Discussion Paper No. 496

A FOUR COUNTRY COMPARISON OF SPITE, COOPERATION AND ERRORS IN VOLUNTARY CONTRIBUTION MECHANISMS

Jordi Brandts
Tatsuyoshi Saijo
and
Arthur Schram

January 2000

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

A Four Country Comparison of Spite, Cooperation and Errors in Voluntary Contribution Mechanisms

by

Jordi Brandts*

Tatsuyoshi Saijo**

Arthur Schram***

June 29, 1999

* Instituto de Analisis Economico (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain and Department of Economics, University of California, Berkeley CA 94720, USA. 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

*** CREED, Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. Email: ArthurS@fee.uva.nl

Abstract

This paper presents data from 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 more complete picture of subjects' behavior than previous studies; it yields information about situations where 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 not large. Second, the evidence for spiteful behavior by Japanese subjects, that has been observed in other studies, is not confirmed by our results. Third, as a whole our data are inconsistent with the explanation that subjects contribute only out of confusion.

This paper is part of the EU-TMR Research Network ENDEAR (FMRX-CT98-0238).

1. Introduction

Traditionally economists have not paid too much attention to the investigation of cross-country differences in behavior. This is probably due to the fact that economic analysis is based precisely on the idea that human behavior everywhere is guided by common principles. This view does not exclude these factors from having some effect on economic behavior; they are, however, considered to be just secondary influences. The question is, however, an empirical one and needs to be analyzed on the basis of data. It is difficult to conduct this kind of analysis using only field data. When comparing behavior across countries on the basis of information from the field one is generally looking at data which have been generated in different environments. This makes a direct evaluation of cultural differences alone problematic.

Experimental methods allow for a very controlled approach to the issue at hand. Examples of experimental 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.

Our objective in this paper is to 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. Our design makes it possible to study, at the same time, two different types of behavior, not exclusively motivated by subjects' own monetary payoffs. In this way we are able to obtain a unifying view of behavior.

We analyze contributions that are made in the case in which they lead to a decrease in the contributor's monetary payoff and to an increase of others' payoffs. There is now abundant evidence of contributions in this kind of situations. Ledyard (1995) and Schram (1999) present overviews of these results. This behavior has often been attributed to some kind of 'altruistic' or 'cooperative' motivation: subjects' utility may depend positively on others' payoffs.

We also study behavior for the case where the individual's direct interest is in full contribution. 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 standard linear public goods environment. Their central result is that subjects deviated to a substantial degree from a dominant strategy of full contribution. This kind of behavior has up to this point not been observed in other countries.

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.[1] 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.

Taken together the two types of deviations from the maximization of own payoff mentioned above point to an alternative interpretation of behavior in terms of decision errors. If one adopts this interpretation, one can view 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 the two general issues, whether behavior varies across countries and what the roles of spite, cooperation and errors are in explaining behavior, call for a more 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 for the two types of situations described above 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.

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

2.1. General features of the design

Our experimental design is based on the linear public good environment first used in Brandts and Schram (1998). The key feature of the design is that subjects do not just make a single decision in each experimental period. Instead, each subject makes 10 decisions, 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 both with respect to cooperative and to spiteful behavior by subjects. This is so because the different mrs for which our subjects had to make decisions were basically of two types. For some of the mrs a spiteful, but not a cooperative, motivation may lead to deviations from dominant strategy play. For the rest of the mrs, the reverse will be true: a cooperative motivation, and not a spiteful one, may lead to deviations from the dominant strategy.

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.[2] 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'.[3] We conducted some baseline sessions and some additional sessions; the difference between them is in the payoff schedule we employed. Table 1 shows the payoff schedules 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 private (public account). The last column on the right of the table gives the different mrs, which are common for all countries.

Note that for the parameter values shown in table 1, it is a dominant strategy 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. 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

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

1 to 8 (9 to 10). For situations 1 and 2 a subject hurts himself 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.[4] For situations 3 to 10, we have the more standard environment that 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 what it increases the total payoff.

