How Universal is Behavior? A Four Country Comparison of Spite, Cooperation and Errors in Voluntary Contribution Mechanisms

by

Jordi Brandts*
Tatsuyoshi Saijo**
Arthur Schram***

August 2002

ABSTRACT

This paper studies behavior in experiments with a linear voluntary contributions mechanism for public goods conducted in Japan, the Netherlands, Spain and the USA. The same experimental design was used in the four countries. Our 'contribution function' design allows us to obtain a view of subjects' behavior from two complementary points of view. It yields information about situations where, in purely pecuniary terms, it is a dominant strategy to contribute all the endowment and about situations where it is a dominant strategy to contribute nothing. Our results show, first, that differences in behavior across countries are minor. We find that when people play "the same game" they behave similarly. Second, for all four countries our data are inconsistent with the explanation that subjects contribute only out of confusion. A common cooperative motivation is needed to explain the data.

Keywords: experimental economics, cooperation, public goods games.

JEL Classification: C90, C91, D63.

*Instituto de Análisis Económico (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain. Email: brandts@cc.uab.es.
**Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567, Japan. Email: saijo@iser.osaka-u.ac.jp
***Corresponding author: CREED, Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. Email: ArthurS@fee.uva.nl

This paper is part of the EU-TMR Research Network ENDEAR (FMRX-CT98-0238). It was completed while Brandts was visiting the Department of Economics of the University of California at Berkeley. He thanks the members of the Department at Berkeley for their hospitality. The authors thank Isabel Busom, Gary Charness and Theo Offerman for helpful comments and the Spanish DGCICYT (PB93-0679, PB94-0663-C03-01 and PB98-0465) for financial support.

1. Introduction

Standard economics is not founded on the view that human behavior is country-invariant. One of the central tenants of economics concerns the internal coherence, rather than the content, of individual preferences. In practice, however, economists have rarely considered cross-country differences in preferences and implicitly assume invariance across countries. An example of this lack of attention is the emphasis on differential factor endowments, rather than on preferences, in international trade theory. To us the question of the universality of behavior is fundamentally an empirical one and should be analyzed on the basis of data. This is our main aim in this paper.

It is difficult to study this issue using only field data, as these are generated in environments where a number of influential factors vary in character and intensity. This makes a direct evaluation of cultural differences problematic. Experimental methods allow for a very controlled approach to the issue at hand. Examples of studies that illustrate the virtues of the experimental approach in this area are Kachelmeier and Shehata (1990), Roth, Prasnikar, Okuwo-Fujiwara and Zamir (1991) and Burlando & Hey (1997). The degree of control that can be achieved in experiments appears to be especially desirable when the objective is the analysis of cross-country differences in motivational factors of human behavior.

In this paper we present a broad comparison of behavior in Japan, the Netherlands, Spain and the United States using a common experimental design. We use a voluntary contributions mechanism (hereafter, vcm) for linear public goods to obtain information about motivation in a context of interdependence. The inclusion in our study of specifically these four countries is motivated by a number of considerations. First, together they cover a good part of the developed world; three different continents and, within Europe, the Germanic North and the Latin South. Second, in previous experiments, reported in Saijo and Nakamura (1995), Japanese subjects were found to deviate from standard behavior in a surprising manner discussed below. Our design allows us to explore this issue in a comparative way.

The third reason for studying the four included countries is that previous non-experimental work does find considerable differences between the four countries we study. Hofstede (1991) reports results from a questionnaire study conducted in 50 countries and finds considerable differences. For instance, the Netherlands and the US both rank highly on his *individualism index* and lowly on his *uncertainty avoidance index*, while Japan and Spain rank

both lowly on the first of the two indices and both highly on the second of the two indices. In contrast, on *a masculinity index* Japan and US both rank highly and the Netherlands and Spain both rank lowly. The results of our study will show whether this kind of differences between the four countries we selected show up in a more controlled environment.

The design we use in this paper makes it possible to study, at the same time, two different types of behavior, both not exclusively motivated by subjects' own monetary payoffs. In this way we are able to obtain a view of subjects' behavior from two complementary points of view. We examine the case where contributions lead to a decrease in the contributor's monetary payoff and to an increase of others' payoffs. Numerous studies report evidence of contributions in this kind of situations. Ledyard (1995) and Schram (2000) present overviews of these results. This behavior has often been attributed to some kind of 'altruistic' or 'cooperative' motivation: subjects' utility may in some way depend positively on others' payoffs.

We also study behavior for the case where deviations from full contribution leads to monetary losses for the deviating individual. Saijo and Nakamura (1995) motivate our interest in this case. They report evidence from a set of experiments conducted at the University of Tsukuba, using a linear public goods environment. Their central result is that subjects deviated to a substantial degree from full contribution, the strategy which was dominant from a purely pecuniary point of view. To date, this kind of behavior has not been observed in other countries.[1]

This surprising result can be interpreted in terms of a 'spiteful' or 'competitive' value orientation. Subjects may be motivated not only by the absolute amount of the payoff they obtain, but also by the relation of their own payoff to the payoffs of others.[2] In the context considered, a subject's decision not to contribute his whole endowment imposes a certain cost on himself, but an even larger cost on the other participants in the vcm; if spite is sufficiently strong, subjects may be willing to give up some of their payoff to hurt others.

Viewed in isolation the two types of deviations from own-payoff maximization mentioned above can be interpreted in terms of motivational factors. However, taken together, they point to an alternative interpretation of behavior purely in terms of decision errors. If one adopts this

[1] From now on we will use the terms *dominant strategy* and *efficiency* in relation to a purely monetary motivation, even though we will, at the same time, be considering other types of motivations.

[2] Saijo, Yamato, Yokotani and Cason (1997) contains evidence of a different type of spiteful behavior by subjects at the University of Tsukuba.

view, one can interpret the two types of departures from the dominant choice in a consistent way: subjects make the mistake of contributing when it is a dominant choice not to do so and they also mistakenly fail to contribute when a dominant strategy prescribes to do it.

We believe that these two issues, whether behavior varies across countries and what the roles of spite, cooperation and errors are in explaining behavior, pertain to one of the most basic questions in the area: do "other-regarding" motivations really exist and how universal are they? To deal with these questions we conduct a systematic investigation in a common framework. To study whether subjects' motivations differ across countries the same design should be used in different locations; to ascertain whether subjects behave similarly in the two types of situations described, a common design is also necessary. As will be explained in section 2, some of the features of our design allow for a very efficient data collection and make it possible to investigate, at the same time, the two issues mentioned above.

The paper is organized as follows. In section 2 we present the experimental design (with special attention to how we control for the effects of between-country variables), the treatments and the procedures. Section 3 describes the results and Section 4 presents a statistical analysis of spite, cooperation and errors. Section 5 summarizes and concludes.

2. Experimental design.

2.1. General features of the design

Our experimental design is based on the linear public good environment first used in Brandts and Schram (2001); an earlier related design was used by Palfrey and Prisbrey (1996). The key feature of our design is that subjects do not just make a single decision in each experimental period. Instead, each subject makes 10 decisions in each round, corresponding to 10 different marginal rates of substitution (hereafter, mrs) between the private and the public good. A set of 10 decisions of this type, which we will refer to as a 'contribution function', will yield information about cooperative and spiteful behavior by subjects, as well as about the presence of decision error. In our baseline design the contribution function will provide a full picture of subjects' motivation. This is so because the different mrs for which our subjects have to make decisions are of three types. For all three types decision error can lead to deviations from the dominant choice. For some of the mrs a spiteful, but not a cooperative, motivation may lead to deviations from dominant strategy play. For

some other mrs, the reverse will be true: a cooperative motivation, and not a spiteful one, may lead to deviations from the dominant strategy. For a third subset of mrs neither spite nor most cooperative motivations can explain behavior and decision error can, therefore, be considered the most likely source of contributions. A second design will yield more detailed information for a specific subset of mrs.

In our experiments every subject is asked to divide 9 tokens by investing each in one of two accounts, A and B. An individual's investment of one token in account A yields an identical amount of money to every member of the group to which the individual belongs. A token placed in account B yields a certain amount of money only to the subject who makes the investment.[3] This division has to be made for 10 different mrs, for which the payoff to the public account was kept constant, while the payoff to the private account varies; in the experiments these different mrs are called 'situations'.[4] We conducted some baseline sessions and some additional sessions; the difference between them is in the payoff schedule we employed. The design used in our baseline sessions includes values of the mrs corresponding to the three different cases described above, while the design for the additional sessions concentrates on the first two types of situations.

Formally, the payoff $_{it}$ to subject i in round t is given by

$$_{it} = r_{s^*t} (\quad - x_{is^*t}) + m \sum_{j=1}^{4} x_{js^*t}, \; i=1,2,3,4, \; s^* \in \{1,..,10\}, \; t=1,..,10 \qquad (1)$$

where s* is the situation selected by the random draw (which varies across t), r_s is the return from the private account in situation s (it varies across countries and across the 10 situations); the endowment, , is 9 tokens; x_{ist} is i's contribution to the public account for situation s in round t; and m is the return from the public good. Note that the mrs is equal to r_s/m.

Table 1 shows the payoff schedules (r_s and m) used in the different countries in the baseline sessions and table 2 gives the ones used in the additional sessions. For each country column A (B) gives the payoff per token in the public (private) account, m (r_s). The last column on the right of the table gives the different mrs, which are common for all countries. Note that for the parameter

[3] Accounts A and B correspond, respectively, to what is often called a 'public account' and a 'private did, however, not use these terms in the experiments.

values shown in table 1, it is a dominant choice to invest all tokens in account A for situations 1 and 2 and to invest everything in B for situations 3 to 10. This is due to the differences between the payoffs in the public and private accounts.

Table 1: Payoff schedules, baseline (symmetric) sessions*

	Japan		the Netherlands		U.S.A.		Spain		
	A m	B r_s	A M	B r_s	A m	B r_s	A m	B r_s	MRS r_s/m
situation 1	8	2	12	3	8	2	8	2	0.25
situation 2	8	6	12	9	8	6	8	6	0.75
situation 3	8	10	12	15	8	10	8	10	1.25
situation 4	8	14	12	21	8	14	8	14	1.75
situation 5	8	18	12	27	8	18	8	18	2.25
situation 6	8	22	12	33	8	22	8	22	2.75
situation 7	8	26	12	39	8	26	8	26	3.25
situation 8	8	30	12	45	8	30	8	30	3.75
situation 9	8	34	12	51	8	34	8	34	4.25
situation 10	8	38	12	57	8	38	8	38	4.75

*The numbers are the payoff per token, in Japanese yen, Dutch cents, American cents, and Spanish pesetas. The numbers m (r_t) are the payoff per token in the public account A (private account B).

We allocate subjects to groups of size 4. It is easy to see that for this case the efficient solution is for all participants to invest all tokens in A (B) for situations 1 to 8 (9 to 10). For situations 1 and 2 a subject hurts himself monetarily by deviating from dominant strategy play, but he hurts the others in the group even more. At the same time, by contributing less than others a subject's payoff will be larger than that of the others.[5] For situations 3 to 10, we have the more standard

[4] The mrs is just the reciprocal of the marginal per capita return (mpcr); we use the term mrs following Palfrey and Prisbrey (1996).

[5] Consider, for instance, situation 1 for a group of four subjects in Japan. If all of them contribute 9 tokens to account A, each of them will obtain a payoff of 8x36=288 yen. In comparison if three subjects contribute 9 tokens

environment where by contributing and, hence, deviating from dominant strategy play a subject reduces his own payoff but increases that of the other subjects. For situations 3 to 8 a subject's contribution lowers his own payoff less than it increases the total payoff.

