Chapter 85

SPITEFUL BEHAVIOR IN VOLUNTARY CONTRIBUTION MECHANISM EXPERIMENTS

TATSUYOSHI SAIJO

Institute of Social and Economic Research, Osaka University, and Research Institute of Economy, Trade and Industry, Tokyo

One of the basic findings in public good provision experiments via the voluntary contribution mechanism is that subjects contribute a considerable amount of their initial holdings to the provision of a public good even when no contribution is the dominant strategy [see Ledyard's survey article (1995)]. This seemingly non-rational behavior has been interpreted as resulting from confusion, kindness or fairness. For example, Andreoni (1995) developed a method to distinguish between confusion and kindness, and concluded that a relatively large part of positive contributions is interpreted as kindness.

On the other hand, there are three fundamental criticisms of voluntary contribution mechanism experiments. First, most experiments to date have used a linear utility function where the marginal return of a public good out of one unit of contribution of a private good is less than one. Therefore, no contribution is the best strategy. Since the solution is a corner solution, the experimental results were biased in favor of contribution. In order to answer this criticism, Saijo and Nakamura (1995) conducted experiments where the marginal return of a public good is greater than one (high marginal return). That is, all contribution is the dominant strategy. They compared the results to experiments where the marginal benefit of a public good is less than one (low marginal return) and found that the deviation from the dominant strategy in the case of high marginal return is larger than that in the case of low marginal return. They also found subjects who did not contribute cared about their relative ranking as well as monetary reward. In the case of high marginal return, if they cared about relative ranking among subjects, they would contribute nothing out of their initial holdings, but if they cared about the monetary reward, they would contribute everything. They termed this ambivalence the *spite* dilemma.

The second criticism is on the linearity of a utility function. In usual theoretical analyses, economists use a non-linear utility function such as the Cobb–Douglas utility function, Saijo, Yamato, and Yokotani (2004) conducted two-subject voluntary contribution mechanism experiments using the Cobb–Douglas utility function. They found that most subjects employed an interior Nash equilibrium strategy, but not quite: assuming that the other subject would choose his Nash contribution, some subjects chose slightly lower contributions than the Nash contribution.[1] The reason for choosing slightly lower con-

[1] This observation is compatible with the data in Andreoni (1993), who studied the crowding out hypothesis in public good experiments with non-linear payoff functions, though he did not refer to spiteful behavior.

Handbook of Experimental Economics Results, Volume 1

DOI: 10.1016/S1574-0722(07)00085-6

tributions than the Nash contribution was because it did not change their own payoffs (since the first order condition of a utility maximization problem is satisfied), however, the other subject's payoff would be lowered considerably.

The third criticism comes from mechanism design theory (Saijo and Yamato, 1999). Mechanism designers implicitly assumed that agents in their mechanisms must participate in the mechanisms. As Olson (1965) noticed, any non-participant can obtain benefit of a public good that is provided by others. This is due to the nature of a public good called *non-excludability*. In other words, almost all mechanism designers after Groves and Ledyard (1977) designed mechanisms for the free-rider problem where every participant decided to participate in the mechanisms, but not mechanisms for the problem whether or not they have incentives to participant in them. This criticism is also valid to the voluntary contribution mechanism. Cason et al. (2004) designed a two-stage game where the first stage is for participation decision and the second is for contribution decision and found that the public good provision game with the voluntary contribution mechanism became a hawk–dove game, but not a prisoner's dilemma game. Although participation is not an equilibrium strategy, in a hawk–dove game experiment with an evolutionary setting, Saijo et al. (2003) observed that among Japanese subjects evolutionarily stable strategies did not appear, but high participation emerged through a transmutation from a hawk–dove game to a game where participation to the mechanism is the dominant strategy. They also found that this transmutation is not due to kindness, but to *spitefulness* among subjects. On the other hand, in Cason, Saijo, and Yamato (2002) these results among Japanese subjects were compared with results among American subjects and found that American subjects followed relatively to evolutionarily stable strategies.

1. Saijo–Nakamura Experiments

Saijo and Nakamura (1995) conducted voluntary contribution mechanism experiments with no communication. Each of seven subjects has ten units of initial holdings for each period, and the total number of periods is ten. There are two parameters in the experiments: the marginal return from the contribution or investment [high marginal return ($a = 1/0.7$) and low marginal return ($a = 0.7$)] and the payoff information (detailed and rough tables). As for the payoff information effect, they found that results with detailed payoff tables are closer to theoretical predictions than those with rough payoff tables. In a detailed payoff table, each column represents the subject's own investment and each row represents the sum of other subjects' investment. That is, a subject can see every payoff from his strategy given the other subjects' strategies, and hence this table is exactly the same as the payoff matrix usually used in game theory. Since the difference of results between rough and detailed payoff tables may be attributed to confusion of the subjects, results with detailed payoff tables are the main concern here. Each subject participated in two marginal return experiments consecutively. For example, a subject attended a ten period experiment with high marginal return and then a ten period experiment with low marginal return. Let us denote this experiment (H, L). (H, L) experiments were repeated four times and (L, H) experiments were repeated four times.