Table 1: Payoff schedules, baseline sessions*

	Japan		The Netherlands		U.S.A.		Spain		
	A	B	A	B	A	B	A	B	MRS
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 in the columns 'A' ('B') are the payoff per token in the public account A (private account B). The mrs is defined as the value for an individual of a token invested in the private account divided by the value for an individual of a token invested in the public account.

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

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

[4] 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 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.

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 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

Table 2: Payoff schedules, additional sessions*

	Japan		the Netherlands		
	A	B	A	B	MRS
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.

the additional sessions to be 'symmetric' 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 two types of sessions are meant to provide different perspectives on the issues we are interested in.

Besides the cross-country comparison and the differense between the AS and the S sessions, we added one more treatment to our design. Since Andreoni (1988) there has been a discussion about the differences in behavior in a 'partners' and 'strangers'

design.[5] 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 recent 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. Our data set provides a good opportunity to test this hypothesis across four countries. Therefore we distinguished partners and strangers sessions in our AS sessions. In the S sessions we only used the partners treatment.

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[6] 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.[7] Subsequently, Schram ran the sessions in Arizona. The sessions in Osaka were run by Saijo and his collaborators with Brandts observing.[8] In this way we ensured 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

[5] 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.
[6] A detailed description of the procedures was used. This is available from the authors.
[7] 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.
[8] We would like to thank Yasuyo Hamaguchi for her help in running the Japanese sessions and Steve

protocol (which included written texts for all interaction with subjects) reduces the potential effects of this kind of influences.

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 all 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.[9]

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 were \$1 $\approx$ ƒ 1,60 $\approx$ ¥ 90 $\approx$ Ptas 114. Note that the numbers in tables 2 and 3 imply that we have used the exchange rate \$1 $\approx$ ƒ 1,50 $\approx$ ¥ 100 $\approx$ Ptas 100. 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.

Backerman and Matt Cox for their help in organizing the U.S. sessions.

[9] 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.

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.[10] 13 subjects per session were recruited according to the usual practice at the various sites. 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 the 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-strangers sessions. Hence we have data from 6 sessions in Japan and the Netherlands, 4 sessions in the United States and in Spain.

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

[10] 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.

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.[11] The relevant payoff table as well as the other information about procedures was known to all subjects in a session.[12]

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.

Table 3 gives some background data for the subjects in the baseline treatments. 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 strongly across countries. In our analysis below we take into account the possible impact of these demographic variables.

[11] 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.
[12] 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.

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. We do not present statistical tests in this section. These 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 and 0 tokens for mrs > 1. The standard prediction, that the contribution function is a stepfunction 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 in a given country. At this point we wish to single out one fundamental trait of our data: all eight aggregate contribution functions for all four countries and both treatments are monotonically decreasing.[13]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.

(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 seem to be smaller than for situations 6-10.

(Figure 2 about here)

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

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

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 it is useful to distinguish between behavior for mrs < 1 and behavior for mrs > 1. In the former case, where the dominant strategy is to contribute all the tokens to the public account, contributions to the private account may be attributed to either some kind of error, as in Palfrey and Prisbrey (1996), or to spite, as suggested by Saijo and Nakamura (1995).

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 in any period and define the *Gross Spite Index* (GSI) as the ratio between this number and the number of tokens available per period for situations with mrs < 1 (18 tokens in A and 45 tokens in S). 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.

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).[14] Parallel to the GSI, we define the *Gross Cooperation Index* (GCI) as the fraction 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 qualify our cooperation index with the term gross with the same intention as for the case of the GSI.

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 treatment and subject background effects and 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=GCI=0. In experiments, however, one has to

[14] 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). Contributions are extremely low in these situations. Moreover, it is not easy to think of a motivation for contributing in those situations, and we

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 Anderson, Goeree and Holt (1998). After referring to different possible sources of error they posit that the cost of a mistake affects its likelihood: relatively costly mistakes are less likely. 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:

$$H_0\text{: GCI=GSI}$$

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 some difference between the GCI and the GSI. 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 $H_{3A,}$ refer to differences in either the GCI or the GSI across countries or treatments.

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

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

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

know of no experimental evidence that shows that subjects contribute for a case like this.

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) and GCI(NL)>GSI(NL) and GCI(US)>GSI(USA) and GCI(SP)>GSI(SP).

