Discussion Paper Series

RIEB

Kobe University

DP2021-07

# Measuring a Normative Expectation: Methodological Decision-making in Intergenerational Social Dilemma Mediation

Yoichiro TSUJI
Junyi SHEN

March 21, 2021

Research Institute for Economics and Business Administration
Kobe University
2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

# Measuring a Normative Expectation: Methodological Decision-making in Intergenerational Social Dilemma Mediation

**Yoichiro Tsuji[*] and Junyi Shen[#, ##]**

## Abstract

The intergenerational social dilemma is a social dilemma that sets up a decision-making process for resource allocation among generations. In principle, there is no future generation; thus, only the current generation can make decisions on resource allocation. Because the current generation is myopic and optimistic, problems due to not taking future generations into consideration arise. To address this dilemma, many experiments have been conducted based on the concept of *future design* to change thoughts and behaviors. In this context, a method is presented for assigning a person to play the role of the future generation (i.e., the imaginary future generation) in the decision-making process of the current generation, and having them participate in the discussion. The impact of this method is significant and effective. However, this imaginary future generation is set up as self-evident to participants in the discussion, and a clear reason for its existence is not given. After reviewing previous studies related to intergenerational social dilemmas and normative expectations, as well as those on *future design*, this paper proposes an experimental design that emphasizes one's internal free-will decision to accept this imaginary future generation, hypothesizing that it can become a shared understanding across generations, and that it will become a normative expectation for the next generation to accept the imaginary future generation.

Keywords: Intergenerational social dilemma; Future design; Social norms; Normative expectation; Imaginary future generation

[*] Corresponding author. Graduate School of Economics, Kobe University, 2–1 Rokkodai, Nada, Kobe 657–8501, Japan yoichiro.2g@gmail.com (Yoichiro Tsuji)

[#] Research Institute for Economics and Business Administration, Kobe University, 2–1 Rokkodai, Nada, Kobe 657–8501, Japan shen@rieb.kobe-u.ac.jp (Junyi Shen)

[##] School of Economics, Shanghai University, 99 Shangda Road, Baoshan 200444, Shanghai, China

## 1. Introduction

How social norms that determine people's behavior are generated is a fundamental question in the social sciences. However, social norms are invisible and difficult to observe. In the fields of economics and sociology, experimental design and quantitative methods have been developed to observe people's behavior under certain conditions in laboratory and field experiments, accumulating a vast amount of research. The same is also the case with observations in the field of social norms research. What factors should be focused on in the generation of social norms have been indicated through accumulated research, and the causal relationships between such factors and outcomes have been evaluated with statistical significance.

Bicchieri (2011) stated that social norms are the rules of people's behavior, and that the premise of such behavior is "normative expectations" that you and the other person can each share one another's expectations. Schram and Charness's (2015) study also showed that, through the disclosure of advice and gains, the choice to be made becomes a "shared understanding" among the group.

In the current study, we propose a way to mediate intergenerational social dilemmas using an experimental design that allows us to observe Bicchieri's idea of normative expectations and Schram and Charness's shared understanding.

The basic structure of social dilemmas is that the rational choices of individuals do not match the optimal choices of society; this is also a major subject of research in the field of social science. Examples of social dilemmas include the problems of global warming and air pollution due to the massive consumption of fossil fuels, and the depletion of forests and water resources due to over-logging and overconsumption, which have great impact not only on human society but also on the global ecosystem. Social dilemmas are treated in economics as external diseconomies or free-rider problems in public goods. In addition, the impact extends across the spatial and temporal scope. In other words, it is a problem that will have a long-term impact not only on the current generation living in the present, but also on future generations that have yet to exist.

The intergenerational social dilemma sets up a decision-making process for resource allocation among generations. In principle, there is no future generation; thus, only the current generation can make decisions on resource allocation. Will the current generation take care of the future generation, or will they impose on them a burden?

To address this intergenerational social dilemma, Saijo (2020) conducted a number of experiments based on several ideas on how to realize *future design*, a decision-making process that takes future generations into consideration. Among them, a method, called "imaginary future generation," assigns a person to play the role of the future generation in the decision-making process of the current generation, and has them participate in the discussion. The impact of this method is

significant and is considered effective. However, the participation of this imaginary future generation is set up as somewhat self-evident and not given a clear reason for existence in the discussion. Therefore, in this study, we propose adding an additional procedure that allows individuals to make decisions on whether to accept this imaginary future generation by their internal free will in the experiments.

The remainder of this paper is organized as follows. Section 2 reviews the design and results of major experimental studies on the structure and structural typology of social dilemmas. Section 3 describes the structure of the intergenerational social dilemma. Section 4 looks at the results of major experimental studies on the purpose of *future design* and methods, such as imaginary future generations. Section 5 discusses the methodological challenges of imaginary future generations. Section 6 defines normative expectations and presents some experiments observing them. Section 7 describes the experimental design and hypotheses proposed in this study. Finally, concluding remarks are presented in the last section.

## 2. Experimental studies on social dilemmas

### 2.1 Structure of the problem

Social dilemmas became an important theme in the social sciences with the pioneering "the tragedy of the commons" (Hardin, 1968), which explained how social resources can be depleted as a result of the rational behavior of each individual.

Hardin describes the process of a common land collapse due to overconsumption as follows: If each herdsman can set cattle loose on a pasture with a limited supply of grass, the herdsman will benefit from each additional cow. However, the loss would be distributed among all herdsmen. If all herdsmen understood this, they would all release their own cows, and the pasture would be devastated. The structure of this problem consists of two elements: first, it is rational for the herdsmen to choose to graze their cattle, comparing the profit and loss per head; second, since all herdsmen choose to do the same, the pasture cannot sustain the amount consumed by the released cattle and will eventually be devastated, even if it regenerates. In other words, it is a structure where the rational actions of individuals have undesirable consequences for the whole.

### 2.2 Formularization

Dawes (1980) stated that the social dilemma is an n-person version of the two-person prisoner's dilemma game, which is formulated as follows:

*(a) The social payoff to each individual for defecting behavior is higher than the payoff for cooperative behavior, regardless of what the other society members do, yet*

*(b) all individuals in the society receive a lower payoff if all defect than if all cooperate. (p.170)*

In other words, betrayal in (a) is a behavioral principle of the domination strategy, which is a rational choice for individuals; however, the gain structure in (b) is such that the gains obtained when everyone cooperates is higher.

If a common resource, such as common land, is considered a "public good," with the property that no one is excluded from its use (non-excludability) and no one competes with other users in its use (non-rivalry in consumption), then there are free riders who do not bear the cost of maintaining the public good (Mankiw, 2017). When the number of free riders increases, there will be no one to bear the cost of maintaining the supply of the public good, and as a result the ability to supply the public good will decrease. This free-rider problem is also established under the behavioral principle and gain structure in social dilemmas, and as a typical problem of social dilemmas, considerable research, especially laboratory experimental studies, has been accumulated.

### 2.3 Types of research

Umino (1991) argued that social dilemmas cannot be solved as long as the gain structure and behavioral principles are taken as given. In the background, rational choice theory states that people, as rational economic agents, act to optimize their own value realization under a recognized choice situation. Based on this, he proposed two ways of solving social dilemmas.

The first is to change the choice situation itself. This is generally manipulated by controlling structural factors and changing the payoff structure (e.g., rewarding those who act cooperatively, punishing those who act treacherously, etc.). The second is to control individual factors such as providing information and knowledge about the situation, increasing expectations of others' behavior and trust in others, increasing group cohesion and sense of group belonging, and publicizing decisions. While these two methods assume that people are rational economic agents, a third method revises this assumption. It brings out elements that do not presuppose rationality (e.g., negative emotions, altruism, and reciprocity) and is considered very important to solving social dilemmas.

In the next two subsections, we will review what findings have emerged from studies aimed at controlling the structural or individual factors mentioned above, or exploring non-rationality such as negative emotions, altruism, and reciprocal behaviors. Table 1 provides a summary of these studies.