Figure 1 shows the patterns of mean investment for (H, L) and (L, H). The horizontal axis denotes periods and the vertical axis expresses the ratio of the observed amount of investments to the sum of all subjects' initial holdings. No investment for all periods is a unique subgame perfect Nash equilibrium with $a = 0.7$ and all investments for all periods are the unique dominant strategy equilibrium with $a = 1/0.7$. Therefore, the difference between the actual investment and the equilibrium investment provides a method of measuring how far subjects are away from theoretical behavior. Figure 1 shows that the difference in high marginal return is larger than that with low marginal return. There are two basic findings with low marginal return. First, the mean investments for all low marginal return experiments are lower than those in experiments by Isaac and Walker (1988) and others. Second, although the decay effect is one of the major findings in the voluntary contribution mechanism experiments, no specific decay toward the end period is observed. These are due to the detailed payoff table effect.

Why did not subjects invest their full initial holdings in the high marginal return case? Consider the voluntary contribution mechanism with two subjects and two strategies. Each subject invests zero or ten. Table 1 shows the payoff tables with high marginal return and low marginal return where $*$ indicates the dominant strategy. Using these tables, we can construct payoff tables where the payoff for each cell is the difference between your payoff and your opponent's payoff. In fact, as Table 2 shows, these two payoff tables become the same.

If $a = 0.7$, as in the prisoner's dilemma case, no investment is still the dominant strategy in Table 2. On the other hand, if $a = 1/0.7$, all investments are the dominant strategy in Table 1 while no investment is the dominant strategy in Table 2. That is, the subject's mind wavers between investment and no investment depending on the relative strengths of the "profit" and "spite" motives. Saijo and Nakamura have named this dilemma the spite dilemma.

Since each subject participated in both high and low marginal return experiments, two mean investments can be plotted in Figure 2. The horizontal axis represents investment with $a = 0.7$ and the vertical axis represents investment with $a = 1/0.7$. When $a = 0.7$, zero investment corresponds to the free-riding side in Figure 2, and 10 corresponds to the altruism side. Similarly, when $a = 1/0.7$, 10 corresponds to the non-spite side, which is called the pay-riding side, and zero investment corresponds to the spite side. The box in the figure is divided into four although this division is arbitrary. Since the "theoretical" solution that is predicted by the dominant strategy is the upper-left corner of the box, i.e., (0, 10), the FP region represents the theoretical prediction. It is less likely to find subjects in the AS region since this region denotes subjects who invest a lot in the free-riding situation and act spitefully toward others when she can receive more than her investment. The focal point is the distribution of subjects among FP, AP, and FS regions. As Figure 3 shows, almost no subjects are found in the AP region, and

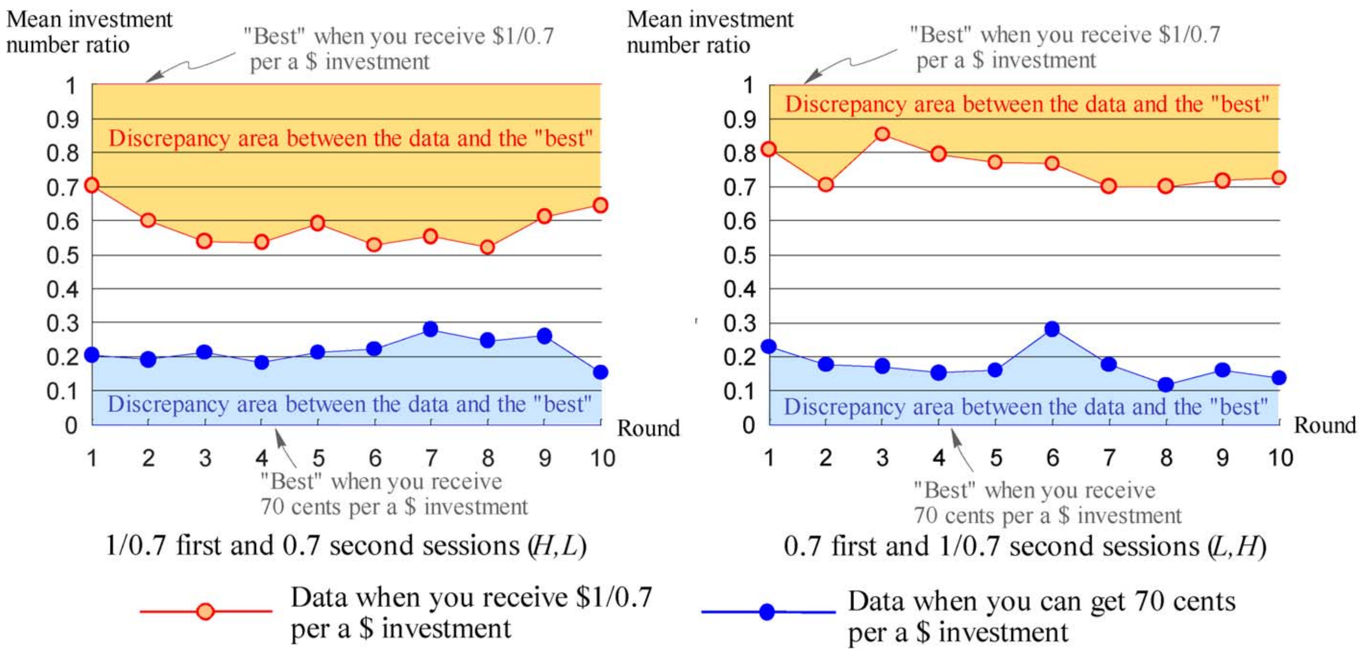


Figure 1. The discrepancies from the “best” and the mean investments.