A common observation when analyzing data from public goods experiments is that cooperation may decline as subjects gain experience with the environment. A similar decay may be present in spite, however. Our conjecture is that, in spite ot these decays, cooperation remains a stronger motivational force than spite throughout the 10 periods of play. We can formulate this hypothesis as:

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

Our two next hypotheses have to be seen as alternatives to an interpretation of the H_0 as stating that the GCI and the GSI are equal across countries and treatments. The Saijo and Nakamura (1996) results on observed spite in Japan mentioned in section 1 lead to:

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

As far as partners and strangers are 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 use data at the group level and 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, while for partners we consider data from period 1 as well as averages over all periods

The GSI and the GCI are the basis of our statistical tests. Table 4 presents an overview of the values of the average GSI and GCI for all the different conditions studied in the AS sessions.

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

	GSI JP	GCI JP	GSI NL	GCI NL	GSI US	GCI US	GSI SP	GCI SP	GSI ALL	GCI ALL
Part	7.8	22.7	4.6	29.0	8.2	25.8	1.6	17.2	15.6	23.7
Stra	14.8	20.3	7.1	27.5	18.3	19.4	4.0	15.5	11.1	20.7
Average	11.3	21.5	5.9	28.3	13.3	22.6	2.8	16.4	8.4	22.2

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

H_{A1A}:

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 data for period 1 (p=.000, N=48) and partners data averaged over all periods (p=.000, N=24).[15]

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 where 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).

These 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 that mistakes depend on the costs of making them? For the S sessions, a special feature of our design implies that the test we have just presented also rejects the modified null. It can be seen in table 2 that, for both countries, the costs of deviating from the dominant choices are symmetric around an 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 of this fact, 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 can be seen directly as a rejection of the costly mistakes hypothesis.

For the AS sessions we need to introduce one modification to be able to make the same argument. 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 before.[16] On the basis of a Wilcoxon Signed-Rank test we can reject the equality of the GSI and the modified GCI, denoted hereafter by GCI34, using both data from period 1 (p=.000, N=48) and averages over all periods for the partners treatment (p=.000, N=24).

H_{A1B}:

Table 5 presents the results of our per country comparisons of GSI and GCI.

[15] N denotes sample size.

[16] 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 through 4.

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	p=.031 N=12	p=.028 N=12	p=.002 N=12	p=.136 N=12
GSI vs. GCI All periods Only partners	p=.116 N=6	p=.028 N=6	p=.028 N=6	p=.046 N=6
GSI vs. GCI34 Period 1	p=.005 N=12	p=.008 N=12	p=.002 N=12	p=.015 N=12
GSI vs. GCI34 All periods Only partners	p=.046 N=6	p=.028 N=6	p=.028 N=6	p=.028 N=6

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

Note first that if we center on the results using GCI34, we can in all instances reject equality of the two indices at the 5% level. If we use the GCI, then there are two cases, US in period 1 and Japan over all periods, for which we can not reject equality at conventional significance levels. 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 GCI34 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.

H_{A1C}:

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

(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 appears to be 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=12) AS sessions in period 10 (p=.000, N=12), S sessions in period 1 (p=.002, N=12) and S sessions in period 10 (p=.004, N=12).[17] We can, therefore, reject the null hypothesis in favor of H_{A1C}.

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 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 group observations for both treatments together and group observations averaged over all periods for partners only.

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

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

[17] 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.

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)

* 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.

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 data from the partners sessions are shown in the third and fourth row of table 6. In each cell, the first number is the coefficient; below it one can first see the absolute t-value and then the p-value.

The 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 four regressions shown in table 6 none of the 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}, i. e. the GSI does not vary across countries and we find no difference between partners and strangers. In addition we also no effect of the country variables on the GCI and no significant effects

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 over period.

of the subject background variables on either GSI or GCI.

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 250 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. Note also from figure 3, 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.

The strong asymmetry of the step function around mrs=1 in the S sessions makes it specially clear that the costly mistakes hypothesis is not consistent with our data.