Table 1. The summary of studies on social dilemmas

| | Themes | Studies | Main findings |
|---|---|---|---|
| Controlling structural factors | Changes in the payoff structure due to punishment | Fehr and Gächter (2000) | ・The possibility of the payoff being reduced through punishment significantly increases the level of cooperation for potential free-riders who are rational economic agents.<br>・Punishing non-cooperators by paying costs is an altruistic act; nevertheless, the punisher does not expect to elicit cooperation from the punishees when meeting the same person in the future. |
| | Emotional effects in punishment | Fehr and Gächter (2002) | ・Negative emotions toward the free-rider provide justification for imposing punishment.<br>・The altruistic act of punishment was shown to be strongly influenced by emotions. |
| | Punishment by a third party | Fehr and Fischbacher (2004) | ・The reason why third parties who did not benefit from the punishment paid the cost to punish is based on the ideas of equality and that both parties should cooperate. Therefore, a dictator who left a lot to himself/herself is more likely to be punished. |
| | Detrimental effects of punishment on altruism | Fehr and Rockenbach (2003) | ・Trustors can use punishment, but refraining from it can be perceived as a fair action, inducing the trustee to increase cooperative behavior and can be explained by the powerful concept of reciprocity. |
| | Marketization of punishment | Gneezy and Rustichini (2000) | ・After the fine was introduced, perceptions changed, as parents might think that the tardiness has a price due to the fine, and they can buy as much as they need as a service.<br>・Once tardiness is priced, it becomes a commodity subject to consumption.<br>・The above changes in perception are regarded as a transformation of the social norm induced by punishment. |
| | Changes in the payoff structure due to the approval mechanism | Saijo et al. (2018); Masuda et al. (2014) | ・Even without private punishment and/or reward technologies, introducing an approval/disapproval stage to the prisoner's dilemma or the voluntary contribution game facilitates the cooperation level among subjects. |
| Controlling personal factors | Conditional cooperation | Fischbacher et al. (2001) | ・The prediction is that all participants would free-ride completely, but even with conditional cooperation, the experiment showed a selfish tendency to contribute less than the average of other subjects in the group. |
| | Reaction to intentions, not results | Falk et al. (2000) | ・The lack of intention to be fair elicited self-interested responses and people take not only the consequences of an action but also the intentions signaled by that action into account. |
| | Intention-based trust reciprocity hypothesis | McCabe et al. (2003) | ・Choosing to leave the decision to the other person signals that the person is trusting the other person, and so the other person understands the intention and makes a reciprocal choice to share. |
| | Generating indirect reciprocity through reputation | Wedekind and Milinski (2000) | ・The history of donation had an important influence on donors' decisions and the image score was used as an index of indirect reciprocity. |
| | The effect of indirect reciprocity through reputation | Milinski et al. (2002a) | ・To maintain cooperative behavior, people need to understand the possibility of receiving rewards by improving their reputation. |
| | The impact of indirect reciprocity through reputation on voting | Milinski et al. (2002b) | ・Donating to UNICEF, an aid organization that operates under a socially recognized philosophy of altruism, is considered socially significant and relevant to being elected as a suitable representative of a social organization. |

## 2.4 Controlling structural factors

One way to change the payoff structure is to introduce punishment. Since punishments are imposed on those who do not cooperate and reduce their gains, rational economic agents do not become free riders but try to cooperate to avoid punishment. This is not to say that rational economic agents turn to cooperation because of their altruism. Rather, such change is because their payoff can be maximized by cooperation.

◆ *Changes in the payoff structure due to punishment*

Fehr and Gächter (2000) conducted a public goods game experiment to observe the effects of punishment. The subjects were divided into two groups: a Partner treatment, in which the members of the group were not changed, and a Stranger treatment, in which the members were randomly assigned. For each treatment, a condition with/without punishment opportunity was attached. In this condition, the subjects were informed about how much each had invested after making the decisions, and whether

they would be given the opportunity to punish others. In the case of punishment, the punisher pays the cost.

The results of the experiment show that the average investment level in the Stranger treatment with punishment was 58% of the initial endowment, which is higher than the 37% in the case without punishment. As time went by, the average investment level in the case with punishment remained the same, but in the case without punishment, it became closer to the free-rider level (i.e., closer to zero). On the other hand, the average investment level in the Partner treatment with punishment was 85% of the initial endowment. Over time, the average investment level in the case with punishment moved toward full cooperation, while in the case without punishment, it moved toward the free-rider level.

In both treatments, in the case with punishment, the punishment increased when the individual's investment level was below the average of the other members of the group and ceased when it was above. In addition, there was no significant difference in investment levels between the cases without punishment in the two treatments. In contrast, when regressing the punishment amount on the investment level of other members in the group, it was found that the latter had a significant effect in the Stranger treatment, but not in the Partner treatment.

From these results, it seems that punishment has two aspects: one is to change the payoff structure, and the other is to elicit altruism. That is, the possibility of the payoff being reduced through punishment significantly increases the level of cooperation for potential free-riders who are rational economic agents because they are aware of a highly possible threat. In the Stranger treatment, punishing non-cooperators by paying costs is an altruistic act; nevertheless, the punisher does not expect to elicit cooperation from the punishees when meeting the same person in the future.

◆ *Emotional effects in punishment*

What caused the altruistic behavior was observed in Fehr and Gächter's (2000) earlier study. They later conducted a similar public goods game experiment (Fehr and Gächter, 2002), in which they asked the subjects afterward why they punished. The subjects were divided into punishing and non-punishing treatment groups. In the case of punishment, the punisher pays the cost.

The results of the experiment showed that in the treatment with punishment, 84.3% of the subjects punished at least once. Punishment increased when the subjects' investment level was below the average of the other members of the group. In the treatment with punishment, the investment level increased over time, but in the treatment without punishment, it sharply decreased. These results are the same as those of Fehr and Gächter (2000), above. It is worth noting that after the experiment, the subjects were given a questionnaire and were asked to express their level of anger in numbers. The question was, "If three people in a group of four, including yourself, have a cooperative investment level, but only one (who is not you) is significantly lower, what do you think of him (her)?" Subjects

were asked to report their level of anger from 1 (minimum) to 7 (maximum). Forty-seven percent of the subjects indicated anger of 6 or 7, and 37% indicated anger of 5. They were also asked, “If you are in a group of four and the other three people in the group have a cooperative donation level, but you are the only one with a significantly low investment level, what do the other three people think of you?” A total of 74.5% of the subjects indicated anger of 6 or 7, and 22.5% indicated anger of 5.

The two questions can be used to observe whether anger is shared socially, because the former asks what you think, and the latter asks what others think. The results indicate that the person with a lower level of investment becomes the target of social anger. It is believed that negative emotions toward the free-rider provide justification for imposing punishment. In other words, the altruistic act of punishment was shown to be strongly influenced by emotions.

◆ *Punishment by a third party*

In Fehr and Gächter’s (2000, 2002) experiments, the punishment was conducted between the parties with direct interests, although subjects were controlled so that they would never be in the same group again for the sake of not creating a reputation for continued cooperation or reciprocity that would result in retaliation the next time they meet. Thus, do disinterested third parties also impose punishment?

To answer this question, Fehr and Fischbacher (2004) conducted an experiment using two games, a dictator game and a prisoners’ dilemma game, in which a third party imposes punishment. In the dictator game experiment, participants were divided into second-party punishment and third-party punishment conditions. The third party here means that the punisher is not the recipient paired with the dictator but a dictator in another group, to avoid reciprocity among the punishers. In the prisoners’ dilemma game experiment, the subjects were observed to see if they cooperated with each other. In both games, the subjects were anonymous and randomly assigned. In each game, a third party could observe how much money has been moved and punish the subjects with the cost paid by himself/herself.

Predictably, just as the second party, if the third party is a rational economic agent, he/she should not punish at a cost. In terms of equality, the severity of the third party's punishment increases when the dictator is at less than half of his initial endowment.

The results of the experiment showed that in the dictator game, about 60% of the third parties punished the dictator when the amount of money from the dictator to the recipient was less than half of the initial endowment. In this game, recipients were asked about the magnitude of punishment they expected from third parties before money was given by the dictator. Their expectations were higher than the actual punishment. When the amount of money from the dictator to the recipient was less than half of the initial endowment in the second-party punishment condition, the punishment was much

stronger than that of the third-party punishment condition. In the prisoners' dilemma game, 45.8% of third parties punished when one subject cooperated but another one betrayed, and 20.8% punished when both subjects betrayed. The difference was statistically significant. Since the probability of punishment was higher when only one subject betrayed than when both subjects betrayed, this experiment indicates that the norm of valuing cooperation was strongly in effect.