Table 1
Two simplified payoff tables

Agent 1 \ Agent 2	Invest nothing ($0)*	Invest all ($10)
Invest nothing ($0)*	10, 10	17, 7
Invest all ($10)	7, 17	14, 14

You receive $0.7 per a $ investment

Payoff = 0.7(the sum of investment)
+ holding at hand
Both invested: payoff = 0.7(20) + 0 = 14
One invested:
Investor's payoff= 0.7(10) + 0 = 7
Non-investor's payoff = 0.7(10) +10 = 17
Neither invested: payoff = 0.7(0) +10 = 10

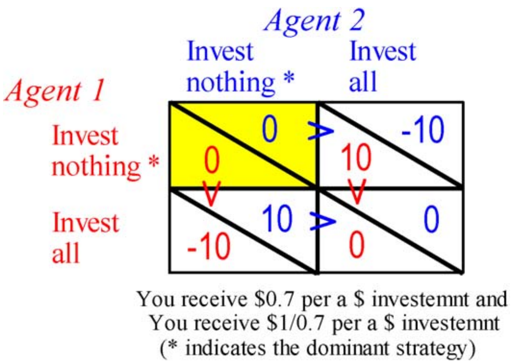
Agent 1 \ Agent 2	Invest nothing ($0)	Invest all ($10)*
Invest nothing ($0)	10, 10	24.3, 14.3
Invest all ($10)*	14.3, 24.3	28.6, 28.6

You receive $1/0.7 per a $ investment

Payoff =(1/ 0.7)(the sum of investment)
+ holding at hand
Both invested: payoff = (1/0.7)(20) + 0 = 28.6
One invested:
Investor's payoff= (1/0.7)(10) + 0 = 14.3
Non-investor's payoff = (1/0.7)(10) +10 = 24.3
Neither invested: payoff = 0.7(0) +10 = 10

(* indicates the dominant strategy)

Table 2
An illustration of the spite dilemma



The number of each cell = Your payoff in either table in Table 1 – the other subject's payoff. When you receive $0.7 per a $ investment, "investment nothing" is the dominant strategy in Tables 1 and 2. However, when you receive $1/0.7, "Invest all" is the dominant strategy in Table 1, but "invest nothing" is. A conflict between "going for money" and the "winning" is called the spite dilemma.

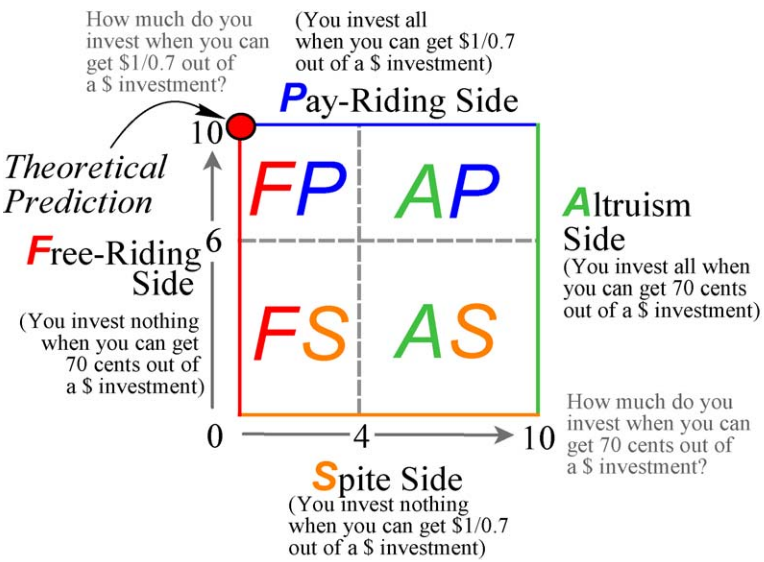
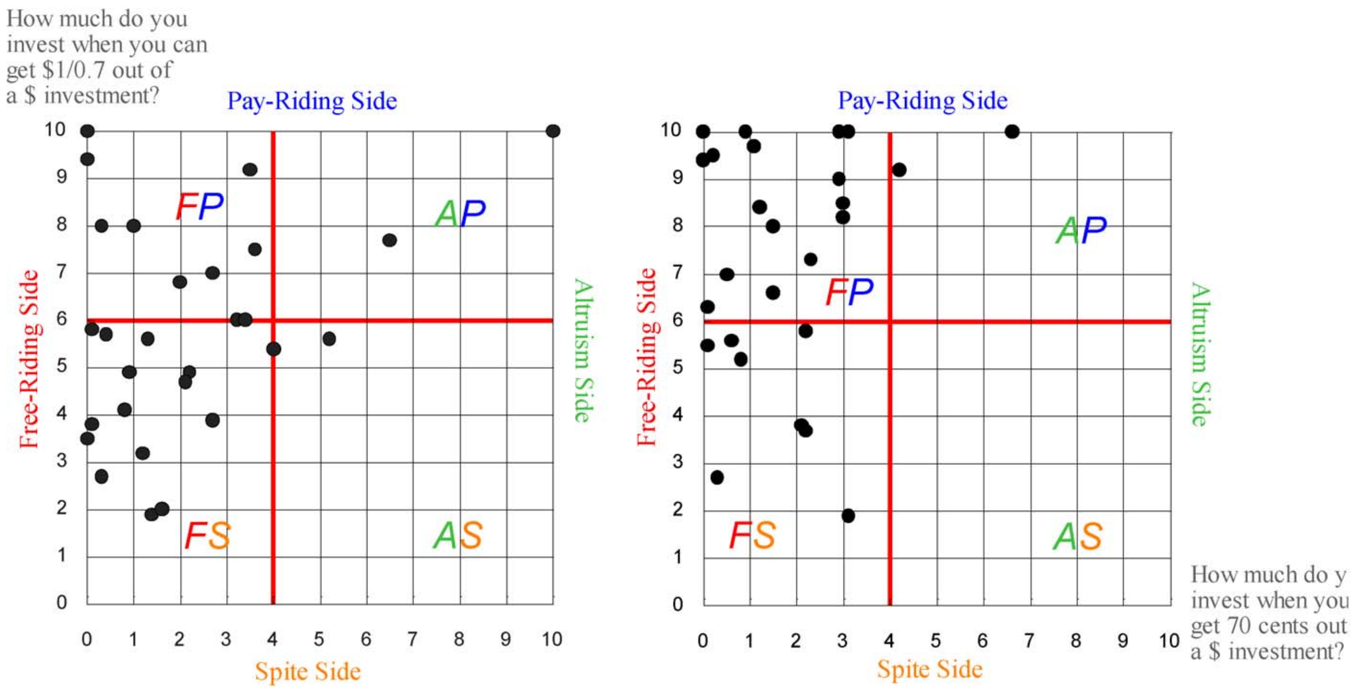