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 may, however, 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.[18]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 out of a linear vcm environment related to ours, albeit with just one, fixed, mrs. At this point, one candidate for an explanation of the discrepancies between our and their data are the differences in the presentation of the payoff information; both their rough and their detailed payoff tables differ from our presentation. It is also possible that subjects at the University of Tsukuba are different from subjects in all four locations we studied in this paper. Another explanation is that the new design we use in this paper elicits more reflective responses from subjects than an environment with just one mrs. Having to make decisions for ten different mrs, subjects may mentally take a step back and think more carefully about their choices. Spiteful behavior may be the result of a very impulsive reaction and that may be the explanation for why it only appears as a rather weak force in the data obtained in our design.[19] Economists are mainly interested in thoughtful behavior and we feel that, in this respect, the use of our design is appropriate for the task at hand.

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. The difference between the two indices is especially stark for the S design, as shown in figure 4. Models

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

based only on decision error, like the one used in Palfrey and Prisbrey (1996) or the one proposed by Anderson, Goeree and Holt (1998) which considers that mistakes depend negatively on their cost, can not accomodate our evidence. Some kind of cooperative motive is needed to accomodate the data. If we attribute part of the GSI to the presence of spite then our data allow us to make the following statement: spite is not as strong a motivation as cooperation and this difference is large enough not to be overwhelmed by the presence of error.

In this paper we have concentrated on showing that cooperation has a significant weight in our data across countries and have not attempted to identify more specifically subjects' cooperative motives. Some recent research, however, has combined specific cooperative motives with the presence of error to explain observed behavior. Palfrey and Prisbrey (1997) present an econometric analysis of experimental data on the basis of a model that combines 'warm-glow' and linear altruism. Anderson, Goeree and Holt (1998) present a theoretical model that combines linear altruism and error to explain a number of specific features of behavior found in public goods experiments. The evidence we present in this paper is, in principle, compatible with this kind of modeling. However, Brandts and Schram (1998) and Bolton, Brandts and Katok (1999) present data that are at odds with these linear formulations of other-regarding behavior.

The absence of cross-country differences may surprise some readers, since national 'cultures' appear to be such strong forces in shaping human behavior. Our results are, of course, very 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 in a systematic manner and to generate a rich data set of thoughtful decisions that needs to be taken into account in the debate.

[19] The contribution levels we observed for mrs > 1 are, however, in line with what is typically found in this type of vcm experiments. The same holds for the decline in contributions over periods and the fact that this decline is not converging to zero.

References

Anderson, S. P., J. K. Goeree and C. A. Holt (1998): "A theoretical analysis of altruism and decision error in public good games", *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, pp.291-304.

Bolton, G.E., J. Brandts and E. Katok (1999): "How strategy sensitive are contributions? A test of six hypotheses in a two-person dilemma game", *Economic Theory*, forthcoming.

Brandts, J. and A. Schram (1997): "Cooperation and noise in public goods experiments: applying the contribution function approach", mimeo, Instituto de Análisis Económico (CSIC), Barcelona.

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, pp. 41-60.

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

Kachelmeier S., and M. Shehata (1990): "Culture and competition: a laboratory market comparison between China and the west", *Journal of Economic Behavior and Organization* 19, pp. 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 (1995): "Public goods: a survey of experimental research", in *The Handbook of Experimental Economics*, by J. Kagel and A. Roth (eds.), Princeton University Press, pp. 111-194.

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

Palfrey, T.R. and J.E. Prisbrey (1997): "Anomalous behavior in linear public goods experiments: how much and why?", *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, pp. 1068-1095.

Saijo, T. and H. Nakamura (1995): "The 'spite' dilemma in voluntary contribution mechanism experiments", *Journal of Conflict Resolution* 39, 1, pp. 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. (1999): "Sorting Out the Seeking: Rents and Individual Motivation", *Public Choice*, forthcoming.