Situations 9 and 10 differ in an important way from all other situations: contributing is inefficient. To see this observe that the mrs of both these situations are larger than 4, the number of subjects. The payoff from the private account is more than 4 times the payoff from the public account and, hence, the decrease in the payoff of anyone who contributes is larger than the increase of the payoffs of the other group members together.

In contrast, in the design used in the additional sessions the efficient solution is for all to invest all tokens in A in every situation. This can be seen in table 2. For situations 1 to 5 (6 to 10) the dominant strategy is to invest all tokens in A (B). We chose the situations in the additional sessions to be 'symmetric' in their mrs, around mrs=1. Therefore, from here onward we will denote the baseline sessions as the 'Asymmetric' (AS) sessions and the additional sessions as the 'Symmetric' (S) ones. The AS sessions yield a more complete picture of behavior while the S sessions make it possible to check the robustness of our results.

Besides the cross-country comparison and the difference between the AS and the S sessions, there is one more dimension to our design. Since Andreoni (1988) there has been a discussion about the differences in behavior in a 'partners' and 'strangers' design.[6] His surprising result that partners cooperate less (free-ride more) than strangers has not always been replicated (see Keser and van Winden, 1997, for a discussion). Weimann (1994) finds the opposite result and Burlando and Hey (1997) attribute the differences to national differences. We believe that this issue is of interest because it yields information about the importance of strategic thinking in subjects' behavior; a lack of differences in behavior between partners and strangers would point to the absence of the kind of strategic thinking that is characteristic of the game-theoretic analysis of repeated interaction. Our data set allows us to study this issue across four countries. We distinguished partners and strangers sessions in our AS sessions. In the S sessions we only used the

---

and the fourth subject contributes 0 to the A account, then each of the three subjects who fully contribute will earn 8x27=216 yen and the remaining subject will earn 216+18=234 yen.

[6] During a typical experimental session, subjects are allocated into different groups; a group is defined by the set of subjects that are connected through the public goods payoff function. In partners, groups are constant across the different periods of an experimental session, whereas in strangers subjects are reallocated to new groups in every period.

partners treatment.

Table 2: Payoff schedules, additional (asymmetric) sessions*

	Japan		the Netherlands		
	A m	B r_s	A m	B r_s	MRS r_s/m
Situation 1	20	2	30	3	0.1
Situation 2	20	6	30	9	0.3
Situation 3	20	10	30	15	0.5
Situation 4	20	14	30	21	0.7
Situation 5	20	18	30	27	0.9
Situation 6	20	22	30	33	1.1
Situation 7	20	26	30	39	1.3
Situation 8	20	30	30	45	1.5
Situation 9	20	34	30	51	1.7
Situation 10	20	38	30	57	1.9

*The numbers are the payoff per token, in Japanese yen and Dutch cents. The numbers in the columns 'A' ('B') are the payoff per token in the public account A (private account B). The mrs is defined as in table 1.

2.2. Controlling for between-country variables.

Following Roth et al. (1991) we consider three aspects of the design which require special attention when conducting a multi-national experiment: experimenter effects, language effects, and currency effects.

The term *experimenter effect* refers to the possibility that different sessions of the same experimental treatment may yield different results, due to possible effects of uncontrolled procedural differences across locations, including the fact that the sessions were conducted by different experimenters. We dealt with these difficulties by first establishing a protocol for the procedures[7] in the initial sessions in Amsterdam, and then using this standard in the sessions at the other locations. The first Amsterdam sessions were conducted by Schram in Amsterdam with

Brandts observing. Brandts then ran the sessions in Barcelona, with Schram observing in some of these.[8] Subsequently, Schram ran the sessions in Arizona. The sessions in Osaka were run by Saijo and his collaborators with Brandts observing.[9] In this way we tried to ensure that the standard established in Amsterdam 'trickled down' to the other locations. We did not explicitly deal with possible uncontrolled personal differences among the experimenters. We believe that the fact that all our sessions were computerized and that all communication with subjects was explicitly included in the English description of the protocol (which included written texts for all interaction with subjects) reduces the potential effects of this kind of influence.

Second, to control for unwanted *language effects* the instructions for the experiment were initially written in English, and then translated into Dutch, Spanish and Japanese. The experimenter responsible for each translation is a national of the country in question who has lived extended periods in the United States. We, therefore, believe that all the translators had a clear understanding of the possible connotations and nuances present in the English instructions.

The translation of our instructions into Japanese raised some special design problems. In the sessions run in the Netherlands, in Spain and in the United States we translated the computer interface of the experiment into the language of the country; the instructions that appeared on the computer screen were also in the language of the country in question. Doing this for the Japanese case would have presented us with some very important difficulties having to do with the use of Japanese characters. The solution we chose for the Japanese sessions was to have the English version of the instructions on the computer and, at the same time, to give a Japanese translation to the subjects. The translation consisted of a separate sheet for each screen of the computerized instructions. The layoff of each page reproduced exactly the layout of the corresponding screen. We believe that this is a reasonable solution, even though it does introduce a certain asymmetry between the Japanese sessions and the rest.[10]

---

[7] A detailed description of the procedures was used. This is available from the authors.

[8] Joep Sonnemans of CREED also observed some of the first Amsterdam sessions. He later went to Barcelona and helped run some sessions there. We thank Joep for his help and note that his presence creates an additional link between the two locations in terms of procedural control.

[9] We would like to thank Yasuyo Hamaguchi for her help in running the Japanese sessions and Steve Backerman and Matt Cox for their help in organizing the U.S. sessions.

[10] If this asymmetry is responsible for an observed difference in behavior between countries, a reasonable conjecture may be that it led to more confusion among Japanese subjects than among subjects in the other three countries. In this case, one should probably observe more deviations from the dominant strategy, independently of the mrs. As will be discussed below, this does not appear to be the case in our data.

With respect to currency effects, first note that one of the most important variables in our design is the mrs, for which the same values are used in all countries. At the time of the experiments, the exchange rates (for the currencies in Japan, the Netherlands, Spain, and the United States, respectively) were ¥ 90 ≈ *f* 1,60 ≈ Ptas 114 ≈ $1. Note that the numbers in tables 2 and 3 imply that we have used the exchange rate ¥ 100 ≈ *f* 1,50 ≈ Ptas 100 ≈ $1. Hence, using the exchange rates to compare payments in other countries with those in the United States, the stakes were slightly higher in Japan (11%) and slightly lower in the Netherlands (6%) and Spain (12%). These differences are small and in the direction of our observations of purchasing power in these countries. Therefore, we do not expect that differences in payoff saliency will affect our results.

2.3. Experimental procedures and subject pools

The experiments were all conducted in computerized laboratories. The Japanese sessions were run at the Institute of Social and Economic Research at Osaka University, the Dutch sessions at the CREED laboratory at the University of Amsterdam, the Spanish sessions at the LEEX laboratory of the Pompeu Fabra University in Barcelona and the United States sessions at the ESL laboratory at the University of Arizona.[11]

For each session 13 subjects were recruited among students; the procedures were the same in all countries.. One of these subjects (randomly selected) was appointed monitor. The monitor was allowed to observe the experiment to make sure everything was done as explained in the instructions. In addition, he made the random draws (see below).

In each session 13 subjects were brought into the laboratory and randomly appointed seats. They were separated by partitions and no communication was allowed. Instructions were computerized and subjects could read them at their own pace; see the appendix for the English version of the instructions. After subjects had finished reading the instructions, two practice periods were run, in which it was made clear that there was no interaction: the computer randomly determined the 'decisions of the other participants'. Then the 10 periods of the experiment began.

The S sessions were only run in Japan and the Netherlands and all used the partners design (two sessions in each country). In each country, we ran two AS-partners and two AS-

[11] We would like to thank Antoni Bosch of LEEX and Vernon Smith of ESL for allowing us to use their facilities. More details about the sessions are available from the authors.

strangers sessions. Hence we have data from 6 sessions in Japan and the Netherlands, 4 sessions in the United States and in Spain. The total number of subjects was 240.

Each experimental session consisted of ten periods; we used a multi-period procedure to follow standard practice. In each period, subjects were anonymously allocated to three groups of four, in accordance with either the strangers or the partners treatment. In each period each subject selected a complete contribution function. In the sessions the payoff schedule was presented in tabular form on a handout, on an overhead projector, and on the computer screen. Each subject could determine in what order he wanted to decide on the 10 situations. The decision for any situation could be revised as long as all decisions had not been finalized.

After all the participants had finalized the division of tokens for all situations for a period, one situation was randomly selected to determine payments. The monitor selected one situation per group by throwing a ten-sided die. Thus, the investments in the group account and the payoffs were determined for the same mrs for each of the four subjects in a group. The monitor entered the result for each group in his computer, and the subjects were told the result for their group privately, through their computers. In summary, each period consisted in each subject making his decisions for each of the situations and in one random draw of a situation to determine payoffs.[12] The relevant payoff table as well as the other information about procedures was known to all subjects in a session.[13]

Table 3 gives some background data for the subjects in the AS sessions. We obtained these data in a post-experimental questionnaire we used at all locations. Observe that males are overrepresented in all countries and that the fraction of economics students varies considerably across countries. In our statistical analysis in section 4 we control for the subject background variables' possible impact on behavior.[14]

---

[12] Cubbitt, Starmer and Sugden (1998) investigate the impact of a random lottery incentive system, similar to the one we use, on subjects' behavior in an individual choice experiment. They find no significant effect. In section 3 we will show that, for some of the values of the mrs for which there are results from previous experiments, our data are generally in line with those results.

[13] All calculations and registration were computerized. Subjects received an overview of all previous periods when the results for a period were given and could simply recall this information at any time. Per period, the information included the situation selected for that subject, her/his division of tokens for that situation, the group investment in A, and the earnings. It also gave the total earnings at that point. In addition, subjects could easily recall their decisions for all situations in the previous period.

[14] The use of student subjects may be considered a drawback of our study: comparing the behavior of young, highly educated, people may not give the "national culture matters" view its best shot. However, previous

Table 3: Subject Background; AS sessions*

		Japan	Netherlands	U.S.A.	Spain
Gender	male	94.3	74.5	60.5	66.7
	female	5.7	24.5	39.5	33.3
major in economics		1.9	65.9	57.6	89.6
	minimum	18	19	18	18
Age	mean	21.35	22.8	21.9	19.6
	maximum	25	31	33	24

*The numbers in the cells are percentages from the post-experimental questionnaire, for the baseline treatments. Missing cases have been dropped.

3. Presentation of the results using contribution functions

In this section, we present a general overview of our results. We do so by using average *contribution functions* as a convenient device for describing behavior across countries and treatments. Statistical tests are reported in section 4. A contribution function gives the number of tokens invested in the public account as a function of the mrs. Note that standard theory predicts an individual's contribution to be 9 tokens if mrs < 1 ($r_s < m$) and 0 tokens for mrs > 1 ($r_s > m$). This prediction, that the contribution function is a step function with only one step from 9 to 0 at mrs = 1, serves as a benchmark to which actual behavior can be compared.

Figure 1 (a-d) presents contribution functions per country for the AS design, where for each mrs the contributions have been averaged over all participants and all periods in a given country. At this point we wish to single out a few fundamental traits of our data. First, all eight aggregate contribution functions for all four countries and both treatments are monotonically

experimental studies that did find differences, like Roth et al. (1991), did also rely on student subjects. In addition, the Hofstede (1991) questionnaire study compared behavior across countries, but for similar occupations. The fact that our subjects may have a common background makes our study more comparable to previous work.

decreasing.[15] Second, at first sight, the eight contribution functions also appear to be rather similar in a more quantitative sense, i.e. for given mrs contribution levels are rather similar. Note also that for situations 9 and 10 the average number of tokens contributed to the public good is always less than one token.

