action within each set but changes actions across the sets. The proportion of such mixing follows the probability distribution of the states.
4. A HM always plays an hedging action that provides a positive payoff at any cases.

Each type has a different play pattern in the set and the game-wise level. The following table 3.3.1 shows possible choice patterns of each type in Game 1.

The RO type subject is expected to play an action U all the time since U maximizes the expected payoff. The PM type subject is expected to play action U three or four times at each set because the PM type is expected to mix his/her play to match with the given probability within each set. The HM type subject is expected to play the intermediate action M all the time.

3.3.2 Strategic Decision Making Experiment

We design a strategy decision making (SDM) experiment to identify individual strategic decision making process with the consideration of the individual mixing propensity. Entire experiment consists of eight sets and each set consists of five rounds of the beauty contest game. In each set, two anonymous subjects are randomly matched and play a whole set with the same partner. The beauty contest game in each set will be also the same for all rounds of the set. As the set

changes, each subject will be randomly rematched to another anonymous partner. The game they play will be also changed to another game randomly. Eight games have different structure in terms of players' choice intervals and target numbers. Each game will be played only at the one set. The game subjects played in the previous sets will not be played in the following sets.

Similar to the ODM experiment, the realized outcome of their choice action will not be informed during the experiment. That is, subjects play the game without feedback and the final outcome of their choice actions will be informed only at the end of entire SDM experiment. We adopt this restriction to prevent subjects from a retrospective or an experience-based learning. Subjects will earn payoffs at each round according to a payoff function and monetary compensation will be paid according to sum of payoffs at the end of experiment. All such structure of the SDM experiment will be informed to the subjects by written form and spoken by experimenters.

We designed eight games to be paired into four pairs. And each subject are assigned to play the two different positions of the same game within each set. For example, the game $\alpha n_2 \beta n_4$ and $\beta n_4 \alpha n_2$ is paired game. So, the player 1 of the game $\alpha n_2 \beta n_4$ plays exactly the same role of the player 2 of the game $\beta n_4 \alpha n_2$. Similarly, the player 1's role of the game $\beta n_4 \alpha n_2$ is exactly same to the player 2's role of the game $\alpha n_2 \beta n_4$. We designed each set to be consisted of four pairs which converese the role. So each subjects will play both sides' role at the same set. This feature will be also informed to the subjects at the instruction stage of the SDM experiment.

To suppress subjects' experience-based learning and routinzed choice pattern, we adopted two randomization devices. At each experiment, subjects play 8 different beauty contest games and the basic structure of games will be the same. Repeating the same game 5 times at each set allows subjects revealing their belief strructure under the assumption that their mixing propensity in the ODM experiment will be also applied to the SDM experiment. However, repeating the same game may involve the experience-based learning that subjects show change of iterated dominance level. Since we focus on capturing subjects' initial belief and strategic decision, we need to aviod such experience-based learning as much as possible.

Form	Target Structure	# of Iteration	Pattern of Iteration	End with Dominance
$\alpha n_2 \beta n_4$	Mix / High	17	A	Y
$\beta n_4 \alpha n_2$	Mix / High	18	A	N
$\delta n_3 \beta n_1$	Mix / Low	4	A	Y
$\beta n_1 \delta n_3$	Mix / Low	5	A	N
$\beta n_1 \beta n_2$	Low	4	S	Y
$\beta n_2 \beta n_1$	Low	4	S	Y
$\delta n_3 \gamma n_3$	High	2	A	N
$\gamma n_3 \delta n_3$	High	2	A	Y

Table 3.5: A Structure of the Beauty Contest Games

The first device is imposing the variation of the game structure. We adopted eight different combinations of choice intervals and target numbers. α denotes a choice interval [100, 500], β denotes [100, 900], δ denotes [300, 900], and γ denotes [300, 500] respectively. A target number is defined as $n_1 = 0.5$, $n_2 = 0.7$, $n_3 = 1.1$, and $n_4 = 1.5$ respectively. At each set, subjects face different pairs of target numbers. Moreover, we designed the sames to be differed in the target structure, the number of iteration to arrive to Nash equilibrium choice, the pattern of iterated strategies, and the location of Nash equilibrium choice. So choosing always the biggest and/or the smallest number may not maximize their payoffs. This fact will be also notified to the subjects at the instruction stage. Combined with no-feedback policy, such variation will prevent subjects to change their strategic/behavioral pattern after they observe the outcome of the past choice. As a result, we can expect subjects to concentrate on their own strategy and belief to maximize their payoff at each game. [Table 3.5] summarizes more details about structure of games[2].

"Target Structure" describes combination of the pair of target numbers. "High" denotes the case in which both p^1 and p^2 are bigger than 1. Similarly, when both p^1 and p^2 are smaller than 1, we denote it as "Low." "Mix" denotes the case in which $(1 - p^1)(1 - p^2) < 0$. So, either $p^1 > 1 > p^2$

[2]"End with dominance" = whether Nash equilibrium action is located at the end of choice interval. Since Nash equilibrium corresponds to infinite times of dominance iteration, "end with dominance" means the interval's end corresponds to such iterated dominance.

or $p^2 > 1 > p^1$. "Mix/High" denotes "Mix" case where $p^1 \cdot p^2 > 1$. "Mix/Low" denotes another "Mix" case where $p^1 \cdot p^2 < 1$. "# of Iterration" describes how many steps of the iterated dominance is required to arrive to Nash equilibrium. Every game has a different number of iteration to find an exact Nash equilibrium choice. "Pattern of Iteration" describes whether the number correspond to each step of the iterated dominance strictly (or monotonely) increases/decreases as the level of iterated dominance increases. In case of strict increasing or decreasing, it is denoted as "S." Otherwise, they will change alternatively (increase and then decrease or vice versa), and it is denoted as "A." For example, a game $\alpha n_2 \beta n_4$ corresponds the case "A." Start with L1 strategy (1st level of iterated dominance) 419.4, L2 strategy decreases to 361.1. And then, L3 strategy increases to 440.3. On the other hand, the game $\beta n_1 \beta n_2$ corresponds to the case "A" that L1, L2 and L3 strategies decrease strictly. "End with Dominance" describes the location of the Nash equilibrium number on the choice interval. If Nash equilibrium is located at the end of each interval (i.e, a^1 or b^1), we denoted is as "Y." Otherwise, the Nash equilibrium is located at the interior of the choice interval and is denoted as "N."

Given this structure, [Table 6] shows numbers correspond to Nash equilibrium, each level of the iterated dominance, and the remaining intervals correspond to each round of the iterated deletion respectively. $L1$ implies the first level of the iterated dominance when the subject considers his/her opponent to choose numbers with equally same probability over the all interval ($L0$ player of CGC (2006), CHC(2004)). $L2$ and $L3$ corresponds to the second and third level of iterated dominance at which the subject considers their opponent to play $L1$ and $L2$ strategy respectively. NE implies the choice corresponds to Nash equilibrium of the game.