Figure 2. The mean investment distribution box.

most subjects are distributed in the FP and FS regions. That is, pay-riding subjects are "theoretical" and spiteful subjects free-ride.



1/0.7 first and 0.7 second sessions

0.7 first and 1/0.7 second sessions

Figure 3. The data of mean investment distribution of all subjects.

Table 3

A part of the detailed payoff table used in sessions A and B (no circles or shades are marked in the real experimental payoff table)

Your Investment Number

Your Opponent's Investment Number

Your Payoff	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	… 24
0	706	871	1072	1297	1536	1775	2003	2210	2386	2523	2615	2658	2648	2585	2470	
1	905					2183	2427	2641	2816	2944	3019	3039	3001	2905	2755	
2	1186					2658	2913	3129	3297	3411	3465	3456	3385	3252	3061	
3	1554					3202	3463	3675	3831	39						
4	2017					3817	4078	4281	4420	44						
5	2578	3010	3432	3831	4193	4507	4762	4950	5064	51						
6	3244	3718	4171	4590	4960	5272	5515	5681	5766	57						
7	4018	4529	5008	5440	5812	6115	6339	6478	6526	64						
8	4904	5447	5944	6383	6751	7038	7237	7340	7345	725						
9	5907	6475	6984	7422	7779	8043	8209	8271	8225	8073	7816	7458	7007	6472	5867	
10	7031	7616	8130	8561	8897	9132	9257	9270	9168	8951	8624	8193	7664	7051	6367	
11	8278	8873	9384	9800	10109	1030							35			
12	9653	10250	10750	11142	11416	1156							09			
13	11158	11749	12229	12589	12820	1291							86			
14	12796	13372	13824	14144	14323	1435							66			
⋮																
24																

Spiteful Strategy Region (you reduce your own payoff to reduce your opponent's payoff more)

The Voluntary Participation Case: The best respnse when the other subject does not participate is 11. However, if you choose 7 rather than 11, your own payoff is reduced from 2658 to 2210 (448 units), but the other subject's payoff is reduced from 8278 to 4018 (4260 units).

The Mandatory Participation Case: assume that the other player chooses 8. If you choose 7 rather than 8, your own pay-off is reduced from 7345 to 7340 (5 units), but the other subject's payoff is reduced from 7345 to 6526 (819).