In summary, third parties punished, although they did not have a stake and did not benefit from paying the cost to punish. In addition, based on the ideas of equality and that both parties should cooperate, a dictator who left a lot to himself/herself is more likely to be punished.

◆ *Detrimental effects of punishment on altruism*

While we have seen that punishment changes the payoff structure and elicits cooperative attitudes, some experimental results also indicate the possibility that punishment has detrimental effects. Fehr and Rockenbach (2003) conducted a trust game experiment with the option of punishing or refraining from it. In their experiment, the trustor first decides on a transfer amount from his/her endowment. This amount is then tripled by the experimenter and given to the trustee. Afterward, the trustee decides how much of the tripled amount to return to the trustor. When the transfer is made, the trustor also informs the trustee of the desired return amount. The trustor decides whether to punish the trustee based on the amount returned. The subjects were divided into a trust condition without punishment and an incentive condition with punishment. In the trust condition, there is no incentive for the trustee to return the money, and therefore, a positive return indicates the trustee's altruism or trustworthiness. In the incentive condition, the trustor can either punish or refrain from punishing. Since punishment can be imposed on the trustee in the amount of four points, a return of more than four points implies the trustee's altruism. In a world where there are only rational economic agents, the trustee has no incentive to return, and the trustor knows this, so it follows that he/she will not transfer any amount in the first place.

The results of the experiment showed that 79% of the trustees gave a positive return in the trust condition, and 64% of the trustees returned more than four points in the incentive condition. In both cases, the trustees were shown to have demonstrated altruism. In the incentive condition, when trustors voluntarily refrained from punishment, none of the trustees chose zero as the return amount, and 47% of the trustees chose a return of 15 points or more[1], with a mean of 12.5 points. In contrast, when penalties were imposed, only 13% of trustees chose a return of 15 points, with a mean of six points. This difference was statistically significant. In other words, when trustors were ready to punish the trustees, they did not trust the trustees and transferred less before the punishments were imposed; as a result, they received fewer returns from the trustees. The trustors who imposed punishment transferred

[1] Note that the initial holding of the trustor is 10 points and the transfer amount is tripled by the experimenter.

an average of 6.8 points, while those who refrained from punishment transferred an average of 8.7 points.

This increase in returns to trustors who refrain from punishment can be explained by the powerful concept of reciprocity. That is, trustors can use punishment, but refraining from it can be perceived as a fair action, inducing the trustee to increase cooperative behavior. On the other hand, using punishment to correct an inequitable distribution of income due to the small return amount, when the trustor originally transfers only a small amount (the transfer is multiplied by three, which appears disproportionate from the trustor's point of view), may be perceived as an adversarial action and may induce the trustee to reduce altruistic behavior.

◆ *Marketization of Punishment*

Gneezy and Rustichini (2000) conducted an experiment in which punishment was perceived as a priced service and did not elicit cooperative behavior. They conducted a field experiment in daycare centers in Israel to observe whether fines for parents who were late to pick up their children would reduce tardiness. Ten daycare centers were divided into two groups: a test group with fines and a control group without fines. The experiment lasted 20 weeks, and in the first four weeks, the experimenters only observed the number of tardy parents. Six daycare centers introduced fines at the beginning of the fifth week. The fines were canceled at the beginning of the 17th week without explanation.

The results showed that in the test group, the instances of tardiness increased at a constant rate in the first three or four weeks after the introduction of the fine. Its rate stabilized at a high level and was double that of the initial period without fines. After the fine was canceled, tardiness remained stable. In the control group, there was no noticeable change after the fourth week. These results indicate that the introduction of the fine significantly increased the number of tardy parents. Canceling fines did not affect the number of tardiness instances compared to when fines were present. Moreover, this number was significantly higher in the treatment group than in the control group, despite there being no significant difference between the test group and the control group in the first four weeks.

The introduction of punishment changes people's perceptions of the environment in which they operate. Before fines were introduced, parents would interpret the teacher's waiting until their children were picked up as a generous and non-marketable activity. However, after the fine was introduced, perceptions changed. Parents might think that the tardiness has a price due to the fine, and they can buy as much as they need as a service. Furthermore, once tardiness is priced, it becomes a commodity subject to consumption. Thus, this change in perception is regarded as a transformation of the social norm induced by punishment.

◆ *Changes in the payoff structure due to the approval mechanism*

Saijo et al. (2018) applied the *approval mechanism* (AM) to the prisoner's dilemma (PD) by adding an approval stage after the PD, where each subject can approve or disapprove the other's choice of strategy in the first stage. If both approve the other's strategy, the outcome is the result of the chosen strategies in the PD; however, if ether disapproves, the outcome is the same as if they had mutually defected from the PD. To compare the performance of the compensation mechanism (CM) proposed by Varian (1994), which gives players an opportunity to offer transfers contingent on cooperative action prior to playing the underlying game, the authors conducted a laboratory experiment with three treatments: the AM treatment with three sessions, the CM treatment with three sessions, and the PD treatment with one session. In each session, 20 subjects played the game in 19 periods, and the pairings of the subjects in each period were anonymous and determined in advance so that no two subjects were paired more than once.

The experimental results exhibited that the AM achieved a cooperation rate of 90% in the first period and an 93.2% across periods, while the CM yielded a cooperation rate of 63.3% in the first period and 75.2% across periods. The period-by-period Chi-square test results indicated that the gap in the cooperation rate between the AM and CM was sustained, especially in the first 10 periods. As a control treatment, the PD obtained an average cooperation rate of 7.9%; specifically, it was 11% for the first five periods, and declined to 6% in the last five.

Furthermore, Masuda et al. (2014) designed a minimum AM for a two-person voluntary contribution game (VCM) for the provision of a public good. They experimentally found that introducing the minimum AM facilitated almost full cooperation among subjects. Contributions in the first period averaged 76.9% of endowments, and then rose repeatedly for the first five periods, achieving 97.8% in period 5. Then, cooperation was sustained with an average contribution rate of over 95% in every period except for the final one. In contrast, in the control treatment of VCM without a minimum AM, subjects contributed 15.3% in the first period amd gradually decreased to 5.7% in the last period.

Both studies mentioned above suggest that introducing an approval/disapproval stage to the PD or VCM game facilitates the cooperation level among subjects. A noteworthy feature of the *approval mechanism* is that it does not utilize private punishment and/or reward technologies. However, since personal punishment (or bribe) is generally prohibited in modern societies and legal systems, this feature is viewed as a strength of the *approval mechanism* (Saijo et al., 2018).

### 2.5 Controlling personal factors

- *Conditional cooperation*

Fischbacher et al. (2001) conducted a one-shot public goods game. The subjects were assigned to two types of contribution decisions. The first type was called an "unconditional contribution" and the second "contribution table." A group of four people was formed, and only one was assigned to the "contribution table" type who then must indicate for each of the possible average contribution levels of the other group members how much he/she is willing to contribute to the public good. The other three did not know who that person was.

A rational economic agent would choose to free ride, where zero contribution is optimal for both the subject who contributes conditionally ("contribution table" type) and the others who decide their contributions unconditionally ("unconditional contribution" type). In other words, the prediction is that all subjects in a group will free ride completely.

The results of the experiment showed that 50% (22 persons) of the subjects conditionally cooperated to match the contributions of other subjects while 30% (12 persons) were free-riders, a non-negligible proportion. Even with conditional cooperation, they showed a selfish tendency to contribute less than the average of other subjects in the group. This result shows that there are a certain number of people who do not necessarily free-ride like a rational economic agent whose optimal contribution is close to zero.

◆ *Reaction to intentions, not results*

Falk et al. (2000) conducted a moonlight game experiment that has the advantage of testing the effect of fairness intention. In this game, player A chooses an action $a$ from −6 to 6 for player B. If $a$ is greater than or equal to 0, player B is given $3a$ tokens that are tripled by the experimenter, and if $a$ is less than or equal to 0, the experimenter can take $|a|$ tokens from player B and keep them for player A. Player B, after observing $a$, can choose any action $b$ from −6 to 18; choosing 0 or more is a reward to player A, choosing 0 or less is a punishment for player A. In the case of punishment, player B pays a cost of $|b|$ and can reduce player A's gain by $3|b|$. In the experiment, player A was divided into two types: intention treatment, where player A decides $a$ and his/her decision is meant to send a signal of good or bad intentions, and no-intention treatment, where player A's action is determined by a random device and the intention of this action is unknown.