The second device we adopted in this study to prevent subjects' experience-based learning is a permutation in the order of play. Even we changed the game structure over the game, repeating the same game may allow subjects to learn about the optimal strategy that will be played at the paired game. For example, subject who played game $\alpha n_2 \beta n_4$ may learn to play another strategy at $\beta n_4 \alpha n_2$ (the paired game of $\alpha n_2 \beta n_4$) with reflection of own play at the game $\alpha n_2 \beta n_4$. Our

Form	L1	L2	L3	NE	1st Round	2nd Round	3rd Round	4th Round
$\alpha n_2 \beta n_4$	419	360	440	500	100, 450	105, 500	105, 472.5	110.25, 500
$\beta n_4 \alpha n_2$	515	629	540	750	150, 750	150, 675	157.5, 750	157.5, 708.75
$\delta n_3 \beta n_1$	678	363	373	300	300, 900	300, 495	300, 495	300, 300
$\beta n_1 \delta n_3$	330	339	181	150	150, 450	150, 450	150, 247.5	150, 247.5
$\beta n_1 \beta n_2$	303	209	106	100	100, 450	100, 315	100, 157.5	100, 110.25
$\beta n_2 \beta n_1$	419	212	146	100	100, 630	100, 315	100, 220.5	100, 110.25
$\delta n_3 \gamma n_3$	463	550	550	550	330, 550	363, 550	399.3, 550	439.3, 550
$\gamma n_3 \delta n_3$	500	500	500	500	330, 500	363, 500	393.5, 500	439.3, 500

Table 3.6: A List of Strategic Choices with respect to the Iterated Dominance

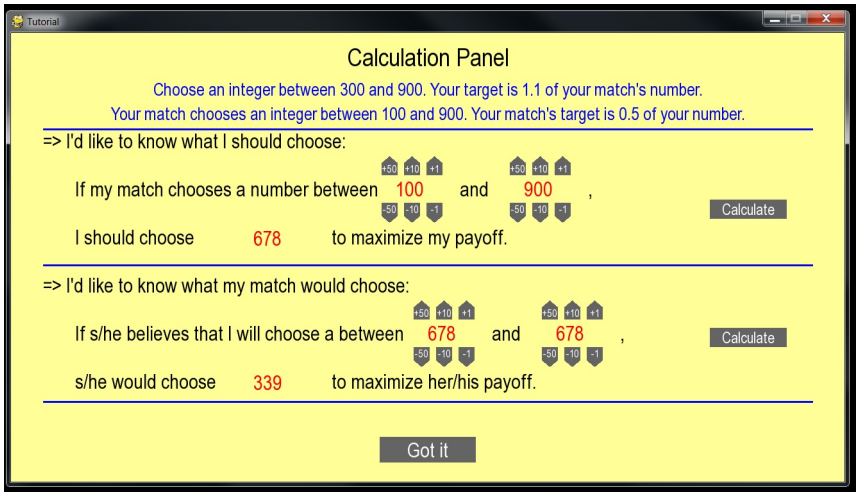


Figure 3.1: Calculation Panel Screen Example

concern is that playing the paired games in a row may reinforce effect of such experence-based learning. To aviod such change in choice from experinece-based learning, we will intentionally permutate the play order of the games two paired games are not to be played in series. And the assignment of the order will be random by individual. This permutation and randomization will be helpful to prevent experience-based learning that may occur when subjects face paried games. [Randomization for the interval weight is not updated: will be added]

3.3.3 Calculation Panel

Subjects can use the calculation panel to find the exact number that corresponds to their best response for partner's choice. There are two modules in at each game's calculation panel. The

above module ("A module") gives result for the best response for the player him/herself with respect to the prediction of the opponent's choice. The below module ("B module") provides the best response for the opponent with respect to the prediction of the opponent's prediction about the player's choice. For each module, there are two range numbers. In the range of A module, subjects can put in the minimum and maximum range within which their opponent's choice might be placed. Then clicking the gray boxed "Calculate" button may generate the red-colored number that maximizes the own payoffs. The distribution over the range is fixed as the uniform distribution over the settled range and subjects will be informed of it. The [Figure 3.1] considers the player 1 at the game $\delta n_3 \beta n_1$. If player 1 want to find L1 strategy with belief of partner's range from 100 to 900, he/she may enter 100 for MIN and 900 for Max and activate [Calculate] button. Then, A module will generate the result 678 as the closest intezer approximate of the best response 678.3.

This calculation panel also allows subjects to find the number for the higher order of iterated dominance (Lk with k > 1). To have the L3 result, the player 1 need to know his/her partner's L2 choice in advance. Then, the player 1 enter MIN 100 and MAX 900 which corresponds to his/her own range of choice to have own L1 best response. Then, put the L1 best response to the range of B module to calculate the L2 best response of the opponent. For example, 678 in the game $\delta n_3 \beta n_1$ corresponds to his/her opponent's range for the L2 choice. From that result, the player 1 can put in 678 for both MIN and MAX (since s/he alrealy have point prediction) and click [Calculate] to have a result of L2 response of the opponent, which is 339 in the [figure 1]. Then, putting 339 for A module's range will generate player 1's L3 best response.

Moreover, a usage record of the calculation panel allows us to track the individual subjects' decision process. That is, having tractable records for calculation process at each step of iterated dominance may help us to figure out the paths subjects might followed to have the final decision. Comparing the lastly confirmed result from the calculation panel and actual decision making also will help us to understand how the subjects use their calculation process when they making the strategic decision. At the calculation panel, subjects will be informed of the recording of their usage of panel and the result of usage will not be related to their monetary outcome at the end

Date	150420	150421	150914	150915-1	150915-2	150915-3	SUM
Participated	19	13	17	11	13	13	86
Effective	11	12	12	8	11	8	62

Table 3.7: Details of Laboratory Sessions

Type	Count	%
RO	28	45.2
PM	16	25.8
HM	18	29.0
Summation	62	100

Table 3.8: Overall Distribution of Mixing Propensity in the ODM experiment

of session. They will be confirmed that their payoff from action choice is the only factor that determines their own monetary compensation.

3.4 Results

Six sessions of laboratory experiment were conducted at Missouri Social Science Experimental Lab (MISSEL) of Washington University in St. Louis with 86 participants ([Table 3.7]). From all paricipants, we collected 62 effective subjects from those who passed screening tests of both experiments. We estimated their individual type by using experimental data from the ODM experiment. Using MLE method[3], we found that RO, PM, and HM type subjects share 45.2%, 25.8% and 29% respectively from all effctive samples. ([Table 3.8])

Using this basic group categorization, we will consider their behavioral pattern in the SDM experiment. We propose a main hypothesis that subjects' behavioral patterns ODM experiment are inherited to the SDM experiment. To identify it, we separatedly tested two sub-hypotheses:

(1) RO types are more likely to show a higher level of cognition level with less dispersion than PM types

[3]For detailed statitical process, please see the appendix.

	$E[\mu_i]$	$E[(\sigma_i)^2]$
RO	3.17	0.61
PM	2.71	1.03
HM	2.32	0.65

Table 3.9: Mean of Distributions of Individual Means and Varinaces in Cognition Level

(2) HM types have different distribution of variances of cognition level from PM types, but have a similar distribution to RO types.

3.4.1 (Result 1) RO types are more likely to show a higher level of cognition level with less dispersion than PM types.

First, we compared two groups' distributions in the SDM choices' cognition level. From individuals' choices, we could find variance of their own distribution of cognition level. Collecting such individual variances, we can find the distribution of variances of cognition level for each group.

[Table 3.9] shows a summary result of two distributions. RO type subjects showed 3.17 for an average cognition level of choices and 0.61 for mean of variances respectively. PM and HM type subjects showed 2.71 and 2.32 for an average cognition level of choices, and 1.03 and 0.65 for mean of variances respectively. Interestingly, RO and HM type subjects showed relatively lower variance than PM type subjects. On the other hand, RO type showed relatively higher average cognition level than other type subjects. This result briefly implies that three types of subjects are effective to describe their behavioral pattern.

To consider such difference more detailed way, we conducted test of distributions between each pair.

[Table 3.10] shows a summary result of a test between different two groups. We used the Fisher method (F-test) to measure a similarity between two groups' distributions. We tested null hypothesis that two distributions are from the same population by using mean and variance respectively.

	Mean Test			Variance Test		
Types	RO&PM	RO&HM	PM&HM	RO&PM	RO&HM	PM&HM
P-Values	0.012	0.004	0.11	0.044	0.439	0.024

Table 3.10: Test Result (p-values) Between Distributions

Test of means between RO and PM type group resulted p-value 0.012 (one-sided) and 0.024 (two-sided) respectively. These p-values provide enough evidence to reject the null hypothesis at high enough significance level. Test of variances showed p-values 0.044 (one-sided) and 0.089 (two-sided). Even though such a result is less than result from mean test, it is enough to satisfy 5% level of statitical significance. Combining these two test results, we can reject the null hypothesis that RO and PM type subjects have the same mean and/or variance for level of cognition.

[Figure 3.2] shows a proportion of total choices from each type's group. In the figure, RO types shows distinctive single-peaked shape at 4 or more level while PM types shows multi-peaked distribution for all levels. The result of hypothesis (1) supports our presumption that PM and RO types are inherit their behavioral pattern showed in the ODM experiment to the SDM experiment. RO types, who consistently choosed actions that maximize expected payoff, also showed consistent choice behavior at certain level of cognition. On the other hand, PM types, who matched their actions to a given probability disctributions of opponent's action, also distributed their actions to several different cognition level.