Altruistic Strategy Region (you reduce your own payoff to increase the other subject's payoff)

Nash Best Response Region

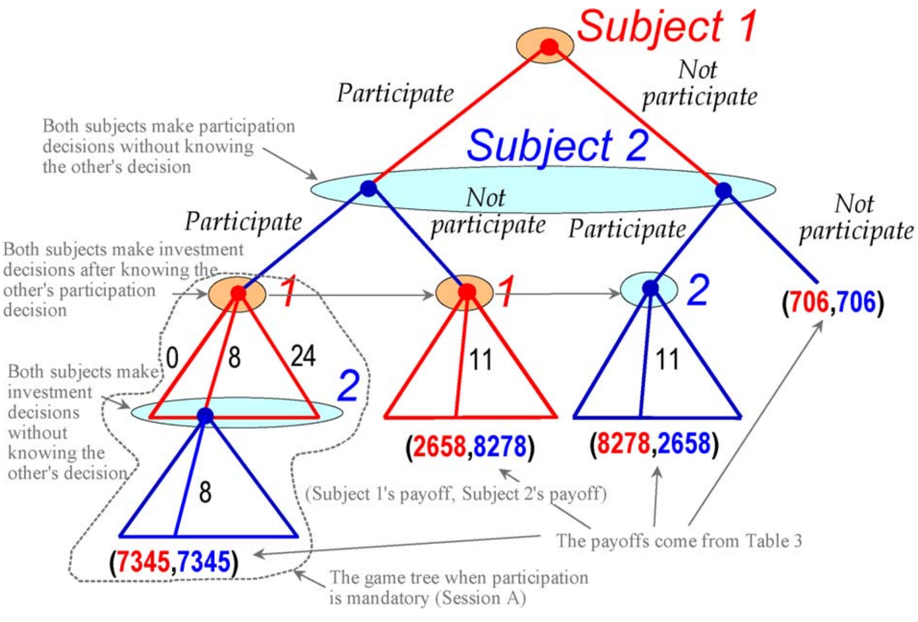


Figure 4. The two stage game tree when subjects can choose their participation in the voluntary contribution mechanism.

2. Non-excludable Public Good Experiments

Cason et al. (2004) and Cason, Saijo, and Yamato (2002) conducted voluntary contribution mechanism experiments with two subjects who had the same Cobb–Douglas utility function,

$$u_i(x_i, y) = \frac{(x_i^{0.47} y^{1-0.47})^{4.45}}{50} + 500,$$

where x_i is subject i's consumption of a private good and y is the consumption of a public good. They assumed that the sum of contribution of two subjects becomes the level of the public good and each subject has 24 units of initial holdings; see Figure 4. Both subjects are required to participate in the mechanism in Session A. On the other hand, each subject can choose her participation decision before her contribution decision in Session B. That is, in the first stage, subjects choose whether or not they participate in the mechanism at the same time, and in the second stage, knowing the other subject's participation decision, subjects who selected participation in the first stage choose contributions to the public good. Therefore, Session A includes the second stage while Session B includes both stages. Each session has twenty subjects and each subject is randomly paired with the other subject. The same game was repeated 19 periods, 4 for practice and 15 for monetary reward, so as not to pair the same two subjects more than once. They used detailed payoff tables in both sessions (see Table 3).

Table 4
The payoff table becomes a hawk–dove game

1 \ 2	p_2 *Participate*	$1-p_2$ *Not Participate*
Participate p_1	7345, 7345	2658, 8278
Not Participate $1-p_1$	8278, 2658	706, 706

(7345, 7345) are based on the Nash equilibrium second stage investments of (8, 8) when both participate in the first stage (see Table 3 and Figure 4). (2658, 8278) is based on the equilibrium investments of (11, 0) when only subject 1 participates. The subgame perfect Nash equilibria are participation probabilities for subjects 1 and 2 ($p1$, $p2$) of (1, 0) (0, 1), and (0.68, 0.68). The unique evolutionary stable strategy (ESS) is ($p1$, $p2$) = (0.68, 0.68).

Transforming this two-stage game in Session B to a normal form game, they show that the game becomes a hawk–dove game, but not a prisoner's dilemma game introducing participation as a decision variable (see Table 4). This game has two pure strategy Nash equilibria and one mixed strategy Nash equilibrium that is the unique evolutionarily stable strategy (ESS) equilibrium. Therefore, Session B can be considered as a test of the ESS prediction.

Cason et al. (2004) used Japanese subjects and Cason, Saijo, and Yamato (2002) compared the results with American subjects. Both used students in several universities in each country, but the following are summaries of results from University of Tsukuba in Japan and University of Southern California in the United States.

In Session A, as Table 3 shows, the Nash equilibrium investment pair is (8, 8). Since each period had 10 pairs and 15 periods were conducted, each session generated 150 pairs of data. Figure 5 shows the frequency distributions of investment data rearranging each pair (a, b) with $a \geqslant b$ since the order of investment numbers does not matter. The maximum frequency pair is (8, 8) with 36 pairs at Tsukuba and with 29 pairs at USC. Both are similar, but Tsukuba rarely has data such as (a, b) with $a, b \geqslant 8$ and USC has data such as (9, 8) with 14 pairs and (12, 8) with 5 pairs.