One prediction of the moonlight game is the self-interested prediction, based on the assumption that all players are selfish and rational. That is, player B will always choose $b = 0$ with no reward or punishment for player A because of the cost. Player A also knows this, so he/she chooses $a = -6$, which is less than zero, to maximize his/her gain. Another prediction is the fairness prediction, assuming that participants are averse to unfairness. In the unfairness aversion model, participants care about their fair and relative share of the overall gains, but this ignores intention, and only the gains outcome is assumed to explain the reciprocal response.

The results of the experiment showed that in the intention treatment, the mean and median of the rewards increased with *a*. Similarly, as the negative magnitude of *a* increased, player B was more willing to impose punishment. These results clearly reject the self-interested prediction. On the other hand, in the no-intention treatment, the median did not show any reciprocal pattern and was perfectly consistent with the self-interested prediction. The regression analysis, with player B's decision as the dependent variable, showed that the coefficient of the interaction term between *a* and the intention treatment dummy is significantly higher. Regarding player B's choice of *b*, there were no subjects choosing the self-interested response ($b = 0$) in the intention treatment, but 30% of the subjects chose it in the no-intention treatment. It seems that the lack of intention to be fair elicited self-interested responses. People take not only the consequences of an action but also the intentions signaled by that action into account.

◆ *Intention-based trust reciprocity hypothesis*

In an experiment that observed responses to the same intentions, McCabe et al. (2003) played a trading game using an intention-based model based on the trust reciprocity hypothesis. Player A has two choices: that the game ends with a gain of [20, 20] for players A and B, or that the choice to end the game is left to player B. In the latter choice, player B has two alternatives: [25, 25] and [15, 30]. That is, player A gives up [20, 20] by trusting player B, and expects player B to choose [25, 25] with more gains. Player B can either respond to player A's trust by choosing [25, 25] to benefit both players or betrays player A by choosing [15, 30] to maximize his/her own gain. In this experiment, subjects were divided into two conditions: a voluntary trust game (VTG) in which player A has a choice of [20, 20], and an involuntary trust game (ITG) in which player A does not have a choice of [20, 20] (i.e., player A has no opportunity to choose). The difference between these two conditions is that player B's inference of player A's intentions is regarded as an opportunity cost, and this opportunity cost is positive in the VTG and zero in the ITG. The positive opportunity cost in the VTG is assumed to generate greater reciprocity than that in the ITG.

As a prediction based on the intention-based model, because there are no intentions from player A in the ITG, player B will choose [15, 30]. However, if using the outcome-based model to explain player B's behavior, the prediction is that player B should make a decision based on his/her own gains or the relative ratio of his/her own gains to the gains received by player A, regardless of player A's intention.

The results of the experiment showed that out of 27 player As in the VTG, 10 chose [20, 20] and 17 left the choice to player B. Of these, 11 players chose [25, 25], and six chose [15, 30]. In contrast, in the ITG, out of 27 players, nine chose [25, 25] and 18 chose [15, 30].

This result somewhat contradicts the outcome-based prediction and is consistent with the intention-based prediction because the majority (64.7%) of player Bs in the VTG responded to player As's intentions and the majority (66.7%) of player Bs in the ITG chose to maximize their own gains.

◆ *Generating indirect reciprocity through reputation*

Wedekind and Milinski (2000) conducted an indirect reciprocity game experiment in which group members donated to other group members and received donations from other members. This experiment observed that reputation affects cooperative behavior through indirect reciprocity. The experiment was designed to increase a subject's image score by one point when he/she donated and decreased a subject's image score by one point when he/she did not. The subjects were divided into groups and paired with different subjects in each round. In addition, the history of whether a subject donated or not was displayed, and anonymity was maintained.

The results of the experiment showed that the history of donation had an important influence on donors' decisions. On average, the image score of those who received donations was significantly higher than that of those who did not. The results showed that the image score was used as an index of indirect reciprocity.

◆ *The effect of indirect reciprocity through reputation*

Milinski et al. (2002a) conducted an experiment in which a public goods game and an indirect reciprocity game were alternated. They observed that it is necessary to maintain a reputation for donating in order to elicit cooperative behavior through indirect reciprocity. Similar to Wedekind and Milinski (2000), they used image scoring to allow people to develop a reputation and displayed their donation history. Subjects were anonymously divided into groups, and the process of the experiment in each group was different. In the first group, the subjects played the indirect reciprocity and public goods games alternately for 16 rounds, after which they played four rounds of the public goods game. In the second group, the subjects were informed that they would play the indirect reciprocity and public goods game alternately for 16 rounds, then they would only play the public goods game from the 17th round until the end of the experiment. The difference between the first and second groups is whether the subjects knew in advance that only the public goods game would continue from round 17. In addition, the third group played eight rounds of public goods games, followed by eight rounds of indirect reciprocity games, then four rounds of public goods games.

The prediction is that those who raise their image score in the public goods game will be rewarded in the indirect reciprocity game, and thus continue to maintain high levels of public goods contribution.

The results of the experiment showed that in the third group, cooperation was quickly rebuilt during the eight rounds of the subsequent indirect reciprocity game, compared to the continuous

decrease in cooperation during the first eight rounds of the public goods game. In the first and second groups, in which the public goods and indirect reciprocity games alternated and lasted up to 16 rounds, cooperation did not decrease during the eight rounds of the public goods game, and the average cooperation rate was significantly higher than that of the indirect reciprocity game.

A possible reason for the higher cooperation rate in the public goods game, in the first and second groups, might be that in the indirect reciprocity game, subjects did not reward those who did not contribute in the round of the public goods game immediately before. In fact, the probability of not receiving a reward in the indirect reciprocity game was significantly higher for subjects who did not contribute than for those who did in the public goods game round immediately before.

The prediction that receiving rewards in the indirect reciprocity game could maintain cooperative behavior in the public goods games was tested using the data of the public goods games played across 17–20 rounds in the first and second groups. In the second group, cooperation decreased during the four rounds of the public goods game because the subjects were told that only the public goods game rounds would continue from the 17th round until the experiment ended. In the first group, however, cooperation was maintained because the subjects were not told that only the public goods game would continue; thus, there was a possibility that the indirect reciprocity game round would still be played, as in the previous 16 rounds. The possibility of not knowing whether the indirect reciprocity game would continue had the effect of preventing the decline of cooperation in the public goods game. This result suggests that to maintain cooperative behavior, people need to understand the possibility of receiving rewards by improving their reputation.

◆ *The impact of indirect reciprocity through reputation on voting*

Milinski et al. (2002b) conducted an experiment to observe how reciprocity influences voting behavior by playing an indirect reciprocity game and asking people whether they would donate to a real aid agency (UNICEF). At the end of the game, subjects were given the right to vote and choose another suitable subject to represent them on the student council.

If the act of donating serves as a signal of a person's social trust, then the person who donates the most should get the most votes.

The results of the experiment showed that there was no significant correlation between the amount donated in the indirect reciprocity game and the amount donated to UNICEF. Those who donated more to UNICEF received relatively more from others, while those who donated less received relatively less. Those who did not donate to UNICEF received less from others in the indirect reciprocity game than those who did. The sum of donations to UNICEF and to others in the indirect reciprocity game correlated positively with the number of votes for suitable representation on the student council. There was no significant correlation between the total amount donated to others in the

indirect reciprocity game and the number of votes. This indicates that donations to aid agencies have a stronger impact on political reputation than donations to other members of the group.

From the results, it was found that donating to UNICEF, an aid organization that operates under a socially recognized philosophy of altruism, is still considered socially significant and relevant to being elected as a suitable representative of a social organization (i.e., the student council).

## 3. What is the social dilemma between generations?

### 3.1 Real world problems

In the experimental studies on social dilemmas we have seen thus far, participants played the game in the same time frame and made decisions on whether to cooperate or not. In these studies, social dilemmas were resolved by rational, altruistic, and reciprocal cooperative behavior through the means of punishment and reputation. However, social dilemmas often span more than a single generation, and their effects can be far-reaching. For example, in the case of climate change, the current generation consumes fossil fuels, but the effects of global warming will become apparent to future generations. Moreover, the relationship between benefits and burdens in public finance is such that the current generation benefits from fiscal spending and tax burden avoidance, while future generations bear the burden of reduced spending and taxes due to the resulting fiscal imbalance. The fact that benefits and burdens differ from generation to generation has become a real problem. Thus, in the case of intergenerational social dilemmas, will the means of punishment and reputation be effective?