(Figure 1 about here)

Figure 2 shows the results from the S sessions conducted in the Netherlands and Japan. Just as for the AS sessions, cross-country differences in behavior do not appear to be large. Observe also that the deviations from the dominant choice in situations 1-5 (contribute 9 tokens) seem to be smaller than for situations 6-10 (contribute 0 tokens).

(Figure 2 about here)

When using a new experimental design for an old economic problem, the question whether it yields different behavior than previous designs should be addressed. It turns out that our results are quite similar to those of previous studies. For mrs > 1, the contribution levels we find are in line with those typically found in the literature. Our contribution rates are smaller than the ones obtained by Isaac, Walker and Thomas (1984) and larger than the ones obtained by Palfrey and Prisbrey (1996) and Saijo and Nakamura (1995).[16]

All in all, behavior follows a common pattern across countries and treatments and the quantitative differences do not appear to be large. Nevertheless, these differences might be significant. This is tested in the following section.

4. Statistical analysis of behavior across countries and treatments

In this section, we present a more detailed analysis of our results. We will define two indices of behavior that will be used in a statistical analysis of several hypotheses. For this purpose

[15] Sufficiently strong non-monotonic behavior would directly suggest that errors are an important component of behavior.

[16] More detailed information on this comparison is available on request.

we distinguish between behavior for mrs < 1 and behavior for mrs > 1.

To measure deviations from the dominant strategy we define two broad-based measures of behavior. We start with the number of tokens contributed to the private account and define the *Gross Spite Index* (GSI_i) for individual i as the ratio between this number and the number of tokens available for situations with mrs < 1 (180 tokens in 10 rounds in A and 450 tokens in S). The GSI is defined as a percentage. The term 'gross' is used to indicate that, at this point, we do not wish to distinguish between error and spite, or even other motivational forces, as explanations for the lack of contributions when mrs < 1; we do not attempt to attribute specific fractions of the deviations to different factors. Formally, we define the GSI for individual i as:

$$GSI_i = \sum_{t=1}^{10} \sum_{s|rs<m} (9-x_{ist})/\Omega_1 \text{ x } 100\%, \qquad (2)$$

Where $\Omega_1 = 180$ for the AS sessions and $\Omega_1 = 450$ for the S sessions. Note that $GSI_i = 0$, if i plays the dominant strategy for mrs<1 in all periods (*i.e.*, $\forall$ t=1,...,10: $r_s<m \Rightarrow x_{ist}=9$). Finally, we denote the average value of GSI_i, across a group of subjects by dropping the subscript i.

For mrs > 1 (where the dominant strategy is to invest all tokens in the private account) we consider the situations where it is efficient to contribute everything to the public account (mrs < 4).[17] Parallel to the GSI, we define the *Gross Cooperation Index* (GCI) as the percentage of the total number of tokens contributed to the public good for 1 < mrs < 4. We use cooperation here in a general sense without referring to any specific formulation of cooperative preferences. We use the term gross in the same sense as with GSI. Formally, we define the GSI for individual i as:

$$GCI_i = \sum_{t=1}^{10} \sum_{s|m<r_s<4m} x_{ist}/\Omega_2 \text{ x } 100\%, \qquad (3)$$

[17] We do not take into account the third kind of situations for which it is dominant and efficient to contribute nothing to the public account (situations 9 and 10 in A). As shown in section 3, contributions are extremely low in these situations. Moreover, it is not easy to think of a motivation for contributing in those situations, and we know of no experimental evidence that shows that subjects contribute in a case like this.

Where $\Omega_2 = 540$ for the AS sessions and $\Omega_2 = 450$ for the S sessions. Note that $GCI_i = 0$, if i plays the dominant strategy for $1<mrs$ in all periods (*i.e.*, $\forall$ t=1,...,10: $m<r_s<4m \Rightarrow x_{ist}=0$). Finally, we denote the average value of GCI_i, across a group of subjects by dropping the subscript i.

In section 4.1. we discuss the hypotheses to be tested. Section 4.2. presents comparisons of the GSI and the GCI for both the AS and the S sessions. Section 4.3. presents tests for country and subject background effects, as well as for the impact of the partners vs. strangers distinction. Section 4.4. presents a summary view of our data.

4.1 Hypotheses with respect to the GSI and the GCI

According to standard theory, subjects will follow the dominant strategy for all mrs. This implies the prediction that $GSI_i=GCI_i=0$. In experiments, however, one has to allow for the presence of decision error. This introduces the possibility that people deviate from the dominant choices and, hence, that the GCI and the GSI take on positive values. Decision error can be conceived of in different ways. Ledyard (1995), in his review of public goods experiments, refers to subjects making mistakes and choosing their allocations randomly. According to this view, there is no *a priori* reason to expect errors to be larger for $mrs<1$ than for the $mrs>1$. A somewhat different view is the one proposed by McKelvey and Palfrey (1995) and used in the context of public goods experiments by Anderson, Goeree and Holt (1998). They posit that the cost of a mistake affects its likelihood: relatively costly mistakes are less likely. As shown below, our design makes it possible to evaluate both mistakes hypotheses.

The combination of the standard view of purposefully individualistic behavior and purely random errors yields the following null hypothesis:

$$H_0: GCI=GSI,$$

The H_0 pertains separately to each of the cases and treatments we study, i.e. the different countries, the partners/strangers distinction etc.. It posits that the GCI is equal across treatments and the GSI is also equal across treatments. The hypothesis that mistakes are related to costs gives rise to a slight variation of the null. We will return to this in section 4.2..

With respect to alternative hypotheses, we do not just consider the most generic one

stating that there will be a difference between the GCI and the GSI across the different cases. Instead, we formulate a number of specific alternative hypotheses, suggested by previous experimental evidence, and test the null against each of them. We consider two kinds of alternative hypotheses. The first kind refers to the differences between GCI and GSI; the corresponding hypotheses will be denoted as H_{A1A}, H_{A1B} and H_{A1C} below. The hypotheses of the second kind, H_{A2} and $H_{A3,}$ refer to differences in either the GCI or the GSI across countries or treatments.

The first alternative hypothesis we consider is H_{A1A}:

$$H_{A1A}: GCI > GSI.$$

It reflects the notion that cooperation is a stronger motivation than spite and that this difference is not swamped by the presence of error. H_{A1A} is written with reference to average behavior. A more specific version of this alternative hypothesis posits that the above inequality holds for each and every country:

$$H_{A1B}: GCI(JP)>GSI(JP) \text{ and } GCI(NL)>GSI(NL) \text{ and } GCI(US)>GSI(USA) \text{ and } GCI(SP)>GSI(SP).$$

A common observation when analyzing data from public goods experiments is that the cooperation indicators declines as subjects gain experience with the environment; in our case, this applies to the GCI. A similar decay may be present in the GSI, however. Our conjecture is that, in spite of these decays, cooperation remains a stronger motivational force than spite throughout the 10 periods of play. This general notion leads to the formulation of the following hypothesis:

$$H_{A1C}: GCI(\text{period } 1)>GSI(\text{period } 1) \text{ and } GCI(\text{period } 10)>GSI(\text{period } 10).$$

Our two next alternative hypotheses refer to the H_0 in relation to the equality of the GCI and the GSI across countries and treatments. The Saijo and Nakamura (1996) results on observed spite in Japan mentioned in section 1 (and the absence, to date, of similar results for other countries) suggest:

H_{A2}: GSI (JP) > GSI (NL) = GSI (USA) = GSI (SP).

As far as the partners and strangers contrast is concerned, the alternative hypothesis suggested by the Andreoni (1988) results is:

H_{A3}: GCI (strangers) > GCI (partners).

These hypotheses are tested in the following subsections. Subsection 4.2. presents the results of testing hypotheses H_{A1A} , H_{A1B} and H_{A1C} while subsection 4.3. presents results for H_{A2} and H_{A3}.

4.2 Hypotheses tests: comparisons of the GSI and the GCI

We want to conduct all our tests on the basis of statistically independent observations. To ensure this we deal differently with data from partners than from strangers sessions. Recall that in the strangers condition subjects are reallocated to new groups in every period; from period 2 on the reshuffling of the groups contaminates behavior across these groups. In contrast, in the partners condition group observations are independent from each other for all periods. For strangers we, therefore, only use data from period 1 (at both the individual and the group level), while for partners we consider the same kind of data from period 1 as well as group averages over all periods.[18] The use of group data for all cases yields a common perspective on the results.

The GSI and the GCI are the basis of our statistical tests. Table 4 presents an

Table 4: GSI and GCI per country and treatment in the AS sessions*

	GSI JP	GCI JP	GSI NL	GCI NL	GSI SP	GCI SP	GSI US	GCI US	GSI ALL	GCI ALL
Part.	7.8	22.7	4.6	29.0	1.6	17.2	8.2	25.8	5.5	23.7
Stra.	14.8	20.3	7.1	27.5	4.0	15.5	18.3	19.4	11.0	20.7
Avg.	11.3	21.5	5.9	28.3	2.8	16.4	13.3	22.6	8.2	22.2

*The numbers in the cells are percentages. 'part'=partners; 'stra'=strangers, 'SP'=Spain and 'US'=USA.

[18] The data are available from the authors on request.

overview of the values of the average GSI and GCI for all the different conditions studied in the AS sessions; the numbers in this table refer to all rounds.

H_{A1A}: GCI>GSI.

We use the matched pairs of GSI and GCI and the Wilcoxon Signed-Rank test to test the null hypothesis against our first alternative hypothesis, H_{A1A}, that GCI>GSI. We begin with the AS sessions. We can easily reject the null using both partners and strangers group data for period 1 only (p=.000, N=48) and partners data averaged over all periods (p=.000, N=24).[19]

For the S sessions the aggregate GSI is 8.7% in Japan and 9.1% in the Netherlands (8.9% on average) and the GCI is 33.2% in Japan and 38.8% in the Netherlands (36.0% on average). Like for the AS sessions we can use a Wilcoxon Signed-Rank test to compare the GCI and the GSI. Since all the S sessions were based on the partners design we can compare average values using groups as observations. The difference between GSI and GCI is statistically significant when tested at the group level, in spite of only 12 matched pairs of observations ($p<0.01$).

Even these conservative tests reject the null hypothesis presented above, H_0. But how can they be used in relation to a modified null that reflects the notion, suggested by McKelvey and Palfrey (1995) and Anderson, Goeree and Holt (1998), that mistakes depend on the costs of making them? This very influential view of decision error is perhaps more firmly grounded in standard economic thinking than the notion that they are purely random. One of the features of our design makes it possible to evaluate the costly mistakes hypothesis. More precisely we are able to test in a straightforward way whether the behavior in our data can be explained exclusively by the costly mistakes hypothesis. Note that our test does not relate to other aspects of the model presented in Anderson, Goeree and Holt (1998). It considers the costly mistakes hypothesis in isolation and, hence, refers only to errors against the dominant strategy.[20]

For the AS sessions we need to modify somewhat the GCI to be able to make our argument. This can be seen in table 1. Taking, for instance, the payoff schedule for Japan in table 1 one can see that the costs of deviations from the dominant choice are 6, 2, 2 and 6 for situations 1

[19] N denotes sample size.

through 4 and they become larger than 6 from situation 5 on. In order to use aggregate indices that reflect the same range in terms of costs of mistakes we need to restrict the GCI to situations 3 and 4, while we can continue using the same GSI as defined above. On the basis of a Wilcoxon Signed-Rank test we can reject the equality of the GSI and the modified GCI, denoted hereafter by GCI_{34}[21], using both data from period 1 (p=.000, N=48) and averages over all periods for the partners treatment (p=.000, N=24).