Figure 6 shows the average investment pattern of Session A. In Tsukuba's case, the average investment is less than the Nash equilibrium level of investment in all periods except for one. In order to understand the subjects' behavior throughout the experiment, each subject was asked to write her reason for the investment decision in each period. According to this information, there are four subjects who explicitly stated the following: they estimated that their opponent would choose 8, and then they chose 6 or 7 because this would make their opponent's payoff much lower than their own payoff. For example, suppose that your opponent chooses 8 in Table 3. If you choose 8, you and your opponent obtain 7345. If you choose 7, you obtain 7340, but your opponent obtains 6526. By choosing 7, you can make the reduction of your opponent's payoff

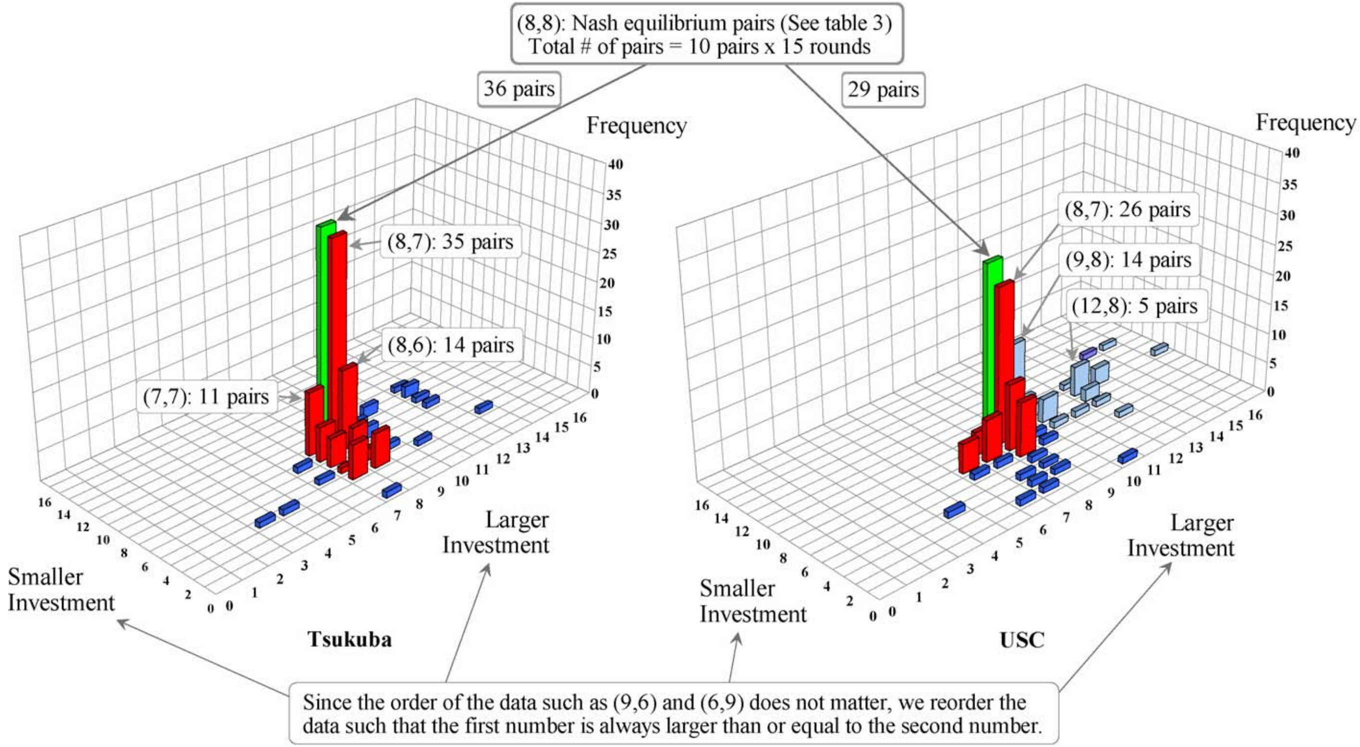


Figure 5. The distribution of investment pairs when two subjects must participate in the mechanism.

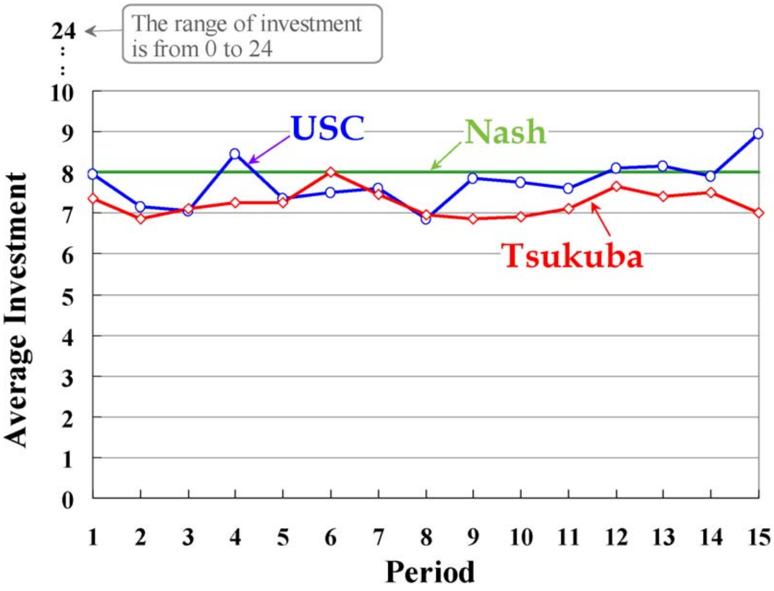


Figure 6. Average investments when subjects must participate in the mechanism.