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

Figure 1a: Japan

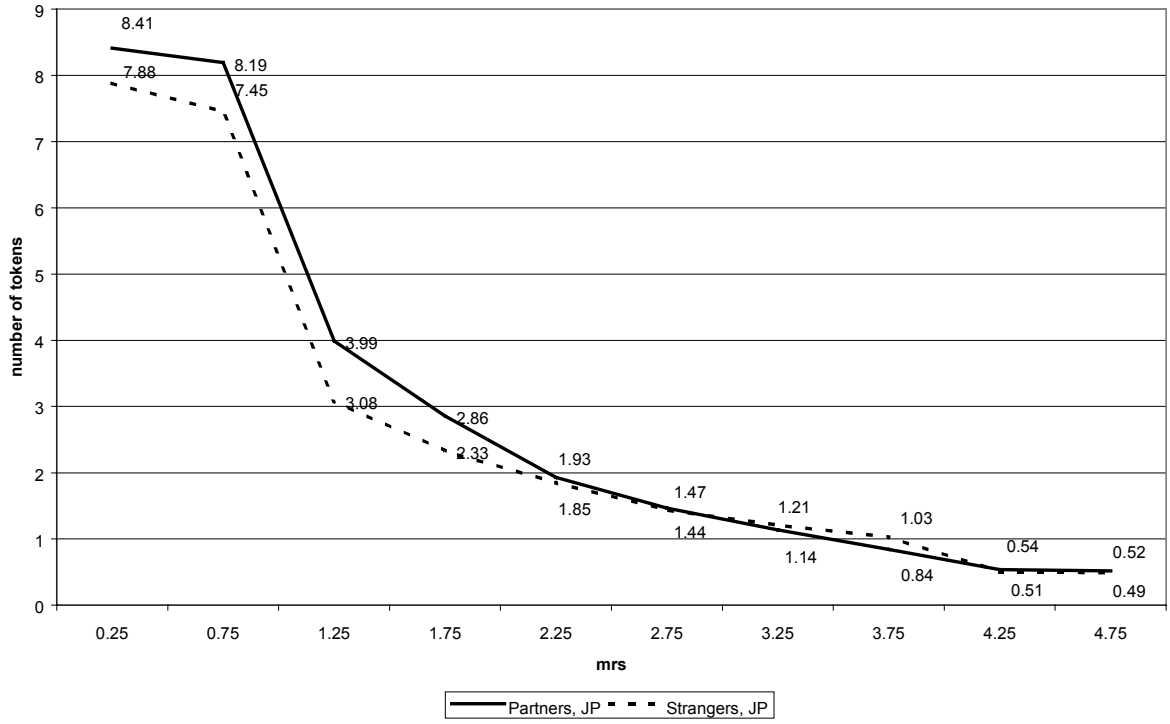

Figure 1b: Netherlands

