論 文 の 内 容 の 要 旨

論文題目　The Infinite Regress Problem in Choice of a Collective Decision Procedure: An Axiomatic Study
（集団的意思決定における手続き正当化を巡る無限後退の解消に関する研究）

氏　　名　　鈴木　貴大

1. Introduction

Consider a situation where a group of individuals faces a decision-making problem without an agreement on the procedure to aggregate their opinions over the set of alternatives X. As Nurmi (1992) points out, different procedures can result in different outcomes even if we fix people's preferences over X and we consider familiar set of voting rules such as plurality, runoff, amendment, Borda count, and approval voting. Saari (1992) shows that millions of different ranking of X can be achieved by the choice of scoring rule when there are ten alternatives. These examples imply that the choice of procedure is no less important than the choice of X.

In the social choice theory, axiomatic study of voting rules has long been made based on the implicit premise that a *good* rule is the one that satisfies normative/intuitive criteria such as Condorcet's criterion, unanimity, etc., while great number of negative results, the best known of which would be Arrow's and Gibbard and Satterthwaite's, suggest that there is no *perfect* voting rule. On the other hand, there is another view that a *good* rule is the one that is favored by the group of individuals themselves, although such procedure might fail to satisfy those axioms that social choice researchers esteem. In his *How to reach legitimate decisions when the procedure is controversial*, F. Dietrich formally states this view as Procedural Autonomy (PA). PA demands that the procedure by which the society aggregates voters' procedural judgments should be entirely determined by the procedural judgments within the group. Once we take PA literally, however, we could face an infinite regress problem as follows. When the society faces a decision making problem X, PA demands that the rule to aggregate their opinions over X must be determined by their opinions over such rules. This means the society faces a new decision making problem: how to choose the rules to choose X. Using PA again, it follows that the society needs to aggregate their opinions over the rules to choose the rule to choose X. This process can go on ad infinitum unless there is an ex ante agreement on some meta-level rule.

The purpose of this study is to find a rational answer for this infinite regress problem. Especially, the author proposes and investigates a new solution concept, named weak/strong convergence, in Chapters 2 and 3. In Chapter 4, the boundary problem, i.e. the determination of the set of individuals who have the right to vote, is discussed. Lastly, Chapter 5 states concluding remarks.

2. Convergence in procedural choice

According to PA, the collective procedural judgments is made (just) by the voters' procedural judgments. the author first models this process with a simple model and shows an impossibility result in the introductory chapter. Assume that each individual has a weak-ordered preference over the set of alternatives, the set of rules F^1, the set of rules to choose rule F^2, and so forth. A (level-L) Procedural Choice Rule (PCR) is defined as a function that maps preference profile over $X, F^1, F^2, \dots, F^L$ into the collective preference over $X, F^1, F^2, \dots, F^L$. The first theorem shows an impossibility of the PCR. A PCR E is said to satisfy Inter-Level Consistency (ILC) if it ranks meta-level rule f at least as good as meta-level rule g if and only if E ranks f's outcome at least as good as g's. A PCR E is said to satisfy Arbitrary Focus (AF) if the collective ranking over level 2, for instance, is entirely determined by individuals' preferences over level $2,3,4,\dots$. The theorem shows that if people's meta preferences are not restricted at all, the combination of ILC and AF allows only a degenerated indifferent PCR; it cannot evaluate any rule as *better* than another no matter what preferences people have.

The difficulty partly stems from the assumption that we allow any type of meta-level preference. The model allows such an individual who prefers plurality as a rule and Hare system as a rule to choose the rule while there is no relationship between them. Koray (2000) restricted to consequential meta-preferences, i.e., everyone judges procedures according to their outcomes, and defined a voting rule f as self-selective if it chooses itself among the other possible voting rules. Koray shows that for neutral and unanimous Social Choice Function (SCF), (universal) self-selectivity is logically equivalent to dictatorship. Later, the notion is extended to menus of voting rules (Diss et al. (2012); Diss and Merlin (2010); Houy (2004)). The results by Diss et al. (2012) show that when the population is large, i.e. $n \to \infty$, the probability that the set of $\{\text{Plurality (P)}, \text{Borda (B)}, \text{Antiplurality (A)}\}$ is not stable is about 84.49% (IC) and 84.10% (IAC).

In Chapter 2, the author proposes an alternative solution concept for a consequential society. The idea can be illustrated with an example. Suppose that a society of 14 individuals chooses one of three candidates, $X = \{a, b, c\}$, and there is an ex ante agreement on the set of voting rules, $F = \{P, B, A\}$. This means that the society agrees on the use of either P, B, or A, but there is still no agreement on which of them is appropriate for the specific agenda. When the preference profile over X is given as $L^0_{1-10}\!: abc$, and $L^0_{11-14}\!: bca$ (i.e. individuals $1,2,\dots,10$ prefer a to b and b to c and individuals $11,12,13,14$ prefer b to c and c to a), the three voting rules P, B, and A yield $\{a\}, \{a\}$, and $\{b\}$, respectively. Suppose now that the same society votes on which rule in F to use. If everyone is consequential and is supposed to submit a linear ordering, it is natural to think that the first 10 individuals submit either "PBA" or "BPA," and the remaining four individuals submit "APB" or "ABP". Suppose that they submit the following: $L^1_{1-4}\!: PBA$, $L^1_{5-10}\!: BPA$, and $L^1_{11-14}\!: APB$. For this profile $(L^1_1, L^1_2 \dots, L^1_{14})$, P yields $\{B\}$ while B and A yield $\{P\}$. Note that each P, B, and A (as a rule to choose the rule) ultimately reaches the same outcome $\{a\}$. The author refers to this situation as "the (original) profile L^0 weakly converges to $\{a\}$".