For the S sessions we do not need to modify the GCI: for both countries, the costs of deviating from the dominant choices are symmetric around mrs=1. For instance, in Japan the costs of a mistake are 2 yen for both situations 5 and 6, 6 yen for both situations 4 and 7 and they keep on taking the same values as one moves away from mrs=1 in both directions. As a consequence, the GCI and the GSI represent aggregate measures of deviations that correspond to the same set of costs of deviations. Hence, a rejection of the equality of the two indices implies that our data cannot be explained by asymmetric costs of deviation: the costly mistakes hypothesis does not explain our data.

H_{A1B}: GCI(JP)>GSI(JP) and GCI(NL)>GSI(NL) and GCI(US)>GSI(USA) and GCI(SP)>GSI(SP).

Table 5 presents the results of our per country comparisons of GSI and GCI for the AS sessions. Note first that if we center on the results using GCI_{34}, we can in all instances reject equality of the two indices at the 5% level. If we look at the tests based on individual data, shown in rows 3 and 6, then we can reject equality at the 5% level for almost all cases, with the only exception involving a p-value of p=.052.

For the S sessions we can reject equality for both Japan and the Netherlands (p=.028, N=6 for both countries). Taking together the results for the GCI_{34} test in the AS sessions and the results from the S sessions, our judgement is that we can safely reject our H_0 vs. the alternative hypothesis H_{A1B} that the GCI is larger than the GSI in all four countries.

[20] It, therefore, does not speak to the possibility of combining decision error with other deviations from the dominant strategy to explain the data.

Table 5: Comparisons of the GSI and the GCI per country for the AS sessions*

Data	JP	NL	SP	US
GSI vs. GCI Period 1 Group data	p=.031 N=12	p=.028 N=12	p=.002 N=12	p=.136 N=12
GSI vs. GCI All periods Only partners Group data	p=.116 N=6	p=.028 N=6	p=.028 N=6	p=.046 N=6
GSI vs. GCI Period 1 Indiv. Data	p=.003 N=48	P=.001 N=48	P=.000 N=48	P=.052 N=48
GSI vs. GCI34 Period 1 Group data	p=.005 N=12	P=.008 N=12	P=.002 N=12	p=.015 N=12
GSI vs. GCI34 All periods Only partners Group data	p=.046 N=6	p=.028 N=6	p=.028 N=6	p=.028 N=6
GSI vs. GCI34 Period 1 Indiv.data	p=.000 N=48	p=.000 N=48	p=.000 N=48	p=.004 N=48

* The numbers in each cell correspond to the p-value for the Wilcoxon test and to the sample size.

H_{A1C}: GCI(period 1)>GSI(period 1) and GCI(period 10)>GSI(period 10).

We now want to analyze whether the difference between the GCI and the GSI

[21] Formally, $GCI_{i34} = \sum_{t=1}^{10} \sum_{s=3,4} x_{ist}/180 \times 100\%$.

persists throughout the ten periods of play; in addition we want to explore whether the GCI is still substantial in the last period of play. The use of the two indices we have defined makes it possible to present the evolution of behavior over time in a simple way. The four panels of figure 3 (a-d) present the development of the GSI and the GCI over time, for both the AS and the S sessions; the averages are taken over all observations.

(Figure 3 about here)

For the AS sessions the GCI clearly declines in all four countries, whereas the change in the GSI can not be described as easily. If, however, one computes averages over countries one finds that the GCI declines from 30% to 16% and the GSI from 13% to 7%. For the S sessions the GSI appears to be constant in time and the GCI exhibits some decline in later periods. Again, the GCI appears to remain larger than the GSI throughout the experiment.

Our tests confirm the above impressions. Since, overall, behavior is similar across countries, we only report results from aggregate tests. Wilcoxon Signed-Rank tests using group data only from the partners sessions find that the differences between GSI and GCI are highly significant in all relevant instances: AS sessions in period 1 ($p=.000$, $N=24$) AS sessions in period 10 ($p=.000$, $N=24$), S sessions in period 1 ($p=.002$, $N=12$) and S sessions in period 10 ($p=.004$, $N=12$).[22] We can, therefore, reject the null hypothesis in favor of H_{A1C}; learning does not lead to behavior that can be explained purely in terms of decision error.

4.3. Hypotheses tests: country and treatment effects

We now move to the two hypotheses that refer to country and treatment effects, H_{A2} and H_{A3}. These can be tested straightforwardly using regression analysis. Table 6 presents evidence about the effects of different variables on the GCI and the GSI, based on the AS sessions; the results

[22] A more complete analysis of behavior over time is beyond the scope of this paper. However, the decline of the GCI over time and the fact that it still has a sizeable value in period 10, together with the absence of a difference for partners and strangers (which will be discussed in section 4.3.), suggest the following simple interpretation of the behavior we observe. Subjects wish to cooperate conditional on others also cooperating. They start out contributing considerable amounts, but are disappointed by the cooperation levels of others. They, therefore, revise their cooperation levels downwards and this goes on period after period. See Brandts and Schram (2001).

from the S sessions will be evaluated later. We present results from linear regressions with GCI and GSI as dependent variables; due to the differences in subject pool composition shown in table 3 we not only consider the effects of country and the partners vs. strangers distinction, but also include variables to capture the possible impact of age, gender and study major. As in the previous section we use period 1 individual and group observations for both treatments together and group observations averaged over all periods for partners only.

Table 6: Effects of country, treatment and individual variables on GCI and GSI; AS sessions.*

Dependent Variable	Const.	Age	Econ	Gender	Stranger	JP	NL	US
GCI Period 1 N=48	.960 (2.44) (p=.02)	-.027 (1.52) (p=.14)	-.169 (1.62) (p=.11)	.065 (.68) (p=.5)	-.022 (.53) (p=.6)	-.074 (.74) (p=.46)	.121 (1.51) (p=.14)	-.027 (.37) (p=.72)
GSI Period 1 N=48	-.443 (1.22) (p=.23)	.021 (1.28) (p=.21)	.059 (.61) (p=.55)	-.048 (.54) (p=.59)	.055 (1.45) (p=.16)	.141 (1.52) (p=.14)	.017 (.23) (p=.81)	.091 (1.34) (p=.19)
GCI All periods Only partners N=24	.546 (.82) (p=.43)	-.11 (.36) (p=.72)	-.166 (.93) (p=.37)	-.028 (.16) (p=.88)		-.086 (.48) (p=.64)	.109 (.84) (p=.42)	.079 (.72) (.48)
GSI All periods Only partners N=24	-.514 (1.36) (p=.19)	.021 (1.24) (p=.23)	.076 (.75) (p=.46)	.169 (1.69) (p=.11)		.137 (1.36) (p=.19)	-.024 (.33) (p=.75)	.000 (.01) (p=.99)
GCI Period 1 N=163[23]	0.611 (3.14) (p=.00)	-.012 (1.31) (p=.19)	-.055 (1.03) (p=0.31)	-.093 (1.95) (p=.05)	-.024 (.57) (p=.57)	-.082 (1.13) (p=.26)	0.091 (1.46) (p=.15)	-.018 (0.27) (p=.79)

[23] The number of observations is less than 192, the total number of participants, due to some missing values on background variables.

GSI	-.126	.010	-.046	-.036	.042	.040	.009	.101
Period 1	(0.77)	(1.32)	(1.03)	(.90)	(1.20)	(0.654)	(.17)	(1.84)
N=163	(p=.44)	(p=.19)	(p=.31)	(p=.37)	(p=.23)	(p=.51)	(p=.87)	(p=.07)

* For the regressions with group data the independent variables are defined as follows: Const.=Regression constant, Age=Average age in the group, Econ=Fraction of economics or econometrics students in the group, Gender=Fraction of women in the group, Stranger=Dummy equal to 1 in case of strangers, JP=Dummy equal to 1 in case of Japan, NL=Dummy equal to 1 in case of Netherlands amd US=Dummy equal to 1 in case of USA.

The first row in table 6 shows result of regressing the average GCI per group in period 1 on average age in the group, the fraction of economics/econometrics students in the group and the fraction of women in the group as well as on a dummy for the strangers treatment and country dummies. The second row shows the same for the GSI per group in period 1 as dependent variable. These regressions are based on 4 (countries) x 4 (sessions) x 3 (groups) = 48 observations. The results of the regressions using only group data from the partners sessions are shown in the third and fourth row of table 6. Finally, rows 5 and 6 show regressions based on individual data from period 1. In each cell, the first number is the coefficient; below it one can first see the absolute t-value and then the p-value.

These results allow us to test H_0 against H_{A2} and H_{A3}; they also provide additional information on some other issues of interest. For all six regressions shown in table 6 only few of the independent variables has a significant effect at conventional significant levels. We can not reject our null hypothesis in favor of either H_{A2} or H_{A3}. The GSI is not larger in Japan than in other countries. In fact, the results show no significant effects of the country variables on the GCI at conventional levels. We find no difference between partners and strangers, possibly an indication that strategic behavior is not prominent in our data. In addition, we also find no effect of the country variables on the GCI. With respect to the subject background variables, the only significant effect is the small negative impact that being female has on the GCI in the regression based on individual data shown in row 5. Overall, observed behavior appears to be remarkably similar across the different categories.

4.4. A final view: step function representations of our aggregated contribution functions.

The analysis of the previous sections has shown that differences of behavior across countries are small. Our data also exhibit a lack of systematic difference between the partners and the strangers treatments. We, therefore, think that it is safe to look at aggregate data without losing

too much information in the process. Figure 4 (a and b) presents the aggregate contribution functions over all sessions, treatments, and periods for the AS and S sessions. It also shows the step functions with steplevels at those integer levels that are closest to the values of the aggregate contribution functions at each mrs.

(Figure 4 about here)

Figure 4 reflects the behavior of a total of 240 subjects: 192 subjects for the AS sessions and 48 subjects for the S sessions, where each of these subjects specified 10 complete contribution functions.

Note that for both the AS and the S sessions the estimated step function is within 1 token of the dominant strategy prediction for all cases where $mrs < 1$ and $mrs > 4$, i.e., in those cases where the dominant strategy coincides with efficiency. The deviation from the dominant strategy is considerably larger for $1 < mrs < 3$, however. The strong asymmetry of the step function around mrs=1 in the S sessions makes it particularly clear that the costly mistakes hypothesis is not sufficient to explain our data.

The comparison between figures 3a and 3b reveals the robustness of behavior across our two designs. Note that the four mrs values in the AS sessions that are comparable with those in the S sessions (mrs = 0.25, 0.75, 1.25, 1.75) show contribution levels which would fit very nicely in the contribution function for S. Hence, the different situations of tables 2 and 3 do not appear to elicit different behavior for comparable mrs values.

5. Summary and conclusions

There are two central features of our data that we wish to highlight in this final section. First, there appear to be only minor differences in behavior between the four countries we investigate. The broad behavioral patterns are common across countries and any quantitative differences are small. Our results give support to the notion that, when people from different countries play “the same game”, they behave similarly. This does not necessarily contradict the general impression that people in different countries do behave differently. Observations of this kind, however, may just reflect the fact that different institutional environments lead to different behavior,

i.e. people in different countries may just be playing according to different rules.[24] Informal observations about alleged cultural differences stem from very uncontrolled environments and, moreover, may be influenced by the observer´s prejudices and biases. In contrast, experimental analysis makes it possible to have different subject pools play literally the same game and to uncover similarities in behavior that are difficult to see with field observations.