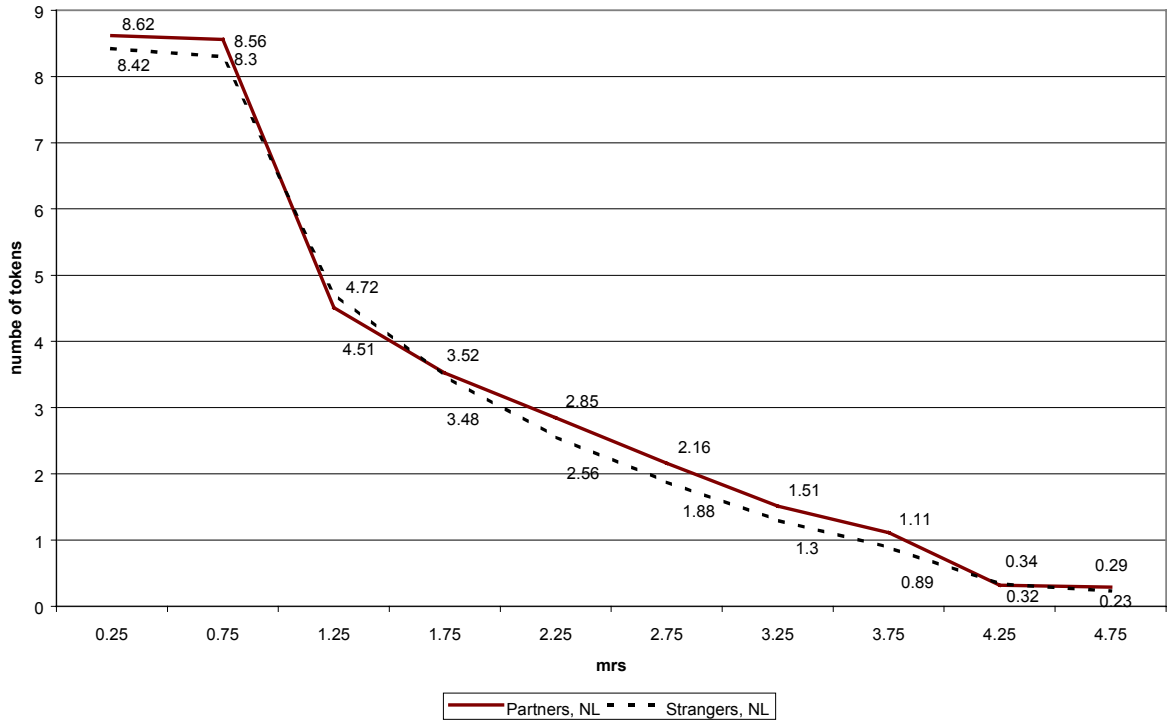

Figure 1c: Spain

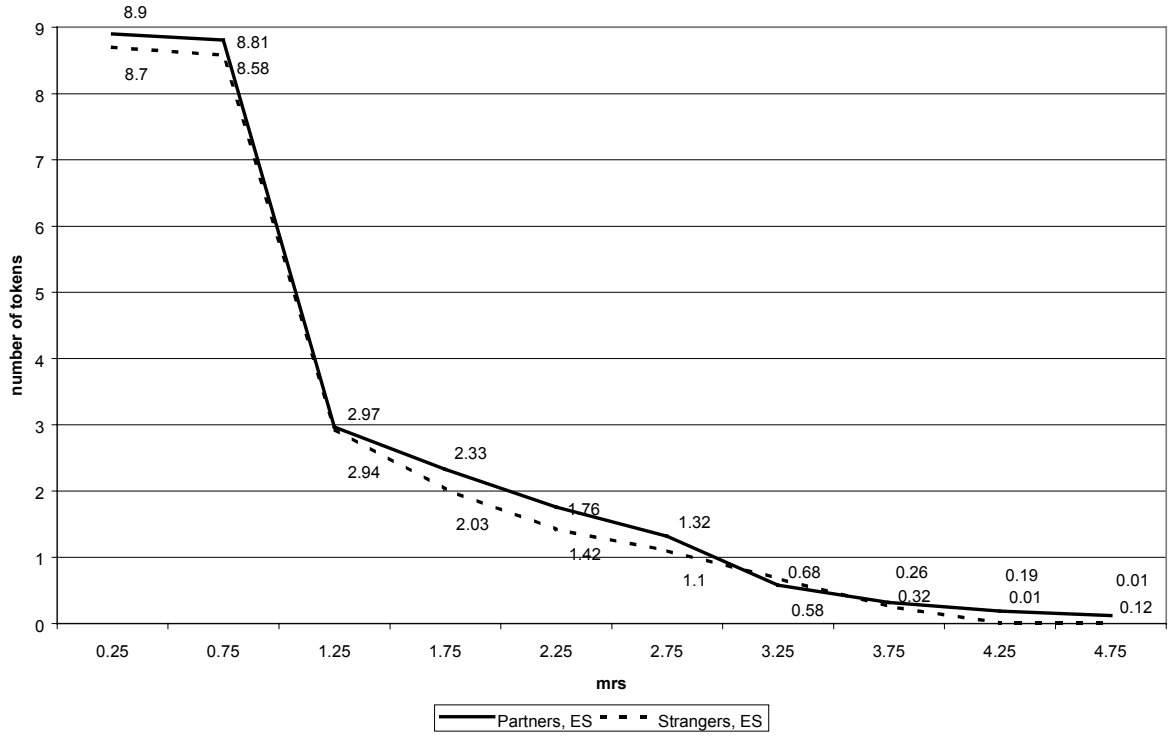

Figure 1d: United States