3.4.2 (Result 2) HM types are less likely to diversify their behavior than PM types, but an average level of cognition is lower than RO types.

The next hypothesis examines whether HM types are distnguished by other types by choosing a certain intermediate level of cognition. In the ODM experiment, HM types choosed actions that gives positive payoffs in any events. We interpreted their behavior as an subjective optimization which is supposed to minimize a risk of wrong (or missed) prediction. So we presumpted that HM

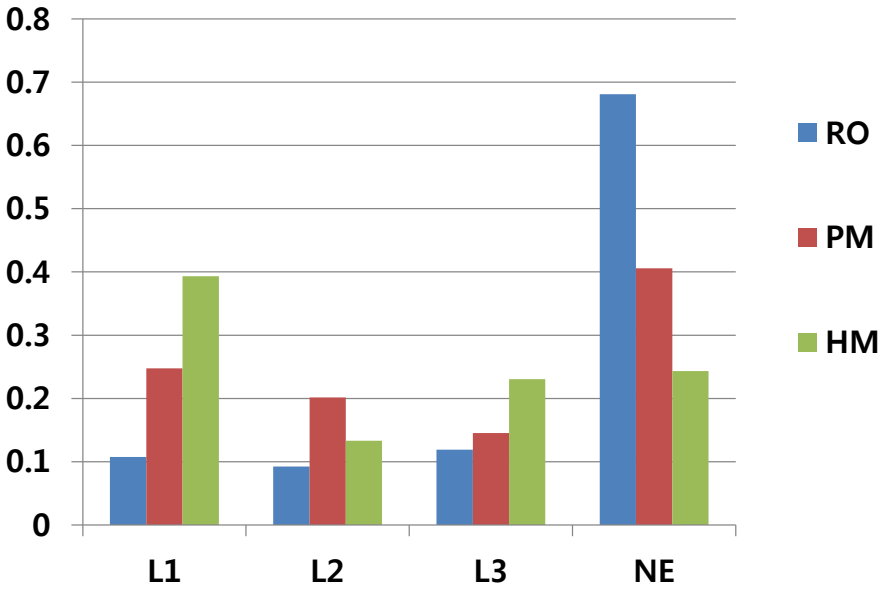


Figure 3.2: Proportion of Total Choices from Each Type

types may choose some intermediate level(s) of cognition within their own support of beliefs. For example, suppose that some subject has a belief about her opponent's cognition level distribution such as L0: L1: L2 = 20: 30: 50 (%). In this case, if L1 choice can give her a positive payoff in any events, she may choose L1 choice consistently. Likewise, we presumpted that their behavioral pattern may show a single-peaked shape as in the RO types, but the average level of cognition will be lower than RO types because they are likely to choose some intermediate level(s) of cognition consistently.

Hypothesis (2) tests whether HM types have a different shape of choice distribution compare to the other types. We first considered overall shape of individual distributions. Since RO types showed obvious single-peaked shape distribution, comparing distributions of variances of individual choices from HM and RO types may provide us a statistical evidence. In the [Table 3.9], RO and HM types showed 0.61 and 0.65 for mean of variances respectively. Via the F-test method, we compared two groups' distributions of individual variances. We set the null hypothesis that two sample distributions came from the population with the same variance. Test resulted p-values

0.439 (one-sided) and 0.879 (two-sided) respectively ([Table 9]). The test results cannot reject the null hypothesis. We also tested a hypothesis that two samples come from the population with the same mean. Test resulted p-values 0.004 (one-sided) and 0.008 (two-sided) respectively ([Table 9]). Test result shows strongly denies the hypothesis that RO and HM type group has the same mean for distribution of individual choices. Combining these results, we would say that HM type subjects are likely to have a single-peaked shape for their individual choice distribution, which is similar to RO type subjects, but their overall choice levels are lower than RO type subjects.

Similarly we tested the difference between HM types and PM types. In the [Table 3.9], we have results from tests for their means and variances from individual distributions. For variances, we have p-values 0.024 (one-sided) and 0.048 (two-sided) respectively. This result also supports our hypothesis that HM types have the single-peacked individual distribution, which is distinguished from behavioral pattern of PM types, who diversify their choices to several different levels of cognition.

We interpret this result as a support for our hypothesis: HM types are likely to choose an intermediate level of cognition consistently. HM type subjects are likely to choose a certain level of cognition level consistently, which is similar to RO types and strongly distinguished from PM types. However, their overall level of cognition is lower than RO types. Such a behavioral pattern not only distnguishes them from other types but also matched to our prediction for HM types.

3.4.3 Recovery of Belief Structure

We now consider belief structure of PM types; From the above result, we observed that three types are likely to show similar behavioral pattern in both (ODM and SDM) experiments. For RO type subjects, this result implies that distribution of RO type subjects, which is mostly focused on a Nash equilibrium action, cannot fully reveal their underlying belief structure. It says most of RO type subjects put the highest weight of their belief on the Nash Equilibrium action, but silent about the other belief they might had. However, PM subjects are more like to diversify their response

Level	L1	L2	L3/3+	NE	Undef.	SUM
Count	97	79	57	159	248	640
W/ Undef. (%)	0.152	0.123	0.089	0.248	0.388	100
W/O Undef. (%)	0.247	0.202	0.145	0.406	-	100

Table 3.11: Overall Distribution of Cognition Level for PM Types in the SDM

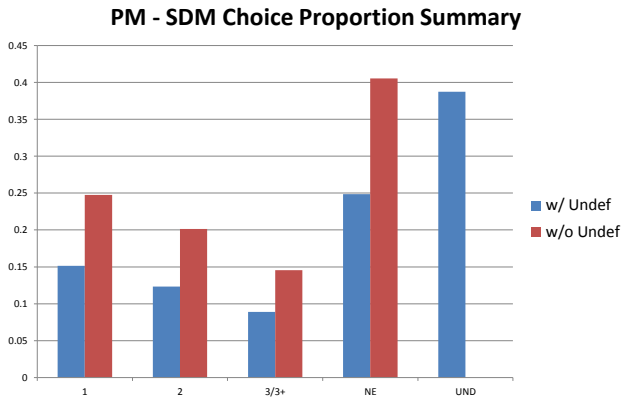


Figure 3.3: Proportion of Total Choices from PM Type

in the SDM experiment, which is expected to be fitted to their underlying belief structure. So we consider 16 PM type subjects' actual responses to recover their underlying belief structure.

In summary result (See [Table 3.11] and [Figure 3.3]), PM type subjects distributed their responses over almost every level of cognition. Even they distributed mostly their responses on the NE responses, they put similar weight on the different level of cognition. Especially, subjects put more weights on the L1 and L2 actions than NE actions. This result implies that most of PM type subjects consider the existence of L0 (or random behaving) subjects. Moreover, as they consider of L0 subjects, they also aware of similar proportion of L1 subjects who may best response to L0 subjects.

Another interesting observation is that as the level of cognition gets higher, the proportion of such a level gets smaller. From [Table 3.11], L1 to L3/+ level share is 24.7 %, 20.2 %, and 14.5 % respectively. When we consider undefined response as a random-like behavior, which correspond

to L0 behavior, this trend is still consistent. This proportion jumps to 40.6% at the NE action. This trend allows us to have two inferences: First, (PM type) subjects consider that less opponents play a higher level of cognition. For this reason, they may strategically assigned a less proportion to the higher level of cognition. Second, subjects may have some "framing" effect on the Nash Equilibrium. That is, even though most of NE actions require higher level of cognition than other actions, they may consider it as some "focal point" and put the highest weight. This interpretation is also consistent to our observation for RO type subjects. Assuming that RO type subjects also have the similar belief structure to PM types, they are mostly best reponse on NE action based upon their "rational optimization" pattern.