($819 = 7345 - 6526$) much larger than your own reduction ($5 = 7345 - 7340$). Employing a random effects error specification, Cason, Saijo, and Yamato (2002) found that the pooled data strongly reject the Nash equilibrium, even when these four spiteful subjects and their opponents' data are excluded. The average USC sequence, however, is close to the Nash equilibrium contribution.

Figure 7 shows that the distributions of investment pairs in Session B are very different from those of Session A. In Tsukuba, the maximum frequency pair is (8, 7) with 28 pairs, and the second is (8, 8) and (0, 0) with 18 pairs, and the fourth is (7, 0) with 15 pairs. At USC, the maximum frequency pair is (11, 0) with 36 pairs, and the second is (0, 0) with 25 pairs, the fourth is (8, 7) with 11 pairs, and the fifth is (7, 0) with 10 pairs.

Figure 8 shows the participation ratio pattern of Session B. In Tsukuba, although the participation ratio starts with 40% in period 1, it gradually increases and is more than 85% in the last 4 periods. On the other hand, the USC data is close to the ESS prediction where the participation ratio should be 68%. Due to the participation ratio difference between the two schools, Tsukuba's average investments toward the end period are higher than those at USC. As a result, Tsukuba provides the public good more than USC does.

In order to understand these rather high participation ratios, consider the case where only one subject participates in the mechanism. As Table 3 shows, the participant should invest 11 to maximize her own payoff. In this case, the maximum frequency investment number is 7, and no participant invested more than 11. If the participant invests 11,

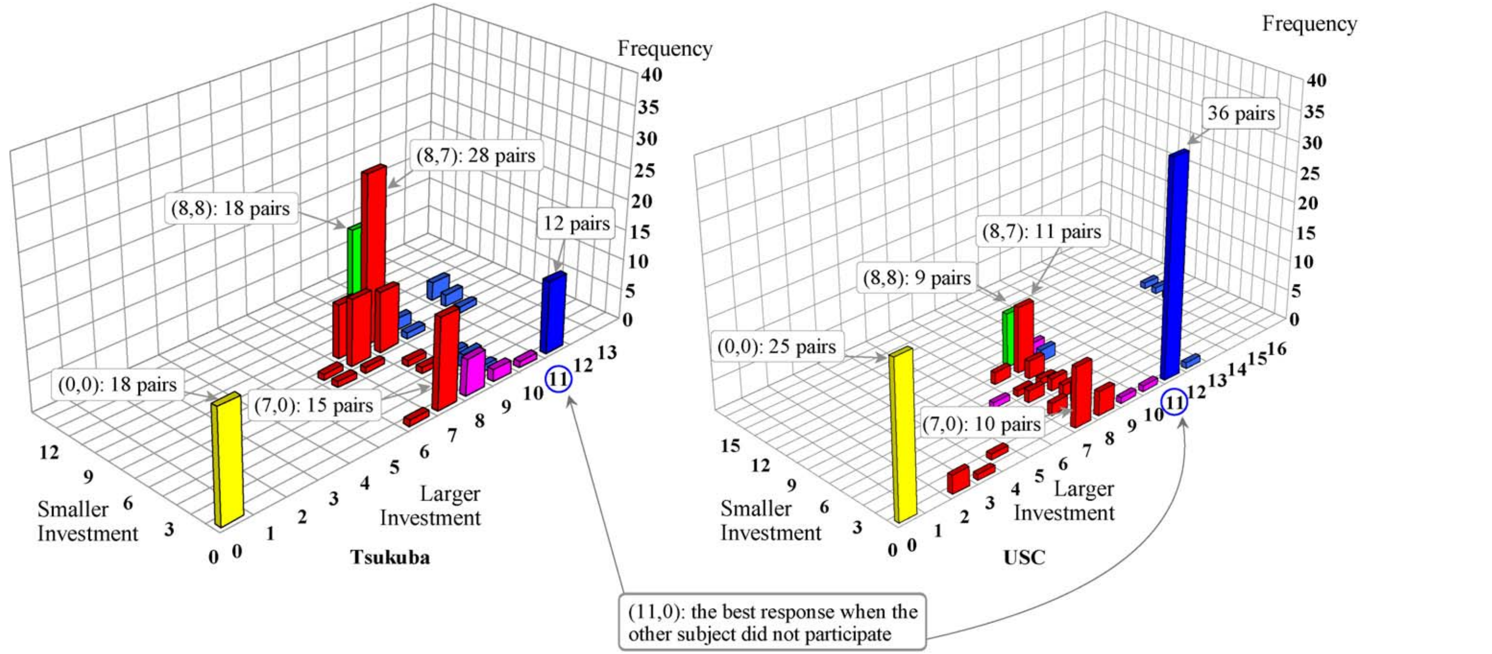


Figure 7. The distribution of investment pairs when two subjects can choose to participate in the mechanism.