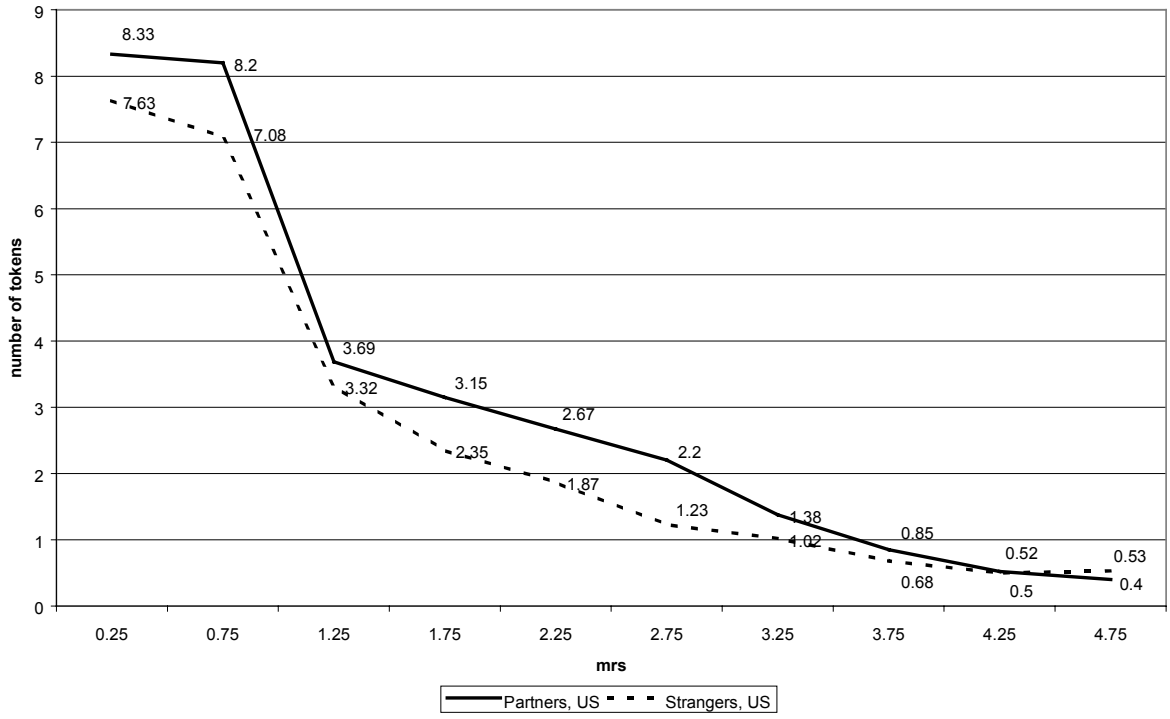

Figure 2: Symmetric Sessions

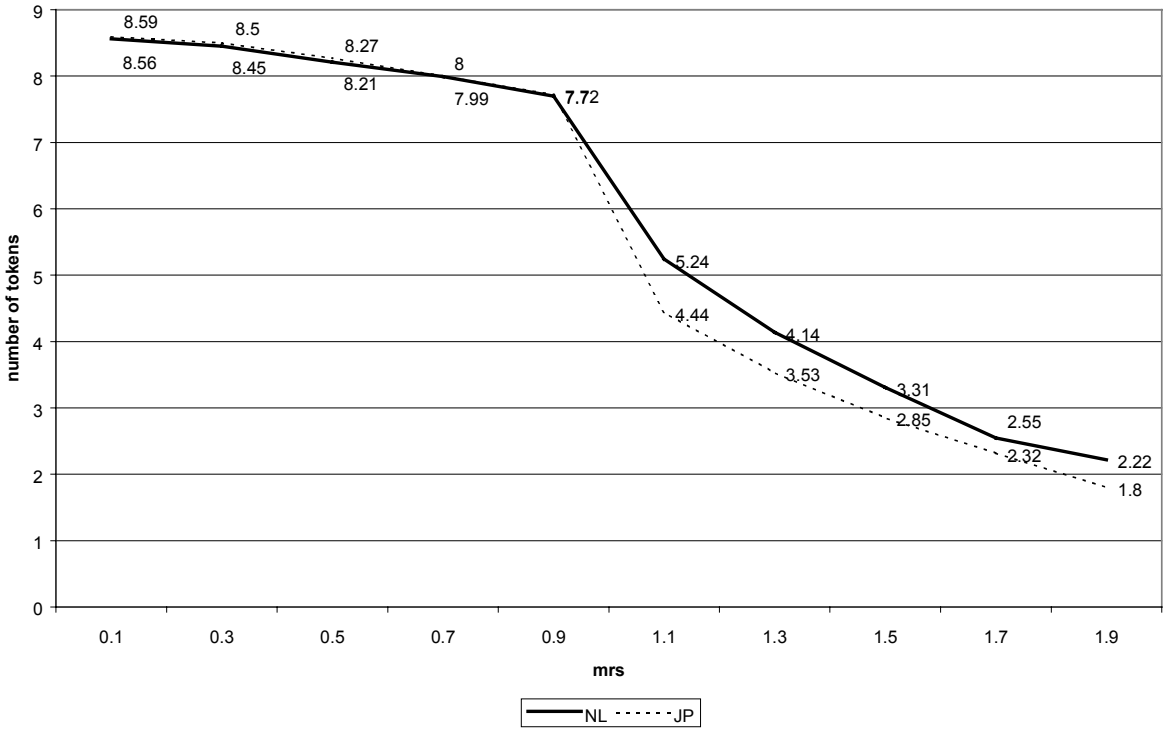