3.5 Conclusion

In this study, we examined how individual optimization pattern is related to their strategic decision making via laboratory experiment. We considered each individual has different ways of optimization when they face a probabilistic event and categorized them into three different types: Rational Optimizer (RO), Probability Mathcer (PM), and Hedging Matcher (HM). Interestingly, we found more than a half of entire subjects showed different patterns from rational optimization. Moreover, they showed different patterns of strategic decision making according to their optimization pattern. While RO type subjects focused on a NE equilibrium action, PM and HM type subjects choosed their actions to a lower levels of cognition. Especially PM type subjects diversified their actions to multiple different levels of cognition as they diversified their actions in the ODM experiment.

Assuming that PM type subjects showed similar optimization pattern in the SDM and ODM experiment, we can have more detailed recovery of underlying belief structure for strategic decision making. In the result, we observed that (PM type) subjects played different actions and assigned less proportion of actions to higher levels of cognition (L1: 15.2%, L2: 12.3%, L3/+: 8.9%). This result suggests that subjects strategically assigned their actions based upon their underlying belief. Moreover, we observed PM type subjects also assigned higher proportion (24.8%) on the

Nash equilibrium action. This result supports previous result from Camerer et al. (2002, 2004) that subjects consider Nash equilibrium as one of focal point. Summing up these observations, we can conclude that subjects actually have heterogeneous belief structure, which consists of several different levels of cognition, but still consider the Nash equilibrium as a plausible focal point.

3.6 Appendix: Statistical Model Specification

3.6.1 ODM Model Specification

In the ODM experiment, subjects make one choice at each round, 64 in total. To analyze the individual subjects' decision-making patterns, we use the maximum likelihood method.

Let $x_{g,j}^{i,k}$ denote the number of subject i's decision that equals to the type k subject's decision in game g of set j. We define $k \in K = \{\text{RO, PM, HM}\}$ and $g,\ j = 1, 2, 3, 4$. We similarly define a vector $x_g^{ik} = (x_{g,1}^{i,k}, \cdots, x_{g,4}^{i,k})$.

We define $\varepsilon^k \in [0, 1]$ the type specific rate of random choice that is independently and identically distributed ("i.i.d."). c_g is the number of actions that each subject has in game g. that is, $c_1 = c_2 = 3$, $c_3 = 4$, and $c_4 = 5$. Since ε^k is assumed to be type specific and identically distributed over all the choices, we formulate the probability that the subject of the type k makes some predicted decision at game g as $1 - \varepsilon^k + \varepsilon^k/c_g = 1 - (c_g - 1) \cdot \varepsilon^k/c_g$.[4] Then, $L_g^{i,k}(\varepsilon^k | x_g^{i,k})$ is the probability of observing $x_g^{i,k}$ when the subject i is of type k :

$$L_g^{i,k}(\varepsilon^k | x_g^{i,k}) = \prod_{j=1}^{4} \left[1 - (c_g - 1) \cdot \varepsilon^k / c_g\right]^{x_{g,j}^{i,k}} \times \left[\varepsilon^k / c_g\right]^{4 - x_{g,j}^{i,k}}.$$

Similarly, we define $\hat{x}_g^{i,k}$ the number of sets of subject i's decision that equals to the type k's decision in game g. That is, $\hat{x}_g^{i,k}$ counts the number of sets in each vector x_g^{ik} such that each set has exactly the same number of the type k subject's decision. For example, consider the RO type subject who may choose the same action in 3 sets (say, set 1, 2, and 3) and mixed two actions in another set (say, set 4). Then, $\hat{x}_g^{i,k} = 3$ since the number of sets that equals to the RO type subject's decision is 3 (set 1, 2, and 3). With the similar notion, we define a vector $\hat{x}^{i,k} = (\hat{x}_1^{i,k}, \cdots, \hat{x}_4^{i,k})$. Then, we define $\hat{L}^{i,k}(\varepsilon^k | \hat{x}^{i,k})$ the probability of observing $\hat{x}^{i,k}$ when the subject i is of type k :

[4] For illustration, suppose that a subject is a RO. With $\varepsilon^{RO} = 0$ or she do not make any mistakes, she will choose the action that maximizes the expected payoff with probability one. If $\varepsilon^{RO} = 1$ or she makes a choice in a completely random manner, then the probability of the optimal choice is $1/c_g$.

$$\hat{L}^{i,k}(\varepsilon^k|\hat{x}^{i,k}) = \prod_{g=1}^{4} \left[L_g^{i,k}(\varepsilon^k|x_g^{i,k})\right]^{\hat{x}_g^{i,k}} \times \left[1 - L_g^{i,k}(\varepsilon^k|x_g^{i,k})\right]^{4-\hat{x}_g^{i,k}}.$$

Now, we define $z^{i,k}$ as a type indicator for subject i where $z^{i,k} = 1$ if subject i is of type k and $\sum_{k\in K} z^{i,k} = 1$. From $\hat{L}^{i,k}(\varepsilon^k|\hat{x}^{i,k})$, subject i's maximum likelihood function can be calculated :

$$\begin{aligned} L^i(\varepsilon, z^i|x^i) &= \prod_{k\in K} \hat{L}^{i,k}(\varepsilon^k|\hat{x}^{i,k})^{z^{i,k}} \\ &= \prod_{k\in K} \left[\prod_{g=1}^{4} \left\{L_g^{i,k}(\varepsilon^k|x_g^{i,k})\right\}^{\hat{x}_g^{i,k}} \times \left\{1 - L_g^{i,k}(\varepsilon^k|x_g^{i,k})\right\}^{4-\hat{x}_g^{i,k}}\right]^{z^{i,k}}, \end{aligned}$$

where $\varepsilon = (\varepsilon^k)_{k\in K}$, $z^i = (z^{i,k})_{k\in K}$, and $x^i = (x_g^{i,k})_{k\in K}^{g=1,\cdots,4}$.

As a result, we can estimate the distribution of $z^i = (z^{i,k})_{k\in K}$ which may assist us to categorize the subject i's individual mixing propensity; we may categorize subject i to one of four types which has the highest $z^{i,k}$.

3.6.1.1 Type Categorization

Each mixing propensity type subject has different model play pattern for each game. Define $x_{g,j}^{i,k}$ is the number of subject i's decision that equals to the type k subject's decision in game g of set j. Similarly, $\hat{x}_g^{i,k}$ is the number of the sets of subject i's decision that equals to the type k subject's decision in game g. Consider an example subject i with play (UUUD, UUUM, UUDD, UUUU) in the game 1. Each entry of the vector corresponds to four actions played at each set. For the RO type subject, x_1^{iRO} = ($x_{1,1}^{i,RO}$, $\cdots$, $x_{1,4}^{i,RO}$) = (3, 3, 2, 4) and $\hat{x}_1^{i,RO}$ = 1. Similarly, for the PM type, x_1^{iPM} = (4, 3, 3, 3) and $\hat{x}_1^{i,RO}$ = 1. At the set 2, the subject played M instead of D (what was supposed to play for the PM type), we count three U actions as the matching actions for the PM type and D as a mismatching action. Similarly, at the set 3, since the subject played D more than one time, we

Set 1-4	Game 1	Game 2	Game 3	Game 4
Round 1	U	U	ALL	ALL
Round 2	U	U	ALL	ALL
Round 3	U	U	ALL	ALL
Round 4	U	U	ALL	ALL

Set 1-4	Game 1	Game 2	Game 3	Game 4
Round 1	U	U	U	U
Round 2	U	D	U	MU
Round 3	U	U	M	MD
Round 4	D	D	D	D

Table 3.12: (Up) The RO type subject's model play pattern example, (Down) The PM type subject's model play pattern example

count the excess number of D action as the mismatching action. For the HM type, x_1^{iHM} = (0, 1, 0, 0) and $\hat{x}_1^{i,HM} = 0$.

(1) Rational Optimizer (RO)

In game 1 and 3, the RO type subject is supposed to play U for every round of every set. In the game 2 and 4, since all actions are expected to give the exactly the same payoff, any choice will be accepted as the optimal choice. We identify the RO type from the other types by the pattern of plays in the game 1 and 3.