A specific feature of our cross-country comparison is that we find less evidence of spite than in Saijo and Nakamura (1995). How can this difference be explained? Their results come from a linear vcm environment related to ours. At this point, our explanation of the discrepancies between the two data sets are the differences in the presentation of the payoff information. Both their rough and their detailed payoff tables differ from the presentation commonly used in this kind of experiments. In contrast, our presentation is close to the ones used in related work. In addition, observe that one of the features of our design is precisely that it allows for a direct cross-country comparison of spiteful behavior.

The second result that we wish to stress is that deviations from the standard prediction are larger for mrs>1 than for mrs<1; in all four countries the average GCI is larger than the GSI and this difference does not disappear over time. Models based exclusively on decision error can not accommodate our evidence. Some kind of cooperative motive is needed to explain the data. If we attribute part of the GSI to the presence of spite then our data allow us to make the following statement: for all four countries spite is not as strong a motivation as cooperation and this difference is large enough not to be overwhelmed by the presence of error. Our data are consistent with a universal cooperative motivation.[25]

The absence of cross-country differences may surprise some readers, since national 'cultures' appear to be such strong forces in shaping human behavior. Our purpose in this paper, however, was precisely to use our design to submit this common sense notion to a rigorous test. Naturally, our results are far from a last word on the matter; a general picture of the impact of country variables on behavior in experiments will only emerge after a substantial body of research will have been completed. We do, however, believe that our design has allowed us to study behavior

[24] For an anthropologist's view of the universal basis of human behavior see Brown (1991).

[25] Combining decision error with a cooperative motive may be a good way to accommodate data from public goods games. Palfrey and Prisbrey (1997) and Anderson, Goeree and Holt (1998) present statistical analyses in this line on the basis of warm glow and altruism as cooperative motives.

in a systematic manner and to generate a rich data set of decisions that needs to be taken into account in the debate. We also believe that it is important that studies that do not find cross-country differences obtain just as much coverage as those that do find them. The last thing the literature needs is a biased sample of experimental results. Before we can even hope to explain differences in behavior across countries we need to get the facts straight. It is this goal that this paper wishes to contribute to.

Journal of Public Economics, 70, 297-323.

Andreoni, J. (1988): "Why free ride? Strategies and learning in public goods experiments", Journal of Public Economics, 37, 291-304.

Brandts, J. and A. Schram (2001): "Cooperation and noise in public goods experiments: applying the contribution function approach", Journal of Public Economics, 79, 399-427.

Brown, D.E. (1991): Human Universals, New York, MacGraw-Hill.

Burlando, R. and J.D. Hey (1997): "Do Anglo-Saxons free-ride more?", Journal of Public Economics, 64, 41-60.

Cubbitt, R. P., C. Starmer and R. Sugden (1998): "On the Validity of the Random Lottery Incentive Experimental Economics, 1, 115-131.

Hofstede, G. (1991): Cultures and Organization. Software of the Mind, McGraw-Hill, London.

Isaac, R.M., J. Walker and S. Thomas (1984): "Divergent evidence on free riding: An experimental Public Choice 43, 1, 113-149.

Kachelmeier S., and M. Shehata (1990): "Culture and competition: a laboratory market comparison Journal of Economic Behavior and Organization 19, 145-168.

Keser, C. and F. van Winden (1997): "Partners contribute more to public goods than strangers", Tinbergen Institute discussion paper 97-018/1, University of Amsterdam.

Ledyard, J. (1995): "Public goods: a survey of experimental research", in The Handbook of Experimental Economics, by J. Kagel and A. Roth (eds.), Princeton University Press, 111-194.

McKelvey, R. and T. Palfrey (1995): "Quantal Response Equilibria in Normal Form Games", Games

and Economics Behavior, 7, 6-38.

Palfrey, T.R. and J.E. Prisbrey (1996): "Altruism, reputation and noise in linear public goods Journal of Public Economics 61, 409-427.

Palfrey, T.R. and J.E. Prisbrey (1997): "Anomalous behavior in linear public goods experiments: how American Economic Review, 87, 829-846.

Roth, A.E., V. Prasnikar, M. Okuwo-Fujiwara, and S. Zamir (1991): "Bargaining and market behavior in Jerusalem, Ljubiljana, Pittsburgh and Tokyo: an experimental study", American Economic Review 81, 1068-1095.

Saijo, T. and H. Nakamura (1995): "The 'spite' dilemma in voluntary contribution mechanism Journal of Conflict Resolution 39, 1, 535-560.

Saijo T., T. Yamato, K. Yokotani, and T. Cason (1997): "Voluntary Participation Game Experiments with a Non-Excludable Public Good: Is Spitefulness a Source of Cooperation?", mimeo, ISER, Osaka University.

Schram, A. (2000): "Sorting Out the Seeking: Rents and Individual Motivation", Public Choice, 103, 231-258.

Weiman, J. (1994): "Individual behavior in a free-riding experiment", Journal of Public Economics, 54, 185-200.