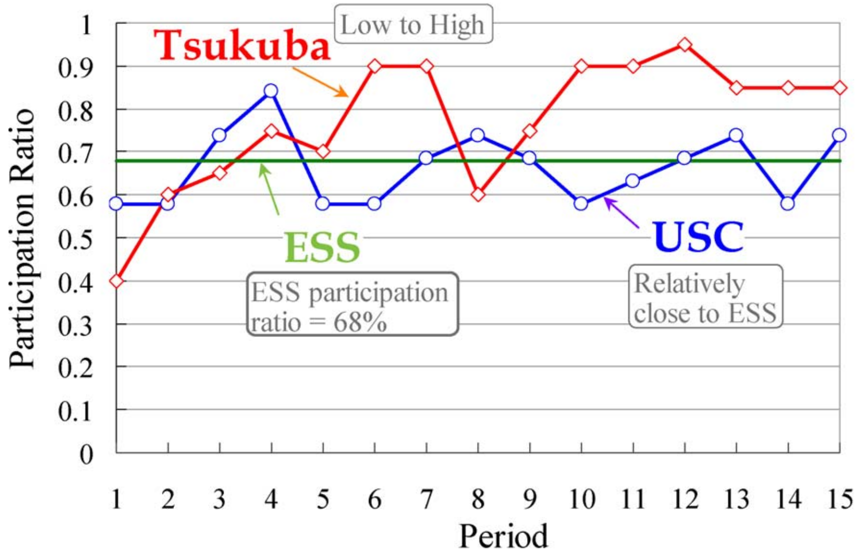
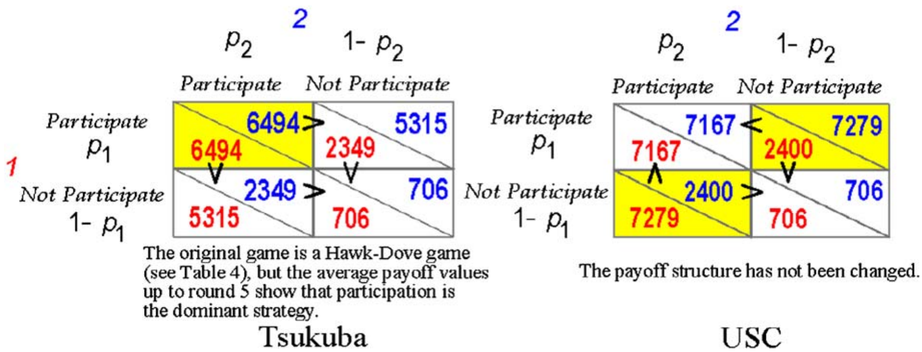


Figure 8. Participation pattern.

Table 5
The average values of payoff data up to round 5



then she earns 2658 and her opponent earns 8278. On the hand, by investing 7 in this situation, the non-participant obtains 4018 while the participant earns 2210. Thus, the reduction of the participant's payoff (448 = 2658 − 2210) is relatively small, while the reduction of the non-participant's payoff (4269 = 8278 − 4018) is relatively large. That is, subjects seemed to punish (or behave spitefully toward) the non-participant choosing 7 although no subjects are paired twice or more. On the other hand, it is hard to interpret the "spiteful" behavior as punishment in Session A since the game in this session has just one stage.

Table 5 illustrates the average values of payoff data up to period 5. In Tsukuba, although no dominant strategy exists in Table 4 representing a theoretically predicted payoff table, Table 5 shows that participation is the dominant strategy. In other words, after non-participants have experienced spiteful behavior from participants, non-participants realize that non-participation is not a good strategy and then find that participation is the dominant strategy. That is, evolutionarily stable strategies do not appear, and participation emerges through a transmutation from a hawk–dove game (Table 4) to a game where a dominant strategy outcome is Pareto efficient (Table 5). On the other hand, in the USC experiments, the basic structure of the game is not changed as the rounds proceeded.

Acknowledgement

This research was partially supported by a grant from the Aid for Scientific Research 08453001 of the Ministry of Education, Science and Culture in Japan.

References

Andreoni, J. (1993). "An experimental test of the public-goods crowding-out hypothesis". American Economic Review 83 (5), 1317–1327.

Andreoni, J. (1995). "Cooperation in public-goods experiments: Kindness or confusion?" American Economic Review 85 (4), 891–904.

Cason, T.N., Saijo, T., Yamato, T. (2002). "Voluntary participation and spite in public good provision experiments: An international comparison". Experimental Economics 5, 133–153.

Cason, T.N., Saijo, T., Yamato, T., Yokotani, K. (2004). "Non-excludable public good experiments". Games and Economic Behavior 49, 81–102.

Groves, T., Ledyard, J.O. (1977). "Optimal allocation of public goods: A solution to the 'free rider' problem". Econometrica 45, 783–811.

Isaac, R., Walker, J. (1988). "Group size effects in public goods provision: The voluntary contribution mechanism". Quarterly Journal of Economics 103, 179–200.

Ledyard, J.O. (1995). "Public goods: A survey of experimental research". In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–194.

Olson, M. (1965). "The Logic of Collective Action: Public Goods and the Theory of Groups". Harvard University Press, Cambridge.

Saijo, T., Nakamura, H. (1995). "The 'spite' dilemma in voluntary contribution mechanism experiments". Journal of Conflict Resolution 39, 535–560.

Saijo, T., Yamato, T. (1999). "A voluntary participation game with a non-excludable public good". Journal of Economic Theory 84 (2), 227–242.