(2) Probability Matcher (PM)

We can distinguish the PM type subject from the other type by observing their mixing proportion of plays: the proportion must be kept across the sets and matched to the given distribution of the virtual players' type. In the game 1, the PM type subject is supposed to play U three times and D one times. The order of play is irrelevant as long as the frequency is kept to U : D = 3 : 1 at every set. In the game 2, U and D is supposed to play two times respectively at every set. This proportion of mixing play is matched to the given distribution of the virtual player's type (L: R = 1: 1). In the game 3, to match the given distribution of the virtual player's type, the action U is supposed to play two times, M and D is supposed to play one time respectively at every set. In the game 4, all

	Game 1	Game 2	Game 3	Game 4
Set 1	M4	M4	B4	B4
Set 2	M4	M4	B4	B4
Set 3	M4	M4	B4	B4
Set 4	M4	M4	B4	B4

Table 3.13: The HM type subject's model play pattern example

actions (U, MU, MD, D) are supposed to play one time at every set.

(3) Hedging Matcher (HM)

The HM type subject is supposed to play the hedging actions which always provide some positive amount of payoff. At each game, such hedging action (Game 1 and 2 : action M, Game 3 and 4 : action B) would provide disounted payoff so that it cannot be beneficial than playing the rational action. The amount of discoutn may vary by the individual result of the risk aversion test. For this reason, the pattern of plays that the HM type subject will be easily distinguished from the other type subject.

3.6.2 SDM Model Specification

3.6.2.1 Information and Payoff Function

At each round, subjects face the information set (a^1, b^1, p^1; a^2, b^2, p^2). Each subject will be informed of their own and their partner's strategic environment respectively. This setting justifies the assumption that the information about the game structure is common knowledge. In every game, subjects will be notified that they play a player 1's role and their partner play a player 2's role. a_1 and b_1 is the minimum and the maximum number that the player 1 can choose respectively. We denote the player 1's choice interval $[a^1, b^1]$. p^1 is the target number for the player 1. x^1 is a choice action of the player 1. At this step, we assume that $x^1 \in [a^1, b^2]$. Notations for the player 2 are defined in similar way. At each round, subjects earn payoffs from their choice action x^1 based on the prediction for the player 2's choice x^2. We define the payoff function $P(x^1|a^1,b^1, a^2, b^2; x^2)$

as following;

$$P(x^1|a^1,b^1,a^2,b^2;x^2) = 100 \times \left[1 - \frac{|log\left(\frac{x^1-p^1\cdot x^2}{|a^1-p^1\cdot b^2|}+1\right)|}{log\left(b^1-a^1+p^1(b^2-a^2)\right)}\right] \equiv 100 \times \left[1 - \frac{|log\left(\frac{x^1-p^1\cdot x^2}{|\underline{e}^1|}+1\right)|}{log\left(\overline{e}^1-\underline{e}^1\right)}\right].$$

$\overline{e}^1$ and $\underline{e}^1$ denotes the largest and smallest possible difference between x^1 and $p^1 \cdot x^2$ respectively. That is, $\overline{e}^1 \equiv b^1$ - $p^1 \cdot a^2$ and $\underline{e}_1 \equiv a^1$ - $p^1 \cdot b^2$. For simplicity, we denote $P^1(x^1|a^1,b^1, a^2, b^2, x^2) \equiv P^1(x^1|e^1, x^2)$ where $e^1 = (\overline{e}^1, \underline{e}^1)$.

Basically the player 1 can maximize own payoff by minimizing the difference x^1 - $p^1 \cdot x^2$ with respect to own prediction about x^2. Suppose that the player 1 has a belief that his/her partner chooses a certain action x^2 in an interval $[\underline{x}^2, \overline{x}^2] \subseteq [a^2, b^2]$ and each x^2 is uniformly distributed within the interval $[\underline{x}^2, \overline{x}^2]$. We denote such belief as a probability distribution $f^1(x^2|\underline{x}^2, \overline{x}^2)$. Then, the expected payoff from the player 1's choice x^{1*} is given by

$$E\left[P^1(x^{1*}|e^1,x^2)\right] = \int_{\underline{x}^2}^{\overline{x}^2} \left[100 \times (1 - \frac{|log\left(\frac{x^{1*}-p^1\cdot x^2}{|\underline{e}^1|}+1\right)|}{log\left(\overline{e}^1-\underline{e}^1\right)})\right] f^1(x^2|\underline{x}^2,\overline{x}^2)\, dx^2.$$

Then, the optimal choice x^{1*} that maximizes $P^1(x^{1*}|e^1, x^2)$ satisfies the equation

$$(x^{1*} - p^1 \cdot \underline{x}^2 + |\underline{e}^1|)(x^{1*} - p^1 \cdot \overline{x}^2 + |\underline{e}^1|) = (|\underline{e}^1|)^2.$$

Observation 2. (Payoff Function)

(1) Separate a point-based prediction and an interval-based prediction

(2) Payoff normalization across different game structure

We inserted the asymmetric concavity by using a log function to effectively separate a point-based prediction and an interval-based prediction. That is, the payoff decreases with concave manner when $|x^1$ - $p^1 \cdot x^2| > 0$ and the slope of function is different when x^1 - $p^1 \cdot x^2 > 0$ and x^1 - $p^1 \cdot x^2 < 0$. This modification is required to distinguish the subject who uses an exact number as a prediction

from the one who uses the interval that may have the same mean under the uniform distribution. For example, suppose that the player 1 has $p^1 = 1.5$ and $[a^1, b^1] = [100, 900]$. Consider two cases in which (1) the player 1 predicts the player 2 plays exactly 300 and (2) the player 1 has a belief that player 2's choice is uniformly distributed within the interval [100, 500]. In the former case, the player 1 may choose 450 to capture $p^1 \cdot x^1 = 1.5 \times 300$. In the latter case, the player 1 may choose $50(\sqrt{205} - 4) \simeq 515.89$ to best respond to own belief. [Costa-Gomes and Crawford (2006)] (henceforth CGC) adopted a kinked linear function that imposes different linear slopes at two different intervals. Even though they avoided a simple linear function, they couldn't distinguished the choice from the point-based prediction and that from the interval-based prediction. In CGC, two different predictions will be led to the same optimal choice 450. [5] Imposing an asymmetric concavity on the payoff function can be useful device to aviod those two belief structures.

Second, we normalize the payoff function by using the largest and smallest difference of prediction $\overline{e}_1$ and $\underline{e}_1$. This normalization will help subjects put similar weights on every game. Since each game has different choice interval, the extent of the prediction error also can be different by games. The normalization adjusts the unit of payoffs so that the extent of payoff loss from the prediction error will be relatively measured.

3.6.2.2 Statistical Model Specification

In the SDM experiment, we focus on the identification of the individual strategy with respect to the mixing propensity which requires the estimation of an individual strategy and the mixing propensity type respectively. For this concern, the estimation will be conducted by a two-layer process. In the first layer, we will fix (or assume) the individual mixing propensity k among four types (RO, PM, UM, or HM). Then, given fixed individual type, we will guess a type-specific strategy

[5] Distinguishing the point-based and the interval-based predction is a senstive issue as long as we are based on the iteratve dominance model. In either Lk or CH model, the L1 strategy will be based on the interval-based prediction rather than the point-based prediction. That is, both models assume that L1 strategy users believe that the L0 player plays randomly over the interval. On the other hand, the point-based prediction is different in that the arbitrary belief anchors on a certain point in the interval. For that reason, separating those two cases will provide a useful evidence to confirm that individuals develop their prediction based on the belief that L0 players exist.

$s^i(k)$ of the subject i with respect to such fixed type. We allows multiple type-specific strategies. In the next subsection we will discuss about how we can guess the type-specific strategies from the actual data. Having such set of strategies $S^i(k)$, we will use the maximum likelihood method to estimate the probability distribution of likelihood for each strategy. From the result of the estimation, we pick the most probable type-specific strategy $s^{i*}(k)$ for type k. In the second layer of the estimation, we collect the most probable type-specific strategies for each type k and define such set of four type-specified strategies $S^{i*} \equiv\{s^{i*}(\text{RO}), s^{i*}(\text{PM}), s^{i*}(\text{UM}), s^{i*}(\text{HM})\}$ as a set of types for the second layer. Among those four types of R^i, we estimate the most probable type by using the maximum likelihood method. In sum, we conduct the estimation in the two stage; first, among the strategies within each type and, second, among the type-specific strategies. For this reason, the below specification generally depicts the maximum likelihood method we will use for each layer of estimation. We generally denote each type s and the set of s as S. In the first layer, $s^i(k)$ and $S^i(k)$ (for each k) will replace s and S repectively. Similarly, $s^{i*}(k)$ and S^{i*} will replace s and S respectively. As a result, two-layer estimation will specify the most probable mixing propensity type and strategy combination for each individual.