### 3.2 Structure of the problem

When social dilemmas cross generations, future generations cannot punish the current one. In addition, there is no opportunity for reciprocal interaction between present and future generations. While the current generation may act altruistically toward future generations, Saijo (2020) states that human myopia and optimism “sparingly deprive” future generations by their engagement in markets and democracy. Myopia here means that we instinctively do not care about the distant future. Optimism means that what happens in the future will not be that bad, and there might exist a so-called optimism bias toward the difference between expected results and actual benefits. The market, on the other hand, is an extremely good device for understanding people’s short-term aspirations but does not allocate resources with consideration for future generations since they cannot participate. Democracy is also a device that benefits people living in the present and does not work for future generations. Even if a candidate for election presents policies that will enrich generations 100 years from now, the present generation will not vote for him/her. That is, in markets and democracies, myopia and optimism reinforce the tendency not to consider the interests of future generations.

### 3.3 An experiment dealing with the intergenerational social dilemma

Sherstyuk et al. (2016) showed that short-sighted behavior increases gains through selfish choices, but it is not best in the long run. Their experiment compared decision-makers who survive infinitely with those who alternate between more realistic generations, observing their gains and behaviors. The gains model emulates a long-term environmental problem, where generation $t$ chooses a certain amount of emission $x$, which affects the accumulation of pollution $S$. In other words, if the amount of emission $x$ is reduced, the amount of increase in pollution accumulation $S$ will also be reduced, and the gains will increase in the next $t + 1$ generation. The subjects were divided into three groups: a long-lived (LL) group, in which the subjects are the same and live longer across all generations; an intergenerational selfish (IS) group, in which the subjects are different and selfish across generations; and an intergenerational long-sighted (IL) group, similar to LL in that the subjects are the same and live longer across all generations; however, they have a long-term perspective. Each generation can leave advice to the next generation on how to behave.

Predictions suggest that subjects' behavior may be myopic about the wealth of future generations. In addition, increasing intergenerational uncertainty may lead to contradictions about one's own behavior and trust in others.

The results of the experiment showed that both IL subjects who live longer and have a long-term perspective and IS subjects who are selfish gain more than LL subjects. However, the IS subjects did not obtain more gains than the IL subjects. The proportion of subjects selecting the best choice for themselves was 32% for IS, 16% for LL, and 24% for IL. Nine percent of IS subjects, 1% of LL subjects, and 2% of IL subjects made choices based on their short-term interests, while 7% of IS subjects, 21% of LL subjects, and 31% of IL subjects made choices based on their long-term interests. If these three groups are replaced by the survival of the government, in the LL group, there exists the possibility of realizing a cooperative gain because the government survives for a long period of time, communication might take place between each other, and optimal environmental policies might be implemented. In the IS group, non-cooperative behavior evolves over generations because if the subjects choose non-cooperative emissions, their advice to the next generation will be the same. Thus, when governments are replaced each generation, it is clear that they are not motivated by future wealth. In the IL group, emissions are coordinated to converge among several generations. If governments are replaced in the short term but have a long-term view, coordination may be possible between governments and between generations.

These results imply that it is best to have the same decision maker living for a long period of time, considering the gains into the future. However, this also means that even the decision makers change from generation to generation, which is acceptable if they are not selfish and do not change the decision to consider those gains across a long period of time into the future by sending their advice.

## 4. What is Future Design?

### 4.1 The purpose

Traditional social science assumes that the nature in which we think and act cannot be easily changed. However, Saijo (2020) argues that institutions, such as markets and democracies, shape the nature of how we think and act, which can also be changed through feedback from those institutions. Therefore, we must change our own thinking, use knowledge from different fields, and design social mechanisms to test their effectiveness. This is the basic framework of *future design*. In other words, the purpose of *future design* is to explore what kind of initiatives would be more effective in changing the thinking and behavior of the current generation to build a sustainable society.

### 4.2 Methods

In *future design*, researchers have conducted a series of experiments to clarify what influences the preferences of individuals and groups. In these experiments, based on group deliberation on resource allocation, the following four specific mechanisms and processes were considered as those that could affect the thoughts and actions of participants in decision-making.

The first is the role of a virtual (imaginary) future generation. Participants were asked to play the role of an imaginary future generation. They were then asked to participate in a group discussion on resource allocation to observe how face-to-face discussions with other participants (the current generation) would affect their decision-making. The second was the introduction of retrospection, in which past policies and their consequences were learned as lessons and reflected in the current decision-making process. Whether the presence of this retrospective process affected group discussions and decision-making regarding resource allocation was observed. The third was the process of accountability of the current generation, communicating the reasons for their decisions to subsequent future generations. How the decision-making of subsequent generations was affected by whether the current generation sent explanations for their decisions and advice to future generations was observed. The fourth was the process of granting the current generation the right to vote for future generations. How the presence of multiple voting rights with additional voting rights would affect other single voting rights holders was observed. Table 2 provides an overview of the above-mentioned methods.

The intergenerational sustainability dilemma game is also fundamental to the above set of experiments. In this game, each generation has two options to choose from: option A, which greatly benefits the current generation and exploits future generations; or option B, which does not benefit the current generation as much but is sustainable by carrying the benefits over to future generations. Suppose that the first generation is given a gain of 3,600 points as an initial endowment. If the subjects choose option A, the first generation gets 3,600, but the following generation gets 2,700 with a fixed

amount of 900 subtracted. If the subjects continue to choose option A, the gains will continue to decrease with each generation and eventually reach zero, which is unsustainable. On the other hand, if the subjects choose option B, the first generation gets 2,700 after a fixed amount of 900 is deducted, and the following generation gets 3,600 after the deducted 900 is added. If the subjects continue to choose option B, the gains are constant and intergenerationally persistent.

In the next subsections, we review the experimental results of each method.

Table 2. The overview of experimental methods in *future design*

| | Past generation | Present generation | Future generation |
|---|---|---|---|
| Imaginary Future Generation (IFG) | | with IFG | Playing the role of the future generation |
| Retrospective | Reading articles about previously developed policies | | Choosing the policies that should be taken now from a position of the future generation |
| Accountability | | | Providing an explanation for the outcome of the decision to the future generation |
| Proxy voting | | | Voting on behalf of the future generation |

### 4.3 The imaginary future generation

◆ *A basic experiment*

Kamijo et al. (2017) conducted an experiment that presented a basic type of imaginary future generation. They asked students in Kochi, Japan to participate in a comparative analysis of whether the presence of an imaginary future generation would make them choose policies that favored the current generation or ones that are sustainable in the future. Some of the participants were randomly assigned the role of the imaginary future generation that would act on behalf of future generations. This experiment had two conditions: a treatment condition, in which an imaginary future generation participated, and a control condition in which there was no imaginary future generation. The intergenerational sustainability dilemma game mentioned above was used for both conditions.

The results of the experiment showed that 60% of the groups in the treatment condition chose option B, compared to 28% of the groups in the control condition. At the individual level, in more than

half of all generations in the treatment group, at least two of the three members of every generation chose option B, and 72% of the subjects in the control condition unanimously chose option A. In terms of the content of the group discussions, subjects who played the role of an imaginary future generation were more likely to prefer option B and were more proactive in making sustainable choices. In terms of duration, the average time spent in discussion in the treatment condition was 351.23 seconds, while in the control condition, the average time spent was 210.80 seconds. This difference indicates that in the treatment condition group members continued the discussion even when they could not reach an agreement.

These results indicate that the presence of an imaginary future generation promotes sustainable choices in both the content of the statements made during the discussions and the actual choices made.

◆ *The impact of capitalization in an urban area*

Shahrier et al. (2017) conducted an intergenerational sustainability dilemma game experiment to observe the effect of the presence of an imaginary future generation on decision-making in two areas of a developing country, Bangladesh. The first was in Dhaka, which is an urban area with an increasingly capitalistic economy, and the other was Bogra, which is a rural area with a rural economy.

The experiment showed that people in the rural area chose option B more than those in the urban area. However, the probability of choosing option B was higher in the rural area, even in the absence of an imaginary future generation. In contrast, in the urban area, there was no effect from the presence of an imaginary future generation on the choice of subjects in the treatment condition (see the previous item). Sixty-three percent of the urban groups had no members with prosocial tendencies and consisted only of individualistic and competitive members. The probit regression analysis, investigating the effects of choosing option B, suggested that the number of prosocial members in the group and the area had significant effects, while socioeconomic factors (e.g., income, education, number of women in the group, occupation, and age) had no effect.