Figure 1a: Japan

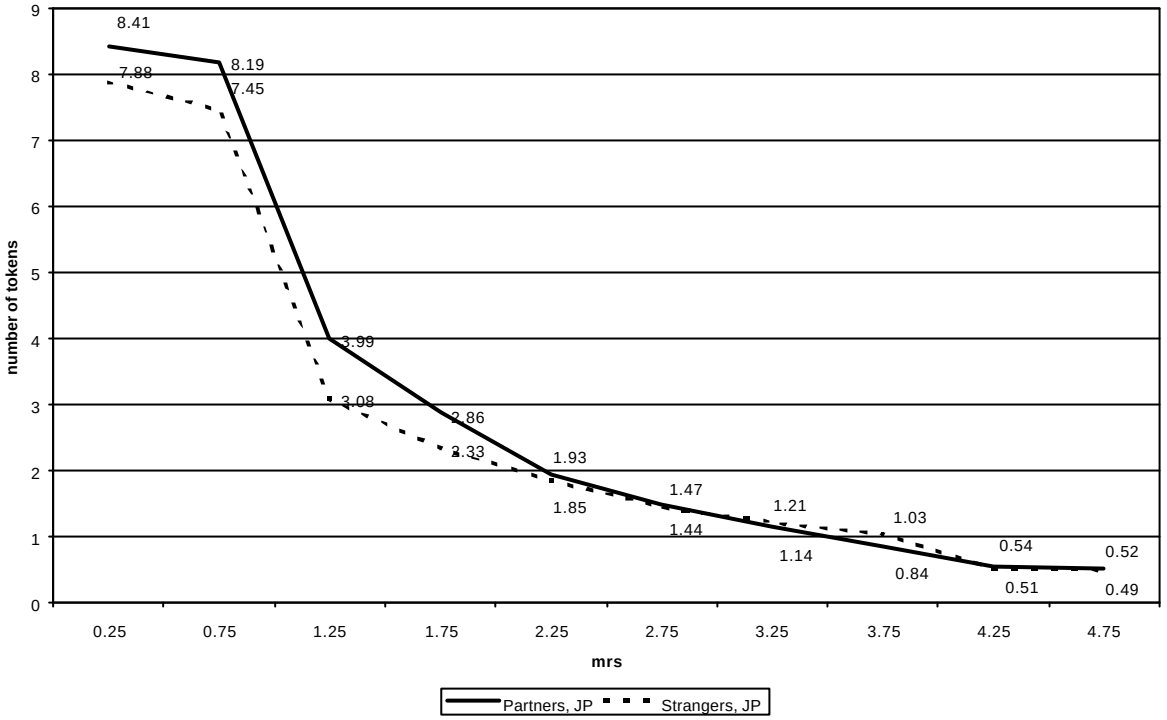


Figure 1b: Netherlands

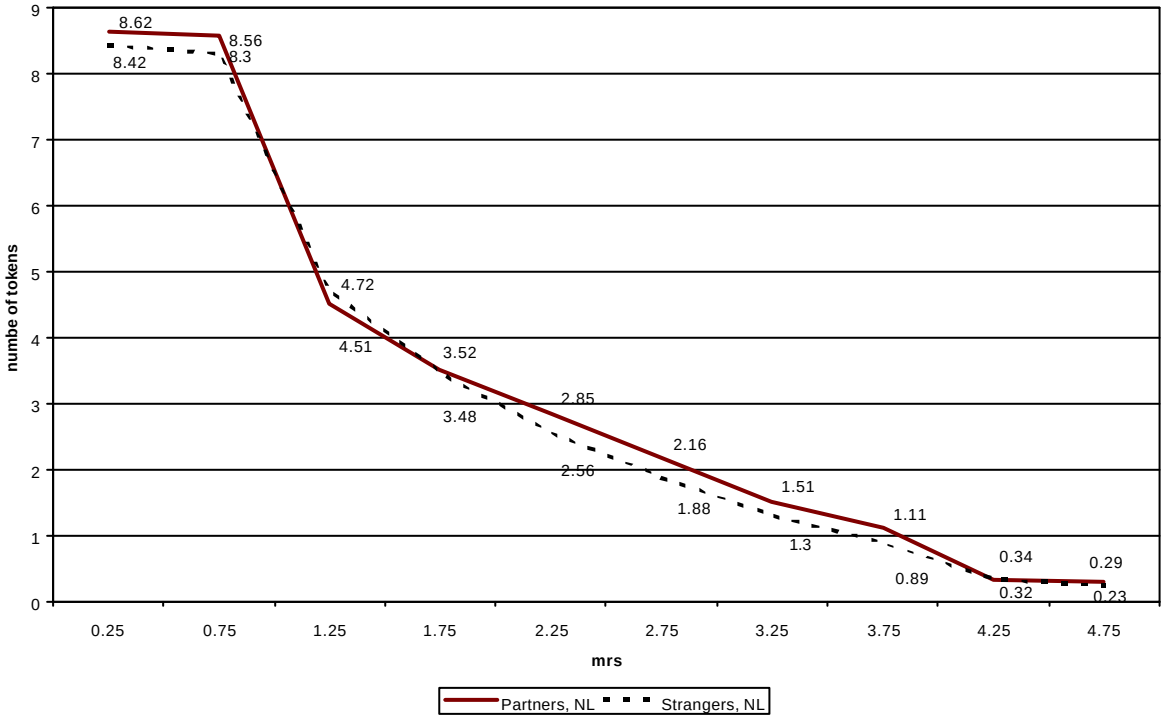


Figure 1c: Spain

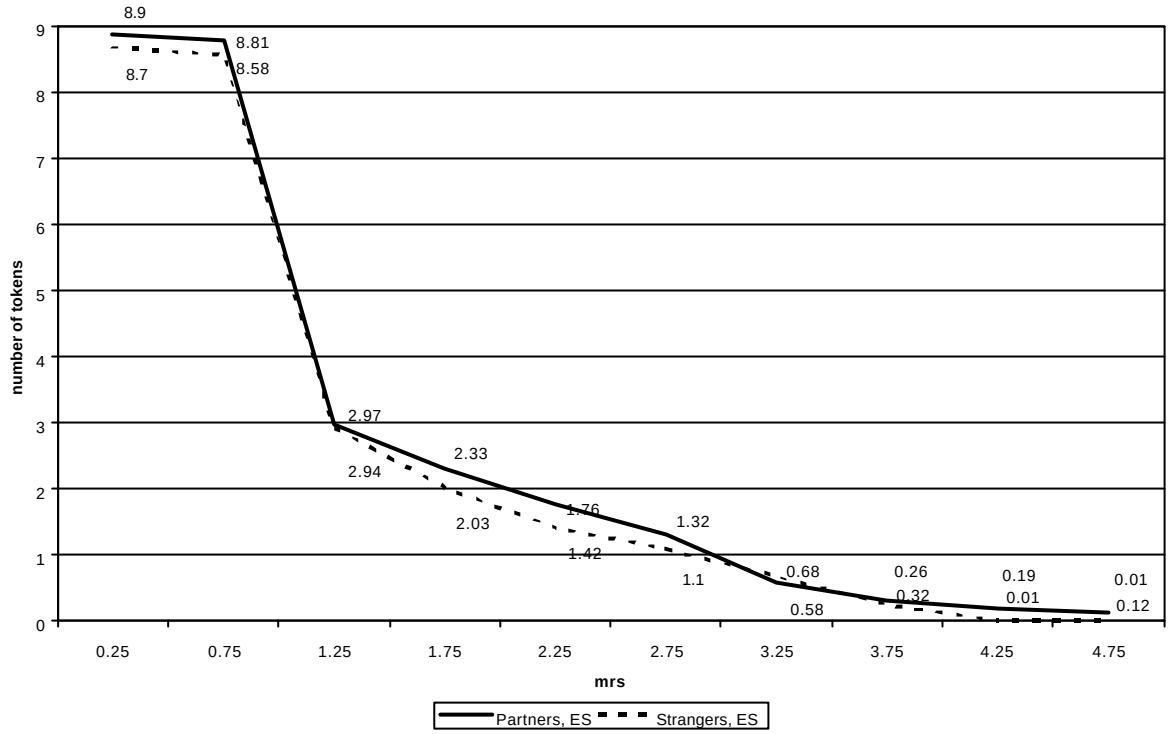


Figure 1d: United States

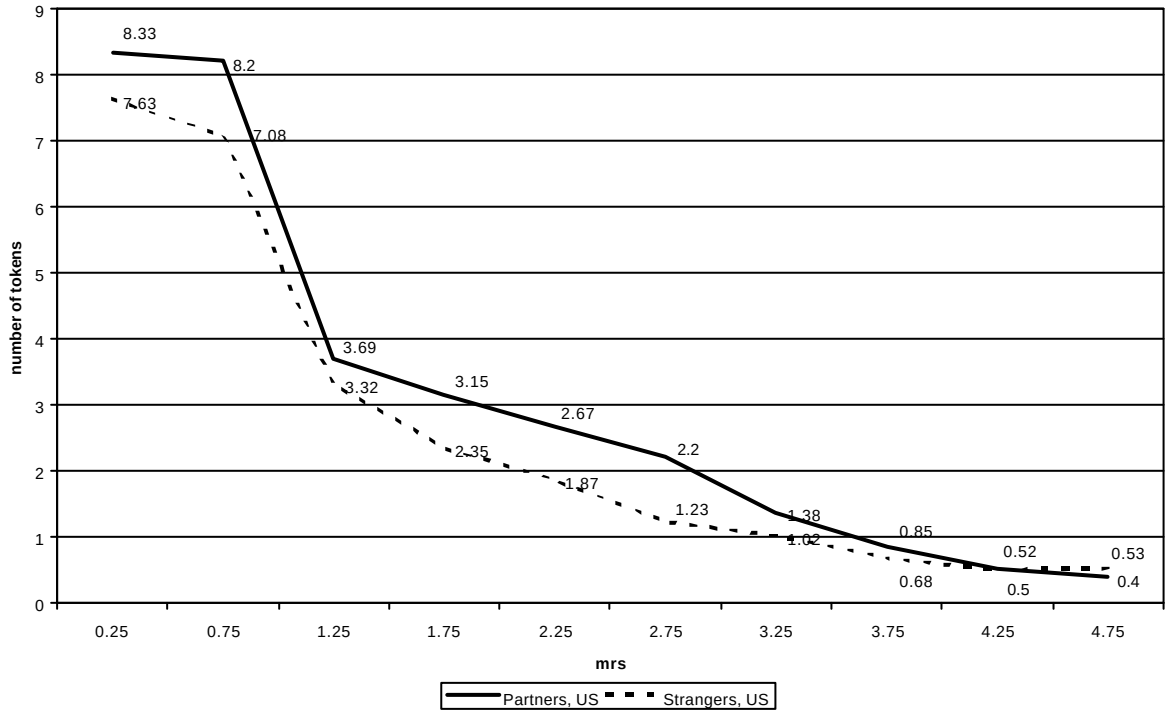


Figure 2: Symmetric Sessions

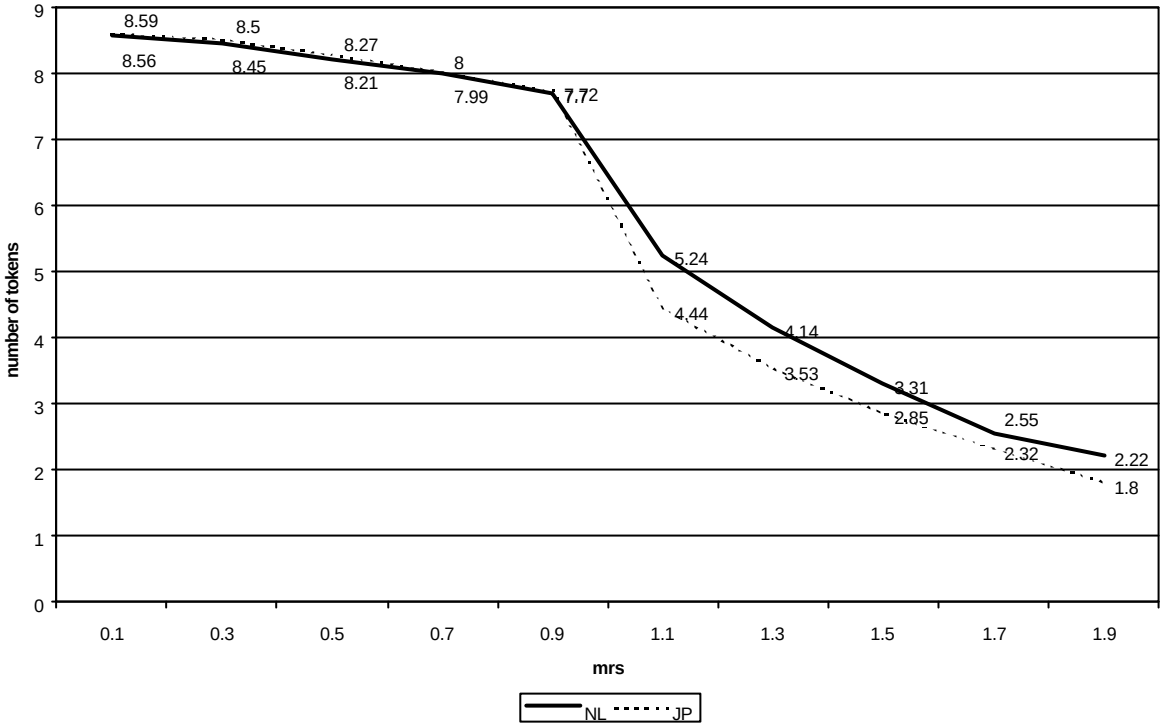


Figure 3a: GCI per country over time; AS sessions

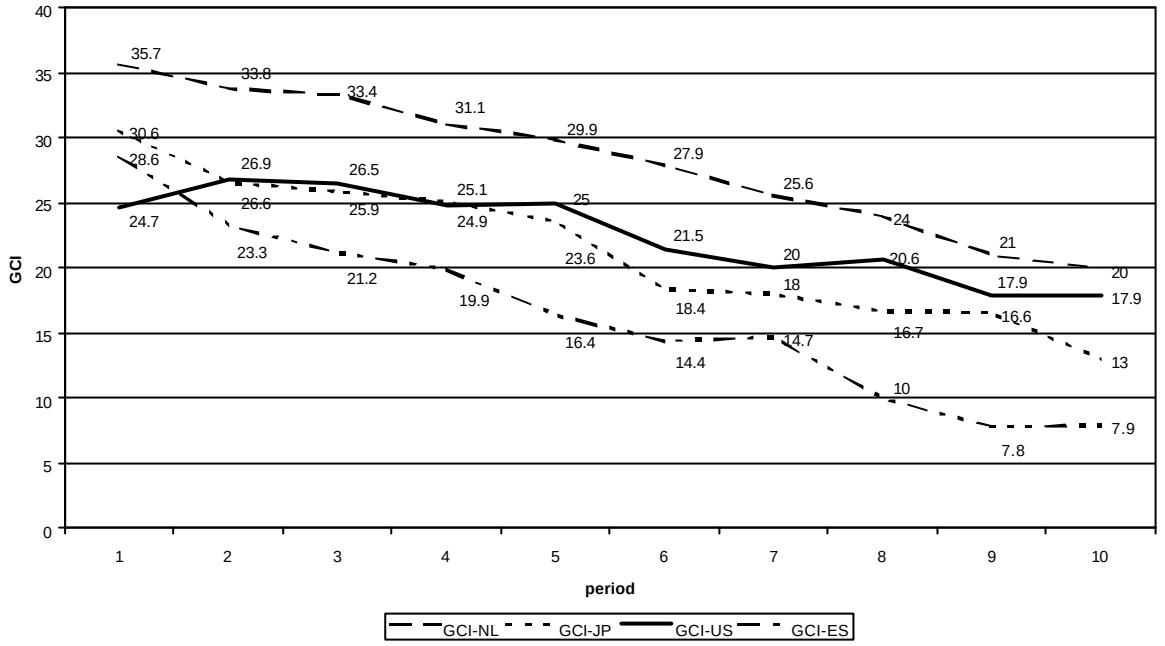


Figure 3b: GSI per country over time; AS sessions

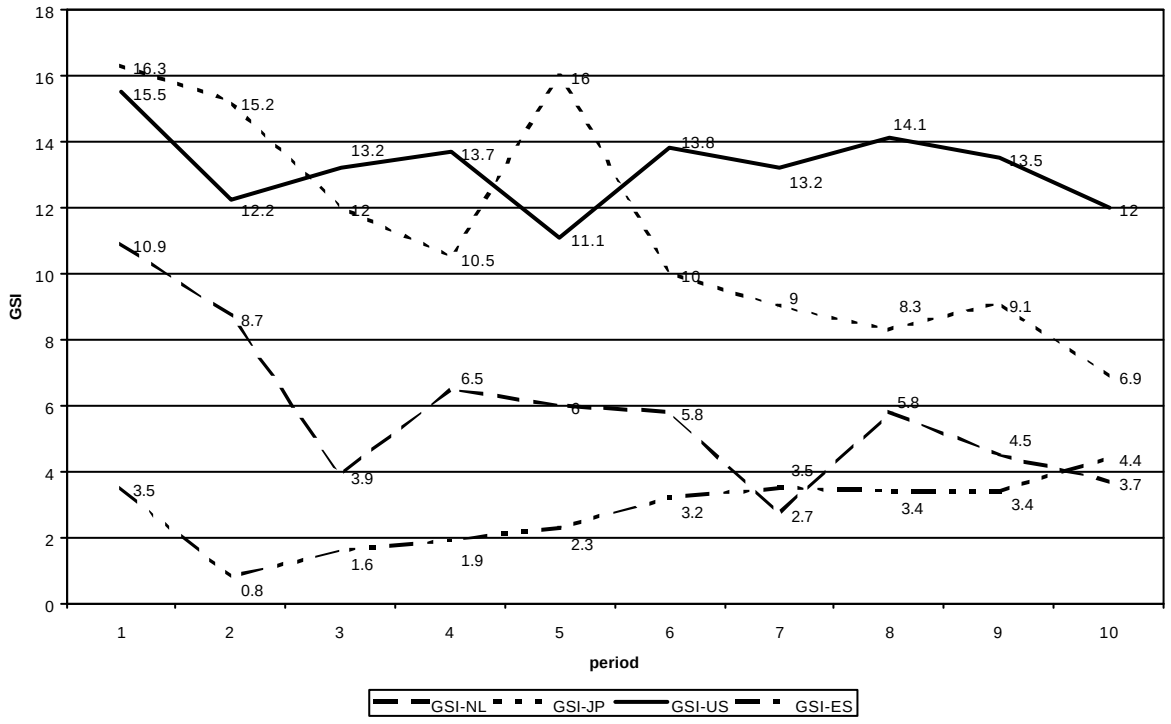


Figure 3c: GCI per country over time; S-sessions

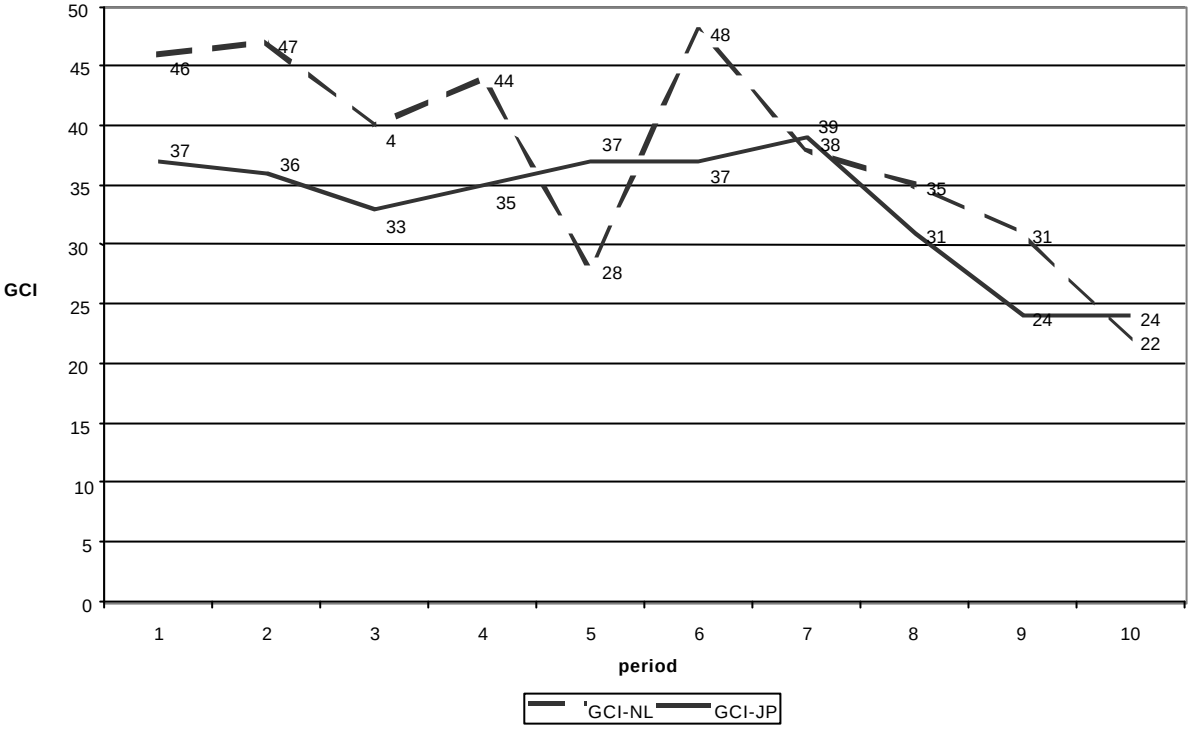