$a^i_{j,l}$ and $b^i_{j,l}$ is subject i's lower and upper bound in the lth round of the set j respectively. $x^i_{j,l}$ is the subject i's unadjusted guess at the lth round of the set j. For the concern that $x^i_{j,l}$ is chosen at the out of bounds, we redefine an adjusted guess $R(x^i_{j,l}) \equiv min\{b^i_{j,l}, max\{a^i_{j,l}, x^i_{j,l}\}\}$ which restricts actual choice into the interior of bounds at each round. I define a target guess for a type s individual at the lth round of the set j $t^s_{j,l}$. That is, $t^s_{j,l}$ is the exact number to be chosen from a type s individual at the game of lth round of the set j. $T^{i,s}_{j,l} \equiv [t^s_{j,l}$ - 0.5, $t^s_{j,l}$ + 0.5] $\cap$ $[a^i_{j,l}, b^i_{j,l}]$ is a target bound for an adjusted guess of the type s individual i in the lth round of the set j. That is, $T^{i,s}_{j,l}$ restricts choosable bound around the exact target $t^s_{j,l}$ with respect to a small possibility of error ($\pm$0.5). Since we assume that all subjects can find the correct guess by using the calculation panel, we only allow very narrow range around the exact target guess.

$\varepsilon^s \in$[0, 1] is a type-specific error rate of adjusted guess and $d^s(R(x^i_{j,l}), \lambda)$ is a type s individual's error density with a precision level λ for the adjusted guess in the lth round of the set j. For preci-

sion level λ, we assume λ to be the same across the sets and rounds. I assume that ε^s is identically and independently distributed ("i.i.d.") over all rounds. Also I assume that all individuals are risk neutral.

$P^i_{j,l}(x|\, y)$ is a subject i's payoff from an own guess x given his/her partner's guess y at the lth round of the set j. From this payoff, we define a type s individual i's expected payoff in the lth round of the set j as $P^{i,s}_{j,l}(x)$:

$$P^{i,s}_{j,l}(x) \equiv \int_{a^i_{j,l}}^{b^i_{j,l}} P^i_{j,l}(x\,|y) f^s_{j,l}(y)\, dy$$

where $f^s_{j,l}(y)$ is a density of y that is distributed according to type s's belief.

We assume that a "spike-logit" shape of error[6]. With this assumption, $d^s(R(x^i_{j,l}), \lambda)$ is defined as :

$$d^s(R(x^i_{j,l}),\lambda) = \begin{cases} \dfrac{exp\left[\lambda P^{i,s}_{j,l}(R(x^i_{j,l}))\right]}{\int_{[a^i_{j,l}, b_{j,l}]\backslash T^{i,s}_{j,l}} exp\left[\lambda P^{i,s}_{j,l}(z)\right]} & for\, R(x^i_{j,l}) \in [a^i_{j,l}, b_{j,l}]\backslash T^{i,s}_{j,l} \\ 0 & for\, R(x^i_{j,l}) \in T^{i,s}_{j,l}. \end{cases}$$

We define $n^{i,s}_j$ the number of rounds that type s subject i plays the exact type s guess at the set j and $N^{i,s}_j$ a collection of such rounds in the set j. We define vectors $x^i_j \equiv (x^i_{j,1}, x^i_{j,2}, \cdots, x^i_{j,5})$ and $R(x^i_j)$ $\equiv (R(x^i_{j,1}), R(x^i_{j,2}), \cdots, R(x^i_{j,5}))$ the subject i's guesses and adjusted guesses in the set j respectively.

By consideration that type s individual i chooses the (adjusted) guess $R(x^i_{j,l})$ with probability 1 - ε^s, we have a sample density for $R(x^i_{j,l})$ in the set j $d^s(R(x^i_j), \varepsilon^s, \lambda)$:

$$d^s(R(x^i_j), \varepsilon^s, \lambda) \equiv (1-\varepsilon^s)^{n^{i,s}_j} (\varepsilon^k)^{5-n^{i,s}_j} \prod_{l \notin N^{i,s}_j} d^s(R(x^i_{j,l}), \lambda).$$

Similarly, we define $R(x^i) \equiv (R(x^i_1), R(x^i_2), \cdots, R(x^i_8))$ as the subject i's adjusted guess for entire experiment and $d^s(R(x^i), \varepsilon^s, \lambda)$ as a sample density function for entire experiment :

[6] I assumed the distribution of error with the consideration that the error rate to be decreased with convex rate. The use of calculation panel may reduce the possibility of error that purely comes from the miscalculation. Moreover, with the consideration of rounding, I allowed the range of exact choice to include the closest integer.

$$d^s(R(x^i), \varepsilon^s, \lambda) \equiv \overset{[}{j} = 1]8 \prod d^s(R(x^i_j), \varepsilon^s, \lambda).$$

Now we define $z^{i,s}_j$ a type s indicator for the subject i where $z^{i,s}_j = 1$ if the subject is of type s and $\sum_{s \in S} z^{i,s}_j = 1$. $\varepsilon \equiv (\varepsilon^s)_{s \in S}$ is a vector of error rates for all types and $z^i_j \equiv (z^{i,s}_j)_{s \in S}$ is a vector of type indicators in the set j. From this definition, we have a subject i's log-likelihood function $L(z^i_j, \varepsilon, \lambda \mid R(x^i_{j,l}))$:

$$L(z^i_j, \varepsilon, \lambda \mid R(x^i_{j,l})) \equiv \sum_{s \in S} z^{is}_j ln \left[d^s(R(x^i), \varepsilon^s, \lambda) \right].$$

3.6.2.3 Endogenous Type Categorization

For categorization of individual type-specific stategies, we use a actual choice data to guess the candidates for type-specific strategies. Different from the ODM experiment, the SDM experiment does not provide explicit belief formation that individuals are expected to follow. The RO type, who might use one single action for whole experiment, is relatively easy to identify. On the other hand, an identification of other types, PM, UM, and HM types, needs to find not only strategies subjects may use but also proportion among those strategies. This consideration enforces us to (theoretically) try the infinite number of different mixing proportions with different strategies. For example, PM type who adopts L1 strategy and L2 strategy with mixing proportion 0.75 and 0.25 and who adopts with mixing proportion 0.50 and 0.50 should be classified as different types. To avoid the difficulties, we need to restrict to our attention to some set of types. To this end, we will exploit the actual choice observation to guess the probable strategies and mixing proportion among them by individuals.

(1) Rational Optimizer (RO) type

RO types always have a fixed set of type-specific strategies. That is, RO - L1 type subject is expected to choose L1 strategy always and RO - L2 type subject is expected to choose L2 strategy always, so on. Since we restrict our attention to only four strategies (L1, L2, L3 and NE), this

Game	RO-L1	RO-L2	RO-L3	RO-NE
$\alpha n_2 \beta n_4$	419.4	361.1	440.3	500
$\beta n_4 \alpha n_2$	515.9	629	541.8	750
$\delta n_3 \beta n_1$	678.3	363.9	373.1	300
$\beta n_1 \delta n_3$	330.8	339.2	181.9	150
$\beta n_1 \beta n_2$	350	173.9	122.5	100
$\beta n_2 \beta n_1$	347.8	245	121.75	100
$\delta n_3 \gamma n_3$	300	550	363	550
$\gamma n_3 \delta n_3$	500	330	500	500

Table 3.14: Rational type's pattern of play in the SDM experiment

assumption restricts the set of the RO type's strategies. That is, for RO type, we construct the set for RO type as $S^i(RO)$ = {RO-L1, RO-L2, RO-L3, RO-NE} for any individual i. From four types, we find the specific strategy $s^{i*}(RO)$ that maximizes the likelihood among them.