In general, a profile L^0 over X is defined to *weakly converge* to $C \subseteq X$ if there exist such consequentially induced preference profiles such that every level-k rule ultimately results in C (for some $k \in \mathbb{N}$). Note that once the weak convergence occurs, further regress has no effective meaning, for the ultimate outcome is exactly the same no matter which higher-level rule is found to be appropriate in the subsequent regress. For a probability model, IAC (impartial anonymous culture) is assumed unless otherwise noted.

The first and fundamental result is on a triplet of voting rules. It says that a large society, i.e. where the population $n \to \infty$, with the menu of three voting rules among Plurality, Borda, Antiplurality, Hare, Nanson, Coomb, Maximin, and Black, has asymptotically only two possibilities; weak convergence or trivial deadlock. Note that this proposition holds for each of the ${}_8\mathrm{C}_3 = 56$ different menus. The latter, trivial deadlock, is defined as the situation where each voting rule results in distinct singletons. It is derived as a straightforward observation that once the trivial deadlock occurs, the regress 'structure' does not at all change no matter how high of a level we might see. Under such a degenerated structure, further regress has little or no effective meaning. To sum, a simple message from the theorem is that the infinite regress problem is asymptotically degenerated under a large society with such a menu of voting rules.

Here we must note that the trivial deadlock does not provide a unique answer. Therefore, the significance of the theorem depends to some extent on how high the probability of weak convergence is. With software Barvinok, the author has calculated the probability of weak convergence for the 56 menus when there are three alternatives, i.e. $|X| = 3$. The result shows that in each menu, the probability is strictly greater than 96.5%. Especially, 98.8% for the menu $\{P, B, A\}$ and 100% for six menus including $\{B, \mathrm{Hare}, \mathrm{Black}\}$. It is noted that this probability is higher than the probability of stability for $\{P, B, A\}$ (84.10% for $\{P, B, A\}$; Diss et al. (2012)).

Several generalizations and discussions on this theorem are provided in the rest of the second part. The most effective of them is the following. We say that a menu of voting rules F has an *asymptotically weak convergent extension* (AWCE) if there exists $G \supseteq F$ such that the probability of weak convergence goes to one under the menu G as $n \to \infty$. The author has established that any menu F made up of what we call concave scoring rules (a scoring rule is defined to be concave if its assignment of scores can be interpreted with a concave function in the usual mathematical sense) has an AWCE.

3. Determination of the society

In the last part, the author discusses the determination of the society itself: who has the right to vote and who does not are assumed to be given before the procedural choice is made. While this implicit assumption is also used in the related literature (Dietrich 2005; Koray 2000, etc.), who should be entitled to have the right to vote is sometimes controversial especially when the results of the decision can affect unspecified number of people. This problem, often called the boundary problem, is a classical problem in political science.

In this dissertation, the author has studied this problem through the framework of collective decision making, i.e. to determine the society based on individuals' opinions. Special attention is paid to the strategic aspect of such situations.

A nomination rule is a voting rule that maps individuals' opinions on who should be in the group into the collective opinion on who are in the group. Such nomination should be distinguished from ordinary voting in the sense that each individual is a candidate as well as a voter. As a result, it is concerned that some self-interested individual might vote strategically so that one's status would be better off. Against such manipulation, Holzman and Moulin (2013) invented the axiom of impartiality, which demands that a nomination rule exclude the voters' ability to manipulate the ballot in their own favor. At the same time, however, they show several impossibilities. One of which shows the incompatibility between anonymity and impartiality.

Based on these studies, the author has studied the design possibility of impartial nomination rules by considering several domains, i.e. what kind of ballots can be submitted, and the codomains, i.e. the number of elected individuals. A systematic comparison has been made of as many as $(n+1)(3n-5)$ voting settings with corresponding pairs of domains and codomains. These settings include: an Approval Voting setting with fixed electorates and fixed alternatives; and those settings with/without self-approval or abstention. By weakening positive unanimity, the author has discovered a series of positive results that characterize the threshold rule as the unique nomination rule that satisfies (a part of) impartiality, anonymity, neutrality, and negative/weak positive unanimity under various domain-codomain pairs.

4. Conclusion

The present dissertation studies the endogenous choice of voting procedures. When there is no ex ante agreement on the procedure, the justification of a specific procedure can often fall into an infinite regress. The results in Chapter 2 show that such infinite regress can be quite often solved through the notion of convergence if the given consequential society (Chapter 4) is equipped with a reasonable menu of possible voting rules, such as {Plurality, Borda, Anti-plurality}. Especially, Chapter 3 shows various ways to avoid the failure of the convergence (trivial deadlock). With these results, we now have new grounds for justifying the use of a specific procedure without falling into the troublesome infinite regress.