Figure 3d: GSI per country over time; S-sessions

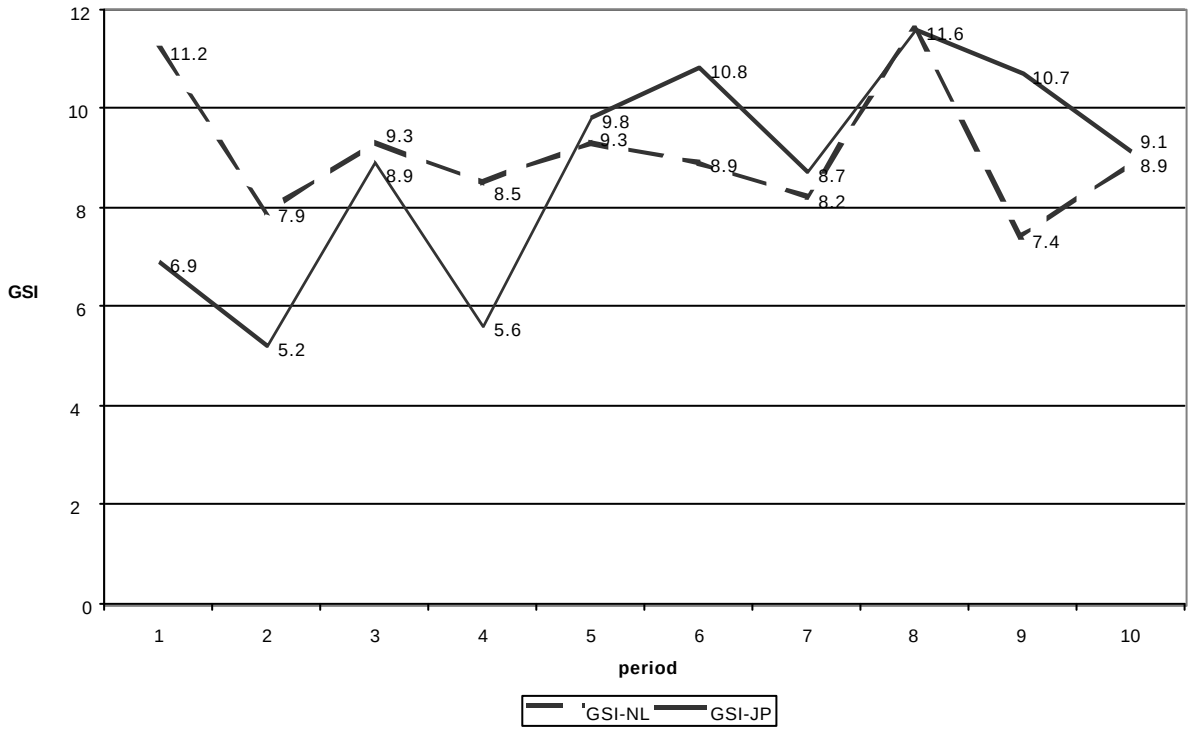


Figure 4a: Aggregate behavior; AS sessions

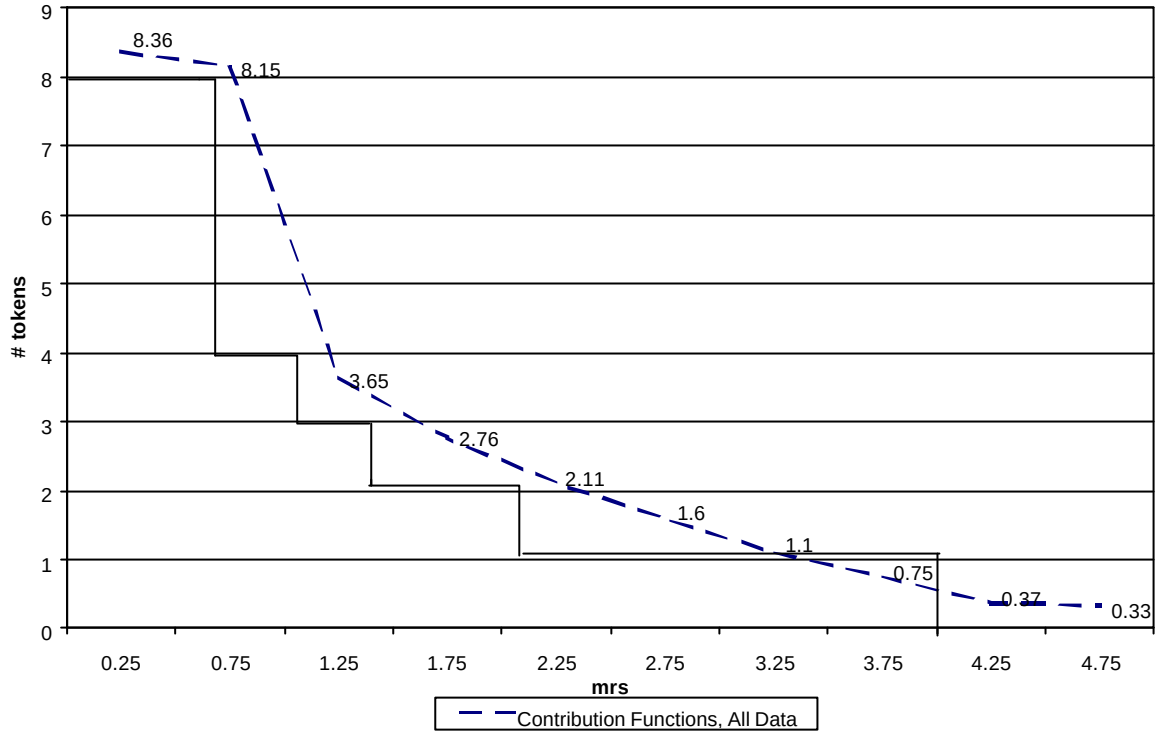
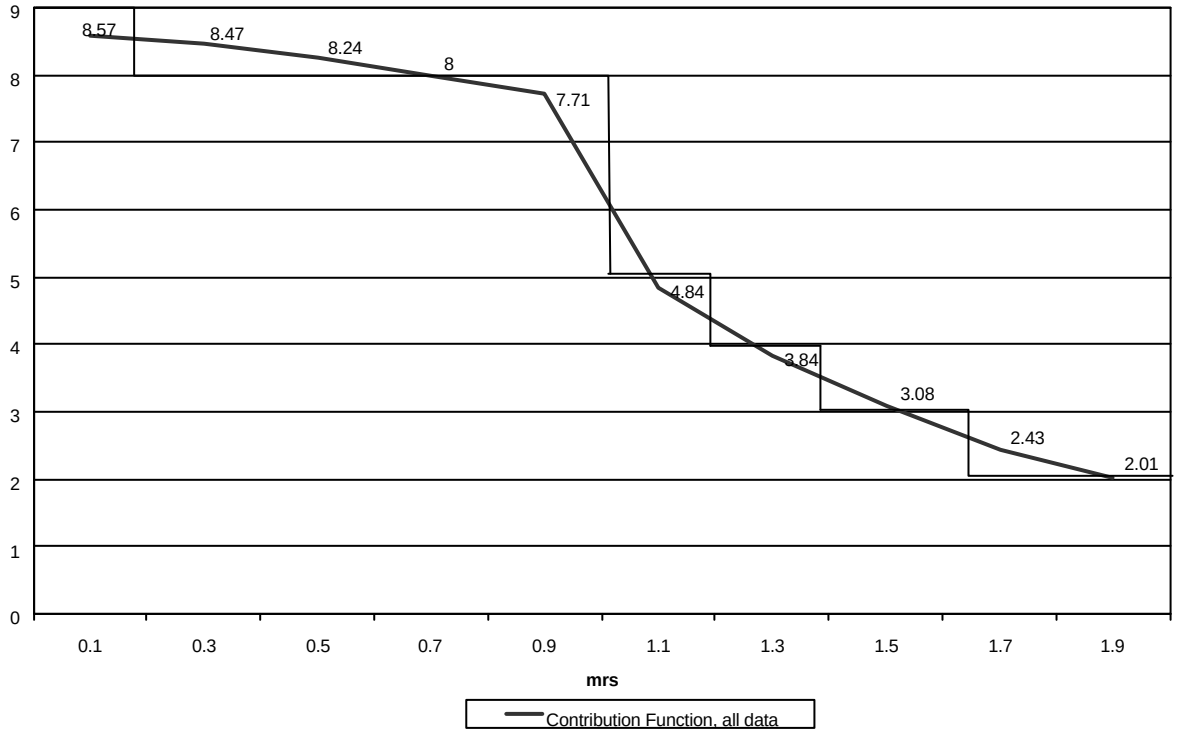


Figure 4b: Aggregate behavior; S sessions



Appendix

This appendix presents the English version of the computerized instructions. Computer pages are separated by lines. In some cases, the instructions included an example screen from the program. In this appendix this is denoted by 'SCREEN[x]'. These screens are given at the end of the instructions. The instructions presented here are for the partners sessions.