At each round of the sets, the RO type subject is supposed to play a certain action correspond to that strategy. For example, the RO-NE type subject may play the action correspond to NE strategy of each round. While the most of the rounds allows distinction of different strategies, the game $\delta n_3 \gamma n_3$ and $\gamma n_3 \delta n_3$ allows sharing the same number for the different strategies. In the game $\delta n_3 \gamma n_3$, L2 player and NE player can play the same choice 550. Similarly, in $\gamma n_3 \delta n_3$, L1, L3 and NE players who may play 500 will not be distinguished. To distinguish them, we need to rely on the records from the calculation panel. In $\delta n_3 \gamma n_3$, L2 player may use the calculation panel to calculate L1 partner's choice 500. And then, by putting 500 into own calculation panel or conducting own calculation may lead to have 550 as a best-response choice. NE player, to arrive to 550, may start with own choice for L1 strategy and have 300 for the initial calculation. Different from other types, NE type may repeatedly use (more than 3 times) the calculation panel to arrive NE strategy.

(2) Probability Matcher (PM) type

For the identification of PM type, we need to specify not only strategies but also the mixing proportion among the strategies. In case of the RO type, picking a certain strategy from the fixed set is enough for idenfication of the type-specific strategy. On the while, PM types are allowed

PM - L1+L2	Round 1	Stg.	Round 2	Stg.	Round 3	Stg.	Round 4	Stg.	Round 5	Stg.
$\alpha n_2 \beta n_4$	419.4	L1	361.1	L2	419.4	L1	361.1	L2	419.4	L1
$\beta n_4 \alpha n_2$	515.9	L1	515.9	L1	515.9	L1	521.7	L2	521.7	L2
$\delta n_3 \beta n_1$	678.3	L1	363.9	L2	363.9	L2	678.3	L1	678.3	L1
$\beta n_1 \delta n_3$	339.2	L2	330.8	L1	330.8	L1	330.8	L1	339.2	L2
$\beta n_1 \beta n_2$	350	L1	173.9	L2	173.9	L2	350	L1	350	L1
$\beta n_2 \beta n_1$	245	L2	347.8	L1	245	L2	347.8	L1	347.8	L1
$\delta n_3 \gamma n_3$	550	L2	300	L1	300	L1	300	L1	550	L2
$\gamma n_3 \delta n_3$	330	L2	500	L1	500	L1	330	L2	500	L1

Table 3.15: A PM type subject's pattern of play at SDM experiment

to use more than one strategy with respect to their own belief structure. For this reason, we need to specify the multiple strategies they may adopt and the frequency how often the strategies are used together. The way how we can guess them at the same time is the main concern for the identification of PM type's type-specific strategy.

For the identification of (pure) strategies, we restrict our focus to 11 different combinations of strategies. Since we assume that subjects use only four pure strategies (L1, L2, L3, and NE), PM type subjects can have (i) 6 different combinations of 2-strategy case : L1 + L2, L1+ L3, L1 + NE, L2 + L3, L2 + NE, L3 + NE (ii) 4 combinations of 3-strategy case : L1 + L2 + L2, L1 + L2 + NE, L2 + L3 + NE, L1 + L3 + NE, (iii) only combination of 4-strategy case. As an example, we consider a L1 + L2 case in the follwing table. The example table describes PM type subject who uses L1 and L2 strategies and mixes them with proportion L1 : L2 = 3 : 2.

Each two columns describes the actions of each round and corresponding strategies. The first column shows choices of the PM type subject of L1 + L2 strategy. The subject may use either pure L1 strategy or L2 strategy. The next column shows the corresponding strategies for each choice. For example, subject's action 419.4 (of the first column) at the game $\alpha n_2 \beta n_4$ of round 1 corresponds to L1 strategy (of the second column). For PM type, we first identify which strategies are used in each set and then find an average proportion among them. In this process, we exclude choices that does not have corresponding strategies from L1, L2, L3, and NE. Once we have average proportion among the strategies, we round up/down it to be fitted with 5-round setting. Similarly, PM

type subject who uses different collection of strategies with different mixing proportion can be classified.

This way of guessing process is based on the assumption that the PM type subject would keep the same mixing proportion across the sets for the same collection of pure strategies. This assumption allows us to guess the mixing proportion from an observed average proportion of choices. From a vector $x^i \equiv(x^i_1, x^i_2, \cdots, x^i_5)$, we can find the frequency of each choice that corresponds to each strategy. Then, we will use the the observed frequency as the guess for the mixing proportion.

By using this process, we will find at most five candidates for PM type-specified strategies from each set-wise observation. For example, consider the subject i's choices in the set 1 that shows the proportion among each strategy as L1 : L2 = 2 : 1 : 1 : 1. We will name a type-specified strategy PM - 1 which shows the mixing proportion L1 : L2 : L3 : NE = 2 : 1 : 1 : 1. Then, we compare the actual choices in other sets (set 2~8) with this PM - 1 strategy. Similarly the actual choices in set 2 shows L1 : L2 : L3 : NE = 1 : 2 : 1 : 1. Then, we may have another type-specified strategy PM - 2 with the mixing proportion L1 : L2 : L3 : NE = 1 : 2 : 1 : 1, so on. As a result, we may have at most eight different guesses for the PM type-specified strategies.

For the matter of counting $n^{i,s}_j$, we will consider all the possible cases. Consider the PM - 1 at the set 2 of the above example. For L1 and L2 strategy, the rounds that corresponds to them will be counted as the fitting ones. For one L3 and one NE strategy, they might be considered to be fitting one. Compare to PM - 1, one more L3 strategy is appeared at the set 2. Then, we pick one of two L3 choices as the deviating choice from NE strategy. For the concern that different choice of deviating choice may induce different estimation result, we may consider both cases as deviation and choose the case that gives higher likelihood result.

(3) Hedging Matcher (HM) type

For HM type, we allows the HM type subjects to choose non-Lk choices and this relaxation leads us to another difficulty; whether to consider such choices as a strategic hedging behavior or not. For this concern, we exploit two assumptions : (1) any hedging behavior will be based on the belief

HM - L1+L2	Round 1	Round 2	Round 3	Round 4	Round 5	L1	L2
$\alpha n_2 \alpha n_4$	380	400	415	375	365	419.4	361.1
$\alpha n_4 \alpha n_2$	516	580	600	550	629	515.9	629
$\alpha n_4 \beta n_1$	400	450	650	550	600	678.3	363.9
$\beta n_1 \alpha n_4$	332	333	333	338	335	330.8	339.2
$\beta n_1 \beta n_2$	350	180	250	200	300	350	173.9
$\beta n_2 \beta n_1$	250	300	333	325	280	347.8	245
$\delta n_3 \gamma n_3$	300	550	350	500	400	300	550
$\gamma n_3 \delta n_3$	350	400	500	450	400	500	330

Table 3.16: A HM type subject's pattern of play at SDM experiment

that consists of multiple Lk strategies, (2) any hedging behavior based on a certain belief will be bounded by the interval formed from the belief. For example, consider some HM type subject at the game $\alpha n_2 \beta n_4$ who has a belief that his/her partner may play either L1 or L2 strategy with some mixing proportion. Then, his/her belief may form a bound for his/her hedging choice and that bound may depend on the L1 (419.4) and L2 (361.1) strategies. Given his/her belief of L1 and L2, choosing any numbers outside of the interval formed by 419.4 and 361.1 (in this case, [361.1, 419.4]) is always weakly dominated strategy by some other number locates in the interior of the interval. From these assumptions, we can infer that any HM type subjects may choose the number that is located within the interval that is bounded by Lk strategies he/she based on. This inference, even though it allows broader range than UM or PM type allows, provides ground to identify whether the subject shows consistent HM type behavior. Consider an example for HM type subject with strategies L1+L2. The table shows that all choices taken by the HM type subject are consistently located in the interval of L1 and L2 strategies. This pattern of play that the HM type subject can be distinguished from randomly playing subjects.