These results may indicate that imaginary future generations have only a limited effect in rural areas. In addition, intergenerational sustainability does not seem to improve in urban areas with capital accumulation. As an interpretation, Shahrier et al. focused on daily learning that propagates from generation to generation. That is, since it is an agricultural society with self-consumption, in the rural area, non-cognitive skills (customs and habits) and agricultural techniques are vertically transmitted from generation to generation. In contrast, in the urban area, cognitive and non-cognitive skills are transmitted horizontally within generations through a formalized education system. The results also suggest that differences in the direction of transmission and the strength of prosocial social preferences influence sustainable choices for the future. Kamijo et al. (2017) observed that the presence of an imaginary future generation has a significant impact, but in developing countries the effect of its presence seems to vary depending on the degree of capitalization.

◆ *The presence of influencers in an urban area*

Timilsina et al. (2018) conducted an intergenerational sustainability dilemma game experiment in urban Kathmandu and rural Chitwan, Nepal. In addition to the decision-making consequences from the presence of an imaginary future generation, they analyzed conversations in group discussions to observe tendencies in the subjects' speech and its impact within the group. The conversations were analyzed quantitatively, and the subjects were categorized into four states with respect to the options for supporting their statements: not expressing an opinion, stating support for option A, stating support for option B, and an ambiguous position. Based on the shifts in these states during the discussion, the subjects' characteristics were further classified into three types: a dependent type, who always supported option A or B without a specific reason; a stable type, who supported option A or B for some reason; and an unstable type, who changed the option they supported. The statements were divided into 15 categories according to whether they considered future generations.

The results of the experiment showed that in the rural area, 16% of the groups chose option A, while 84% chose option B. In contrast, in the urban area, 36% of the groups chose option A, while 64% chose option B. In terms of subjects' characteristics, 14% of the stable type and 15% of the dependent-type subjects favored option A in the urban area. In the rural area, the percentages were 6% and 3%, respectively. In terms of discussion content that took future generations into consideration, there was no difference between urban and rural areas; however, statements that did not consider future generations were more common in the urban area. In addition, some statements not seen in the rural area were seen in the urban area (e.g., "surprise at earlier generations' decisions," in the sense that they were able to get benefits from the beginning, "senses of guilt relaxed by earlier generations' decisions," in the sense that the previous group also chose A, and "non-negligible cost of considering future generations," in the sense that they paid the cost by choosing B).

These results indicate that option A was chosen more often in urban areas, while option B was selected more often in rural areas. This is consistent with the experimental results of Shahrier et al. (2017). The characteristics of the subjects in the group discussion showed that in the urban area, the stable types who supported option A acted as influencers (the "key generation"), who encouraged the dependent and unstable types to also support option A. This means that a significant percentage of urban subjects were of the stable type, who consistently chose option A and influenced others to do the same.

◆ *Realistic fiscal policy choices*

In a study by Hiromitsu (2017), participants were asked why they chose between two scenarios that were close to actual fiscal policy in Japan in a questionnaire administered both before and after the group discussion. In the two scenarios, the consumption tax, income tax, insurance premiums,

medical expenses, pensions, and specific items and amounts/rates borne by individuals were listed as examples of how they would change between now and 30 years from now. This generated two options: fiscal policy 1, which reduces the burden on the current generation, and fiscal policy 2, in which the current and future generations bear the same burden. In this experiment, participants were asked to play the role of an imaginary future generation, and group discussions were held.

The results of the experiment showed that, among all the groups, support for fiscal policy 1 was 20.9%, and support for fiscal policy 2 was 79.1%. Among the groups that only consisted of the current generation, 27.7% supported fiscal policy 1 and 72.3% supported fiscal policy 2. On the other hand, among groups that included an imaginary future generation, 12.3% supported fiscal policy 1 and 87.7% supported fiscal policy 2. From the questionnaire, the most common reasons for changing opinions in the groups that initially favored fiscal policy 2 but eventually changed to fiscal policy 1 were, "Depending on future trends in population, economy, science and technology, the future will be better than what we currently expect" and "It is not productive to worry about future generations because we do not really know what they will value and want." On the other hand, the most common reasons for changing options in the groups that initially favored fiscal policy 1 but eventually changed to fiscal policy 2 were, "We should consider the interests of society as a whole, including the interests of future generations, not just our own" and "Considering my own life, a policy that would significantly increase the burden in 30 years is a big blow to me."

These results indicate that a majority of the respondents supported fiscal policy 2, but they had a strong tendency to make conscious public decisions, as 64.8% of the respondents stated that "the interests of ourselves (the current generation) and the interests of future generations should be considered fairly." On the other hand, the behavior of supporting fiscal policy 1 tended to be strongly self-interested, as 20.9% of the respondents stated that "the interests of ourselves (the current generation) should be considered first." In this experiment, Hiromitsu (2017) recognized that future uncertainty and agnosticism were cited as reasons for supporting fiscal policy 1 and suggested that knowledge for a correct view of the economy is necessary. With regard to optimism bias, which is human nature, the questionnaire confirmed that 57.1% of respondents stated, "since we are uncertain about what the future in 2046[2] will be like, we should focus more on the present benefits."

### 4.4 Retrospective

Nakagawa et al. (2019) asked participants to read an article about a previously developed nuclear policy (i.e., the Monju prototype fast breeder reactor) and how it came about as a retrospective process, and observed how this experience influenced decisions through a group discussion about forest policy in Kochi Prefecture, Japan. The participants were divided into a treatment group that experienced the retrospective process and a control group that did not undergo this process. In the experiment, the

[2] This experiment took place in 2015 and 2016, which means 30 years from then.

participants in the treatment group read the case (about Monju), had a group discussion, and made a group policy choice. In addition, two individual choices were made by the participants in both treatments: a policy choice before the group discussion and another after it. There were five policy options, with options 1-3 maintaining the current inefficient forest management, and options 4 and 5 improving the status quo.

The results of the experiment showed that in the control group, a significant percentage of the participants chose options 3 and 4 as their most preferred policies, at 47.5% initially, rising to 65.0% after the group discussion. Conversely, participants in the treatment group chose option 4 as the most preferred policy over option 3. The percentage was 39.0% initially, eventually reaching 39.1%, remaining virtually unchanged.

The logit regression analysis with the choice to improve from the status quo as the dependent variable showed that the variable for the treatment group that experienced the retrospective process was significant. The other explanatory variables (e.g., age, gender, regular/non-regular employment, education, etc.) were not significant.

These results suggest that the acquisition and experience of an intergenerational retrospective perspective influences individual preferences for forest policies, in the sense that the policies chosen as most favorable in the treatment group were different from those in the control group. Participants in the treatment group tended to choose policies that essentially improved the status quo for the future, while the reverse was observed in the control group.

### 4.5 Accountability

Timilsina et al. (2019) observed how accountability influences decision-making in the next generation of urban Kathmandu, Nepal. The participants were divided into three treatments: intergenerational accountability (IA), in which subjects choose options A or B[3] and are accountable for their choices by passing on their reasons and advice to the next generation; an imaginary future generation (IFG), which includes the imaginary future generation; and a baseline treatment, which includes neither accountability nor IFG. An intergenerational sustainability dilemma game was also used in this experiment.

The results showed that the subjects in the IA treatment were more likely to choose option B than those in the IFG and baseline treatments. IA and IFG were more effective in encouraging subsequent generations to choose option B. In each treatment, the higher the percentage of choosing option B, the greater the number of prosocial members. In the baseline treatment, if the majority of members were

[3] Recall that option A greatly benefits the current generation and exploits future generations, and Option B does not benefit the current generation as much and is sustainable through carrying the benefits over to future generations.

selfish, they would always choose option A. However, in the IA and IFG treatments, they would choose option B, even in the same situation. In the logistic regression analysis, the number of prosocial members and the IA dummy were determinants of maintaining intergenerational sustainability, and the IA dummy was particularly effective.