INTRODUCTION

You are about to participate in an experiment about decision-making. The money for this study has been provided by various institutions. The instructions are simple and if you follow them carefully, you will be able to earn a considerable amount of money. All the money you earn during this experiment will be yours to keep. It will be paid to you personally in cash at the end of today's session. In addition to the money you will earn, you have already received a show-up fee of 500 ptas [5 guilder] when you came in.

Now press F2 to continue.

INTRODUCTION

Please take your time to read these instructions at your own pace. If you have any questions while reading them, please raise your hand and someone will come to your table. Before the experiment, two practice rounds will be played, which will not be paid.

Throughout these instructions, you may return to a previous page by pressing F1 and proceed to the next page by pressing F2.

F1=previous F2=next

PERIODS AND GROUPS

The experiment consists of 10 separate periods. In each period, you are in a group with three other participants. The other participants in your group remain the same in all 10 periods. You will not know which of the other participants is in your group. The group composition is secret for every participant.

In each period, you and the other participants in your group have to make decisions. The amount of money you earn depends on your own decisions and on the decisions of the three other members in your group.

The monitor has a separate role. This will be explained later in the instructions.

F1=previous F2=next

DECISIONS: TOKENS

In each period of the experiment, you will have 9 tokens. You must invest each token either in ACCOUNT A or in ACCOUNT B. The amount of money you earn depends on your division of the tokens, on the division by other group members and on chance.

F1=previous F2=next

DECISIONS: ACCOUNT A

Each token you invest in account A yields 8 pesetas [12 cents] to YOU as well as to EACH MEMBER OF YOUR GROUP. You (and everyone else in your group) also receive 8 pesetas [12 cents] for each token invested in account A by other group members.

F1=previous F2=next

DECISIONS: ACCOUNT B

Your investment in account B yields money for YOU ALONE. Other group members do not receive anything for your investment in account B.

The amount of money you receive per token invested in account B differs per 'situation'. There are 10 different situations that may occur. Which situation holds for you will be determined by chance. This is explained below.

F1=previous F2=next

DECISIONS: 10 SITUATIONS

The amount of money you earn per token in account B is different for each of the 10 situations. We call this amount the B-VALUE for the situation. Note that the amount related to account A is always 8 pesetas [12 cents] per token.

At the beginning of each period you will have to tell us how you want to divide your tokens between the two accounts for each of the 10 situations.

F1=previous F2=next

YOUR DECISIONS ON THE COMPUTER

SCREEN[1]

To report your decisions, you will be able to use a decision screen like this one. Note that the 10 situations are represented by rows on the screen.

F1=previous F2=next

INFORMATION ON THE SCREEN

SCREEN[1]

The first two columns (on the left half of your screen) show the payoff per token in accounts A and B. The amount is 8 pesetas [12 cents] PER GROUPMEMBER for every token invested in A.

In Situation 1, the payoff (B-value) is 2 pesetas [3 cents] FOR YOU ALONE for every token you invest in B.
For Situation 2 the B-value is 6 pesetas [9 cents],
etc.

F1=previous F2=next

INFORMATION ON THE SCREEN

SCREEN[1]

For Situation 10, the B-value is 38 pesetas [57 cents].

These numbers will be given on your screen, handed out on paper, and projected on the overhead-screen.

F1=previous F2=next

DECIDING FOR A SITUATION

SCREEN[2]

The right half of the screen is used for your decisions. Using the ARROW-keys (and then ENTER) you can indicate for which situation you want to enter your decision. If you want to change your decision, you may always re-choose the situation concerned later.

Now choose a situation and press ENTER.

Then you may continue with the instructions by pressing F2.

ARROWS: choose; press ENTER.

DIVIDING YOUR TOKENS

When you are going to enter your decision for a situation, you will be asked to divide 9 tokens over A and B.
Shortly, you will be given the opportunity to practice giving a division of the tokens. First we will go through the various steps.

F1=2 pages back F2=next

DIVIDING YOUR TOKENS

First you choose A or B using the ARROW keys. Then you type the number of tokens you wish to invest in that account. When you are finished, press ENTER. The computer will automatically invest the rest of your tokens in the other account. If you want to change your choice, simply choose A or B and enter a new number.

F1=previous F2=next

DIVIDING YOUR TOKENS

Press ENTER again to enter your choice for this situation. Remember that you may return to this situation to change your decision later.

On the next screen you may practice dividing your tokens for situation 2. This is just to practice using the screen, your choice will not be registered or used in any way.

F1=previous F2=next

DIVIDING YOUR TOKENS

SCREEN[3]

Now try entering an investment decision. First choose A or B (using the arrows). Enter a number between 0 and 9. Press ENTER. Confirm by pressing ENTER again.

F1=previous F2=next

YOUR CHOICE APPEARS ON THE SCREEN

SCREEN[4]

Note that your choice is given in yellow on the right half of your screen. In parentheses, we multiply your investment in tokens with the amount per token, as given on the left half of your screen. We only do this to help you with your calculations.

The amount in parentheses is not equal to your total earnings, for two reasons. First of all, you also earn money from investments in account A by other group members. Second, only one situation is actually played. How this is done will be explained shortly.

F1=dividing your tokens F2=next

CONFIRMING YOUR CHOICES

After you are satisfied with your decision for all situations, you must confirm your choices by pressing F10. You will be asked to finalize these decisions. After that, it is no longer possible to change your decisions for that period.

You may have to wait a little while for other participants to finish deciding. When everybody has finalized their decisions, we will come by and select the situation that will be played. We will now explain how this is done.
F1=previous F2=next

A SITUATION IS SELECTED

After everyone has divided 9 tokens for every possible situation one situation will be chosen. The same situation will be selected for each member of a group.

The monitor will throw a 10-sided die (with sides '0' through '9') separately for each group. The result for your group determines the situation chosen for your group, where a '0' is interpreted as '10'.

You are not permitted to tell anyone about your decisions. We will enter the situation selected in your computer to allow the central computer to determine the outcome for the period.

F1=previous F2=next

YOUR EARNINGS FOR THE PERIOD FROM ACCOUNT A

For each group the separate throw of the die will determine which of the divisions of tokens will be used to calculate earnings. You

will see that the situation selected will be highlighted on your screen. The situation selected determines the ACTUAL amounts of tokens invested in accounts A and B by each of the participants.

The computer will use this information first to determine the total number of tokens invested in A by members of your group. Multiply this number by 8 pesetas [12 cents] to determine your earnings from account A.

F1=previous F2=next

YOUR EARNINGS FOR THE PERIOD FROM ACCOUNT B

Your earnings from account B are determined by considering the situation selected for you (with a corresponding B-VALUE) and the number of tokens you invested in B for that situation. Multiplying your investment by the B-VALUE gives your earnings from account B.
Your total earnings in a period are equal to your earnings from account A PLUS your earnings from account B.

F1=previous F2=next

THE REGISTRATION WINDOW

SCREEN[5]

The computer will do all calculations and registration for you. On the left half of your screen you may see how the latter is done. At all time in the experiment, you may choose to have this information on your screen.
If you press the key 'F1' later, the information will alternatively appear in a small (only a few periods), large (all previous periods) or no window.

You cannot try this now. Later in the practice rounds you may try the options concerning the presentation of the registration.
F1=previous F2=next

CALCULATIONS AND REGISTRATION

SCREEN[5]

Consider the information given in the registration window. For each period, you first see the period number and the situation selected for you in that period.
Next the table shows your division of 9 tokens for that situation.
The fifth column gives the total number of tokens invested in A by members of your group.

The last column gives your earnings for that period. The computer determines this by multiplying the number in column 5 by 8 and adding the earnings from B. The earnings from B are determined by your investment (column 4) and the B-VALUE in the situation of column 2. Your total earnings to date are given at the right bottom corner of the registration table and later on the top right corner of your screen. If you have any questions about your earnings at any time, please raise your hand.

F1=previous F2=next

NEXT PERIOD

After the earnings have been calculated and registered for everyone, we will proceed to the next period.
This will continue for 10 periods.

F1=previous F2=next

RECALLING PREVIOUS CHOICES

When making your decisions, you may want to recall your choices from the previous round. By pressing F2, you will see these previous choices on a window like the one on the right. The numbers in the window now have no particular meaning. We just want to show you what the window looks like.

While this window is on your screen, you cannot enter choices. You must first press F2 again to cancel the window. Then you can enter your choices.

You may try this later, in practice round 2, by recalling your choices from practice round 1.

F1=previous F2=next

PAYMENT

At the end of today's session we will pay you in cash the amount of money that you will have earned in 10 periods. In addition, you have already received 500 pesetas [5 guilder] cash for your participation. Your earnings are private information. You do not need to tell them to anyone.

We stress again, that you are not allowed to talk or communicate with other participants. If you have a question, just raise your hand. We will come to your table.

F1=previous F2=next

MONITOR

The monitor has two tasks. First of all, (s)he will help us with the random selection of a situation. We will explain how this is done when it is time to select a situation.

Besides this, the monitor may check that everything is taking place as described in the instructions. To help her or him with this, we will give the monitor an overview of the group composition at the beginning of the experiment. With this, (s)he may check whether the reported investment in account A is correct, for example.

The monitor is not allowed to talk or communicate directly with other participants. If the monitor believes that something is not taking place in accordance with the instructions (s)he must report to the experimenter.

The monitor's earnings do not depend on the decisions made today. (S)he will receive the average payment of the previous experiment.

F1=previous F2=next

PRACTICE ROUNDS

When everyone has finished with the instructions, we will play two practice rounds. In these practice rounds, you will not be playing in groups. The computer will make up decisions for other members randomly. Therefore, the practice rounds cannot give you information about what you may expect from other participants.
These practice rounds will not be paid.

F1=previous F2=next

RE-READ

This brings you to the end of the instructions. If you would like to reread something, you may choose one of the keys F1-F9. Choose F2 to indicate that you are finished with the instructions. With F10 you will return to this menu.

F1 = previous page
F2 = FINISHED WITH THE INSTRUCTIONS
F3 = Introduction
F4 = Instructions about accounts A en B
F5 = Instructions about reporting decisions
F6 = Instructions about selecting a situation and determining
your earnings
F7 = Instructions about the registration of the results
F8 = Instructions about recalling previous decisions
F9 = Instructions about the monitor

choose from: F1 F2 F3 F4 F5 F6 F7 F8 F9

FINISHED

You are now finished with the instructions. When everyone is finished, we will begin with the practice rounds.

You may still reread parts of the instructions. When everyone has indicated that they are finished, we shall stop the instructions and start the practice rounds, however.

PLEASE WAIT QUIETLY UNTIL EVERYONE IS FINISHED WITH THE INSTRUCTIONS
choose from: F1 F2 F3 F4 F5 F6 F7 F8 F9