For HM type, specifying the adopted strategies will be enough to identify the type. Different from the UM or PM type, we focuses on the identification of the HM subject itself. So we consider 6 different combinations of two strategies that can form bounds : L1 + L2, L1 + L3, L1 + NE, L2 + L3, L2 + NE, and L3 + NE.

Even though the wide range of target guess is allowed, we can effectively identify the consistent

behavior from the HM type. First, every combination of strategies has the different formation depends on the game structure. For example, L1 + NE may have the broadest range in most of games but not for game 1 and 8. For this reason, even the random-likely behaving subject consistently chooses the number between L1 and NE strategies, the subject may show consistent deviation at game 1 and 8 of every round. Moreover, the formation of the interval in the game also changes. In the game 1, 2 and 7, L1 strategy locates lower than NE strategy while NE locates lower than L1 in the set 3, 4, 5, and 6. From this structure, the HM type subjects with a certain belief need to know the exact range that will bound his/her optimal hedging choice. Moreover, the range that is covered by such combinations are at most equal to smaller than a half of entire choice interval and apparent game environment randomly changes across the sets. For this reason, the probability that the subject who consistenly chooses numbers bounded by certain pure strategies is statistically insignificant.

4 References

[1] Arad, A., Rubinstein, A., 2012, The 11–20 Money Request Game: A Level-k Reasoning Study, The American Economic Review 102(7), 3561-3573(13)

[2] Ashworth, T., 1980, Trench Warfare, 1914-1918: The Live and Let Live System, NewYork: Holmes & Meier

[3] Athey, S., Bagwell, K., 2008, Collusion With Persistent Cost Shocks, Econometrica 76, 493–540

[4] Axelrod, R., Hamilton, W., D., 1981, The Evolution of Cooperation, Science 211, 1390-1396

[5] Benoît, J.-P., Krishna, V., 1993, Renegotiation in Finitely Repeated Games. Econometrica 61, 303–323

[6] Bossert and Suzumura, 2009, Consistency, Choice and Rationality, mimeo

[7] Caplin, A. and Dean, M., 2007, Search, Choice, and revealed preference, Theoretical Economics 6, 19 - 48

[8] Camerer, C., Ho., T-H., Chung, J-K., 2004, A Cognitive Hierarchy Model of Games, The Quarterly Journal of Economics 119, No. 3, 861-898

[9] Camerer, C., Ho., T-H., Chung, J-K., 2007, Self-tuning Experience Weighted Attraction Learning in Games, Journal of Economic Theory 133, 177 – 198

[10] Chu, Y. and Chu, R., 1990, The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings, The American Economic Review 80(4) : 902 - 911

[11] Costa-Gomes, M. A., Crawford, V. P., Broseta, B., 2001, Cognition and Behavior in Normal-Form Games: An Experimental Study, Econometrica 69, Issue 5, 1193–1235

[12] Costa-Gomes, M. A., Crawford, V. P., 2006, Cognition and Behavior in Two-Person Guessing Games: An Experimental Study, The American Economic Review 96, 1737-1768

[13] Cox, J. C. and Grether, D. M. , 1996, The preference reversal phenomenon: Response mode, markets and incentives, Economic Theory 7(3), 381 - 405

[14] Crawford, M. A. , Costa-Gomes, M. A., Iriberri, N., 2013, Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications, Journal of Economic Literature 51, No. 1, 5-62

[15] Cripps, M.W., Mailath, G. J., Samuelson, L., 2004, Imperfect Monitoring and Impermanent Reputations. Econometrica 72, 407–432.

[16] Cripps, M.W., Thomas, J.P., 2003, Some Asymptotic Results in Discounted Repeated Games of One-Sided Incomplete Information, Mathematics of OR 28, 433–462

[17] Cripps, M.W., Thomas, J.P., 1995, Reputation and Commitment in Two-Person Repeated Games Without Discounting. Econometrica 63, 1401–1419

[18] Edward Lin, Nov. 1. 2013, "One Of The Best Korean LoL Players Got Banned For 1000 Years." 2P.Com, http://2p.com/2598150_1/One-Of-The-Best-Korean-LOL-Players-Got-Banned-For-1000-Years-by-EdwardsLin.html

[19] Ely, J.C., Välimäki, J., 2003, Bad Reputation, The Quarterly Journal of Economics 118, 785–814

[20] Ely, J.C., Välimäki, J., 2002, A Robust Folk Theorem for the Prisoner's Dilemma, Journal of Economic Theory 102, 84–105

[21] Farrell, J., Maskin, E., 1989, Renegotiation in repeated games, Games and Economic Behavior 1, 327–360

[22] Fehr, E., Schmidt, and K. M., 1999, A Theory of Fairness, Competition, and Cooperation, The Quarterly Journal of Economics 114, No. 3, 817-868

[23] Friedman, J. W., 1971, A Non-cooperative Equilibrium for Supergames, The Review of Economic Studies 38, 1–12

[24] Fudenberg, D., Levine, D., 1983, Subgame-perfect equilibria of finite- and infinite-horizon games, Journal of Economic Theory 31, 251–268

[25] Georganas, S., Healy, P.J., Weber, R.A., 2015, On the persistence of strategic sophistication, Journal of Economic Theory 159, Part A, 369-400

[26] Grether, D.M. and Plott, C. R., 1979, Economic Theory of Choice and the Preference Reversal Phenomenon, The Amerrican Economic Review 69, 623 - 638

[27] Hillas, J., 1994. Sequential Equilibria and Stable Sets of Beliefs, Journal of Economic Theory 64, 78–102

[28] Hörner, J., Sugaya, T., Takahashi, S., Vieille, N., 2011, Recursive Methods in Discounted Stochastic Games: An Algorithm for delta goes to 1 and a Folk Theorem, Econometrica 79, 1277–1318

[29] Hörner, J., Takahashi, S., Vieille, N., 2013, Truthful Equilibria in Dynamic Bayesian Games (SSRN Scholarly Paper No. ID 2370574), Social Science Research Network, Rochester, NY

[30] Humprey, S. J., 2006, Does Learning Diminish Violation of Independence, Coalescing and Monotonocity?, Theory and Decision 61, 93 - 128

[31] Kreps, D. M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic Theory 27, 245–252

[32] Mailath, G. J., Samuelson, L., 2001, Who Wants a Good Reputation?, Review of Economic Studies 68, 415–441

[33] Mailath, G., Samuelson, L., 2006, Repeated Games and Reputations: Long-Run Relationships (OUP Catalogue), Oxford University Press

[34] Manzini, P. and Mariotti, M. , 2007, Sequentially Rationalizable Choice, The American Economic Review 97(5), 1824 - 1839

[35] Masatlioglu, Y., Nakajima, D., Ozbay, E.Y. , 2012, Revealed Attention, The American Economic Review 102(5), 2183 - 2205

[36] Masatlioglu, Y. and Nakajima, D., 2012, Choice by Iterative Search, mimeo

[37] McLennan, A., 1985, Justifiable Beliefs in Sequential Equilibrium. Econometrica 53, 889–904

[38] Neimark, E. D., Shuford, E. H., 1959, Comparison of Predictions and Estimates in a Probability Learning Situation, Journal of Experimental Psychology, Vol 57(5), 294-298

[39] Pearce, D. G., 1987, Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation (Cowles Foundation Discussion Paper No. 855). Cowles Foundation for Research in Economics, Yale University

[40] Ray, D., 1994, Internally Renegotiation-Proof Equilibrium Sets: Limit Behavior with Low Discounting. Games and Economic Behavior 6, 162–177

[41] Rubinstein, A., 2002, Irrational Diversification in Multiple Decision Problems, European Economic Review 46, 1369–1378

[42] Van Damme, E., 1989, Renegotiation-proof equilibria in repeated prisoners' dilemma. Journal of Economic Theory 47, 206–217

[43] Vulkan, N., 2000, An Economist's Perspective on Probability Matching, Journal of Economic Surveys, Vol. 14, No. 1, 101-1

[44] Song, H., Oct. 22. 2013, "Apdo got banned for 1000 years and resigned from Korean National E-sports Team.", Thisisgame.com, http://www.thisisgame.com/webzine/news/nboard/4/?n=50415