These results are consistent with the research finding that prosocial people play an important role when they cooperate to maintain common resources and public goods. Intergenerational accountability is the most effective factor in maintaining or extending sustainability across generations. It is also thought to function as a social device that raises empathy beyond selfishness, spanning generations, as a one-way communication channel from the current generation to the next in the intergenerational sustainability dilemma game.

### 4.6 Proxy voting

Kamijo et al. (2019) conducted an experiment to examine the effect of proxy voting rights on people who have one vote. One group consisted of three participants: two participants, A and B, representing the current generation, and one participant representing the future generation. The current generations choose either option A or B. The future generation only receives the gains and does not participate in voting. Participants were divided into a proxy voting (PV) treatment, in which the current generation Bs explained to As that they had the right to vote for themselves and the right to vote on behalf of future generations for option B; a non-proxy voting (NPV) treatment, in which participant Bs had the right to vote for both themselves and future generations; however, they did not explain this to participant As; and an ordinary voting (OV) treatment, in which participant Bs had the right to vote only for themselves.

In the experiment, 46% of the participants in the PV treatment chose option B, while 30% of those in the NPV treatment chose option B. In the NPV treatment, the current generation A chose option B slightly less often than those in the OV treatment. In other words, participants in the NPV treatment were the most reluctant to choose option B.

The results indicate that not explaining the purpose of the second vote may have caused negative (selfish) reactions among those who only had one vote. It is also possible that a person with only one vote could be interpreted as an over-amplification of the interests and opinions of those with proxy voting rights. To prevent any misunderstanding or distrust from current generation voters, the purpose of proxy voting on behalf of future generations should be clearly explained to them.

## 5. Methodological challenges – the imaginary future generations

### 5.1 As an effective method

Saijo (2015) stated that he forms a group of people who "become" future generations by using our innate mental functions that allow us to infer the state of mind of others. By providing a process for this group and the current generation to deliberate and make decisions on resource allocation, we are attempting to curb myopia and optimism – the structural problems of intergenerational social dilemmas – and introduce an intergenerational perspective to the market and democracy. As we saw in Section 4, there is a strong tendency for decisions to be made with consideration for future generations when imaginary future generations are present rather than the current generation alone. In addition to experiments, *future design* is also being practiced at the municipal level in places such as Yahaba Town, Iwate Prefecture, and Suita City, Osaka Prefecture, all of which are in Japan, with the participation of local residents. Saijo (2017) observed that compared to the current generation, which tends to be obsessed with the issues at hand, when given the role of an imaginary future generation, the participants approached judgments and decisions with the future in mind, viewing and ranking many issues and measures from a bird's eye view. Therefore, the introduction of an imaginary future generation is a powerful and influential way to influence decisions, not only in experiments but also in policy formation practices.

### 5.2 Basis for its existence

What is the basis for the introduction of imaginary future generations? A series of studies in *future design* are as yet unclear on this point. The expectation is that through the democratic process of deliberation, an imaginary future generation will choose policies that are desirable for future generations, assuming that an imaginary future generation already exists and has been accepted. However, does this premise of being based on the common understanding of all participants seem a bit too strong of an intervention? Is such an intervention inconsistent with the process of democratic deliberation? The imaginary future generation seems only given from the outside, within the framework of the experiment or field study but is not adopted in the participants' customs, social norms, or actual institutions.

Saijo (2015) stated that an imaginary future generation is just one example, and there should be various processes, which will be an important issue to consider in the future. In this paper, we will not discuss the methods and effects of the imaginary future generation, but rather what the rationale for its existence is and how it should be embraced.

### 5.3 Into the inner free will of people

We propose that the institution of imaginary future generations could be left to the free will of the participants to make their own choices, rather than from the outside. That is, everyone in each generation decides whether he/she should accept the role of the imaginary future generation by himself/herself. If there exists an idea based on an intergenerationally common understanding within

people, could an imaginary future generation be it? If so, will future generations, perceived as a common understanding, have the same effect on influencing decision-making as they did originally?

## 6. What are normative expectations?

### 6.1 Overview

Social norms are often based on shared beliefs or mutual expectations among players. Bicchieri (2006) argues that they are the "rules of behavior" and functions that map one's own expectations about the behavior of others onto what should be done.

Xiao and Bicchieri (2010) conducted a trust game experiment to observe the behavior of equality and reciprocity among players. The players (and trustees) were divided into two groups: a control group starting with the same initial holdings and a treatment group in which the trustors had initial holdings twice that of the trustees. Suppose you are the trustee, when you are thinking about how much you will pay back to the trustor, you consider it as follows: "how much do I think my counterpart believes I should transfer to her?" This is not about how much you should pay back to the trustor, but how the trustor feels about it. You need to infer not what you think, but the trustor's "how he/she thinks" behavior. Xiao and Bicchieri (2010) examined in their experiment whether equality and reciprocity are recognized and shared. They found that equality and reciprocity could not be established unless he/she and the paired subject shared the expectation that they would do this in this case. This shared expectation is called a normative expectation, and behaviors based on normative expectations are called "rules of behavior."

### 6.2 Observation of normative expectations

How can we observe this normative expectation and the rule of behavior? Schram and Charness (2015) conducted an experiment in which they formed a group of three people in a dictator game, with one being the dictator. They observed how much money was allocated from the dictator to the other two recipients. There was an advice group that only gave advice to the dictator and his/her entire group, and the dictator was divided into two conditions: one in which he/she received advice and one that did not. In each condition, there were two groups: a treatment group in which the outcome of the dictator's decision was made public and a control group in which the outcome was not made public.

The results of the experiment show that the advice creates a shared understanding of the dictator's decision-making and generates conventions. When the decision-making results are made public, the convention becomes a social norm.

In the condition where the dictator was advised, the dictator and his/her entire group received advice from the advice group. Here, "what the dictator should do" is transmitted. Of course, the

dictator can choose not to follow this advice, but this will be a shared understanding. When the dictator's decisions were made public, they became social norms for the next dictator. The advice group also knew the decisions made by the dictator and would advise the next dictator on whether to encourage or reverse them as norms. By repeating this process, shared understanding and social norms are solidified in a certain direction. In contrast, under the conditions where the dictator is not advised, "what the dictator should do" is not transmitted, and no shared understanding is created. Even if the dictators' decisions were made public, they would not serve as norms.

The advice given by the third-party advice group indicates what the dictator should do and generates the normative expectations of the dictator. In addition, when decisions are made public, dictators tend to follow the advice. In contrast, if the outcome of the decision-making process is not made public, the advice is ineffective. Therefore, the combination of taking advice and disclosing the decision-making results is effective. This is because if the person giving the advice knows that the results will be made public, the advice will be altruistic, and the dictator will follow it.

### 6.3 Application to an imaginary future generation

Using Schramm and Charness's (2015) experimental design, we might be able to observe whether normative expectations and rules of behavior that imaginary future generations should adopt emerge. If people accept the imaginary future generation, and if it is repeated, then we will see that this influential idea is a decision that comes from within people.

## 7. Further issues and research methods

### 7.1 Setting up the problem issues

As demonstrated in Section 4, we will not ask again about the effects of the imaginary future generation itself, based on the assumption that it can have a significant effect on decision-making in group discussions. In light of the review described thus far, we would like to address the following two issues as future tasks:

The first is to observe the imaginary future generation as a normative expectation. Can we choose to be an imaginary future generation as freely as possible? If this is indeed adopted, will it be repeated in the next generation rather than in one generation only? If it is not adopted, will it also continue? Alternatively, will it be suddenly adopted? The creation, persistence, and disappearance of certain conventions can also be observed. Schotter and Sopher (2003) repeated a battle of the sexes game over several generations and found that conventions were characterized by "punctuated equilibria," where one convention is replaced by another, "socialization," which is maintained through generations, and "social inertia," where the next generation follows blindly. If such a convention exists, it can be viewed as a normative expectation and a rule of behavior that the imaginary future generation should be

adopted through the shared understanding mentioned in Schram and Charness (2015). The second issue is identifying the consequences of methodological differences. That is, we plan to investigate how the difference between given and selectable imaginary future generations appears in differences between gains in true future generations.

### 7.2 The methods

An experiment to answer the above questions could involve two stages of decision making: in the first stage, a group of subjects decides whether to accept the imaginary future generation; in the second stage, the same group of subjects decides how to allocate resources between the current and future generations using an intergenerational sustainability dilemma game. Another group of subjects receives the gains resulting from resource allocation, and the two-stage decision-making process is repeated.

The second stage is based on the experimental method of Schram and Charness (2015), who observed that the disclosure of advice and payoff promotes a shared understanding among groups. That is, after knowing the advice on whether to adopt the imaginary future generation and the previous generation's reactions toward that advice, the next generation decides whether to adopt the imaginary future generation again and/or whether to follow the decision of the previous generation. This decision becomes a shared understanding among groups, which is then passed on to the next generation. In addition, in the experiment of Schram and Charness (2015), a dictatorship game was used; however, since the decision-making process of *future design* is based on the deliberative process, a group discussion will be considered.

### 7.3 The hypothesis

The assumption that an imaginary future generation has a significant effect on decision-making in group discussions leads to the hypothesis that the imaginary future generation will continue to be accepted throughout each generation. However, this hypothesis does not explain why it is accepted throughout each generation. We believe this is because each generation decides by free will to accept the imaginary future generation and allocate resources among generations to avoid sacrificing the gains of the true future generation. If this is supported, then the acceptance of the imaginary future generation becomes a shared understanding and has a significant effect on the decision to choose it as a normative expectation for the next generation.

Another hypothesis is that the gains of the true future generation under the decision to adopt the imaginary future generation is greater than the given one. This is because free will acceptance is the same as understanding the purpose of the imaginary future generation, and it is believed that they will make decisions that serve their purpose rather than relative to the imaginary future generation as a

given. If this hypothesis is supported, then accepting the imaginary future generation by free will would prove to be the most optimal for future generations.

## 8. Conclusion

We reviewed the experiments related to a social dilemma in which people's decisions are interpreted as the result of rationality, or emotion and altruism. We also identified the problematic structure of the intergenerational social dilemma, in which people with a myopic and optimistic nature are involved in markets and democracy. An intergenerational social dilemma impedes the allocation of resources from the perspective of the future on the time axis. To address this problem, the *future design* method presents ideas such as an imaginary future generation, accountability, retrospective, and proxy voting for future generations, which were found to have a significant impact on people's decision-making. However, to set an institutional premise, such as an imaginary future generation, as self-evident, we should leave it to the people's inner free will by ensuring that it is at least a choice situation.

In this paper, we propose an experimental design based on the hypothesis that the acceptance of the imaginary future generation will become a shared understanding across generations through people's inner free will, and that the acceptance of the imaginary future generation will become a normative expectation for the next one. In the future, we would like to refine and test our hypotheses, check the validity of the experimental procedures, and examine the factors that should be included in the quantitative evaluation.

## Acknowledgements

The authors would like to thank Tatsuyoshi Saijo and Ken-Ichi Shimomura for the extremely helpful comments and suggestions. Financial support from the Japanese Ministry of Education, Culture, Sports, Science and Technology through a Grant-in-aid for Scientific Research (C) 19K01541 and 19K01558 is gratefully acknowledged. All the views expressed in this paper and any errors are the sole responsibility of the authors.

## References

Bicchieri, C. (2006). The Grammar of Society: The Nature and Dynamics of Social Norms, Cambridge University Press

Bicchieri, C. (2011). Social Norms in Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/social-norms (referred to 12/18/2020)

Dawes, R. M. (1980). Social Dilemmas. *Annual Review of Psychology*, 31, 169-193

Falk, A., Fehr, E. and Fischbacher, U. (2000). Testing Theories of Fairness: Intentions Matter. Working Paper No. 63, Institute for Empirical Research in Economics, University of Zürich

Fehr, E. and Fischbacher, U. (2004). Third-Party Punishment and Social Norms. *Evolution and Human Behavior*, 25, 63-87

Fehr, E. and Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90, 980-994

Fehr, E. and Gächter, S. (2002). Altruistic Punishment in Humans. *Nature*, 415, 137-140

Fehr, E. and Rockenbach, B. (2003). Detrimental Effects of Sanctions on Human Altruism. *Nature*, 422, 137-140

Fischbacher, U., Gächter, S. and Fehr, E. (2001). Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economics Letters*, 71, 397-404

Gneezy, U. and Rustichini, A. (2000). A Fine Is a Price. *Journal of Legal Studies*, 29, 1-17

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162, 1243-1248

Hiromitsu, T. (2017), Consideration of Clues to the Solution of Long-Term Financial Problems by Means of Experiments. *PRI Discussion Paper Series*-17A-08 (in Japanese)

Kamijo, Y., Komiya, A., Mifune, N. and Saijo, T. (2017). Negotiating with the Future: Incorporating Imaginary Future Generations into Negotiations. *Sustainability Science*, 12, 409-420

Kamijo, Y., Hizen, Y., Saijo, T. and Tamura, T. (2019). Voting on behalf of a Future Generation: A Laboratory Experiment. *Sustainability*, 11(16), 4271

Mankiw, N. G. (2017), Principles of Microeconomics, 8th ed, Cengage Learning

Masuda, T., Okano, Y. and Saijo, T. (2014). The Minimum Approval Mechanism Implements the Efficient Public Good Allocation Theoretically and Experimentally. *Games and Economic Behavior*, 83, 73-85

McCabe, K. A., Rigdon, M. L. and Smith, V. L. (2003). Positive Reciprocity and Intentions in Trust Games. *Journal of Economic Behavior and Organization*, 52, 267-275

Milinski, M., Semmann, D. and Krambeck, H. J. (2002a). Reputation Helps Solve the "Tragedy of the Commons". *Nature*, 415, 424-426

Milinski, M., Semmann, D. and Krambeck, H. J. (2002b). Donors to Charity Gain in Both Indirect Reciprocity and Political Reputation. *Proceedings. Biological Sciences*, 269, 881-883

Nakagawa, Y., Kotani, K., Matsumoto, M. and Saijo, T. (2019). Intergenerational Retrospective Viewpoints and Individual Preferences of Policies for Future: A Deliberative Experiment for Forest Management. *Futures*, 105, 40-53

Saijo, T. (2015). Future Design. In *Future Design: Society Looking Seven Generations Ahead*, edited by Saijo, T., Keiso-shobo, 1-26 (in Japanese)

Saijo, T. (2017). Future Design. *Economic Review, Hitotsubashi University*, vol. 68(1), 33-45. doi.org/10.15057/28328

Saijo, T., Masuda, T. and Yamakawa, T. (2018). Approval Mechanism to Solve Prisoner's Dilemma: Comparison with Varian's Compensation Mechanism. *Social Choice and Welfare*, 51, 65-77

Saijo, T. (2020), Future Design: Bequeathing Sustainable Natural Environments and Sustainable Societies to Future Generations. *Sustainability*, 12(16), 6467

Schotter, A. and Sopher, B. (2003). Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study. *Journal of Political Economy*, 111, 498-529

Schram, A. and Charness, G. (2015). Inducing Social Norms in Laboratory Allocation Choices. *Management Science*, 61(7), 1531-1546

Shahrier, S., Kotani, K. and Saijo, T. (2017). Intergenerational Sustainability and the Degree of Capitalism in the Society: A Field Experiment. *Sustainability Science*, 12(6), 957-967

Sherstyuk, K., Tarui, N., Ravago, M. V. and Saijo, T. (2016). Intergenerational Games with Dynamic Externalities and Climate Change Experiments. *Journal of the Association of Environmental and Resource Economists*, 3(2), 247-281

Timilsina, R. R., Kotani, K., Nakagawa, Y. and Saijo, T. (2018), Concerns for Future Generations in Societies: A Deliberative Analysis on Intergenerational Sustainability Dilemma. *Social Design Engineering Series*, Kochi University of Technology

Timilsina, R. R., Kotani, K., Nakagawa, Y. and Saijo, T. (2019), Accountability as a Resolution for Intergenerational Sustainability Dilemma. *Social Design Engineering Series*, Kochi University of Technology

Umino, M. (1991). The Scope of Social Dilemma Research. In Order Problems and Social Dilemmas, edited by Moriyama, K. and Umino, M., Harvest-sha, 146-153 (in Japanese)

Varian, H. R. (1994). A Solution to the Problem of Externalities when Agents Are Well-informed. *American Economic Review*, 84(5), 1278-1293

Wedekind, C. and Milinski, M. (2000). Cooperation through Image Scoring in Humans. *Science*, 288, 850-852

Xiao, E. and Bicchieri, C. (2010). When Equality Trumps Reciprocity. *Journal of Economic Psychology*, 31, 456-470