Figure 3a: GCI per country over time; AS sessions

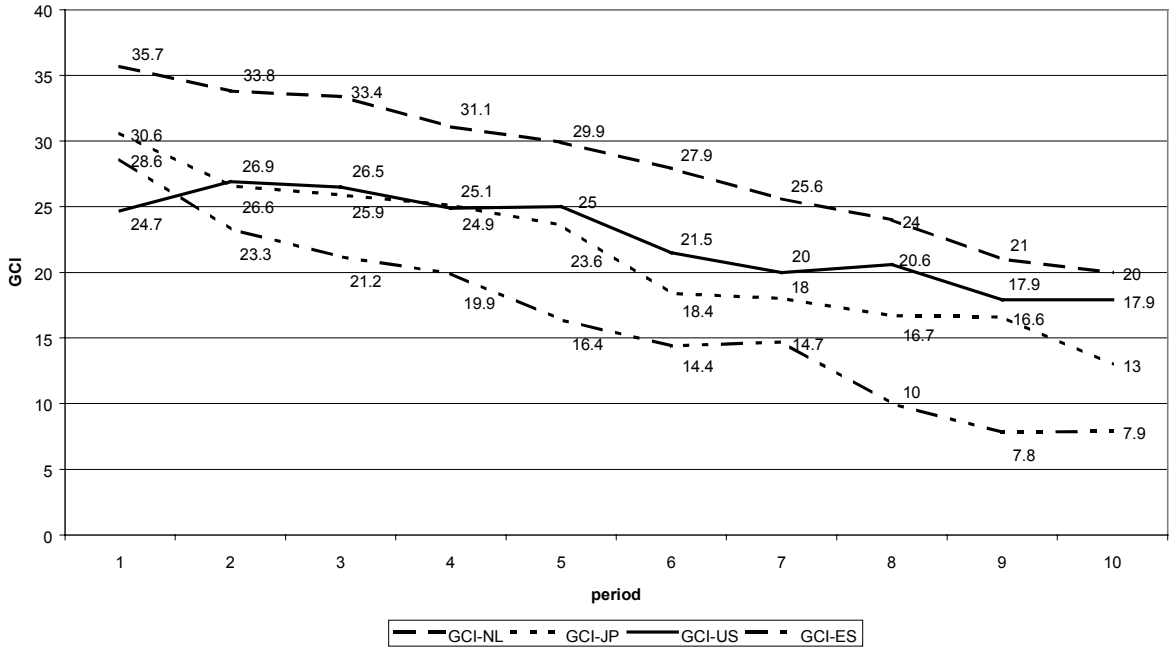

Figure 3b: GSI per country over time; AS sessions

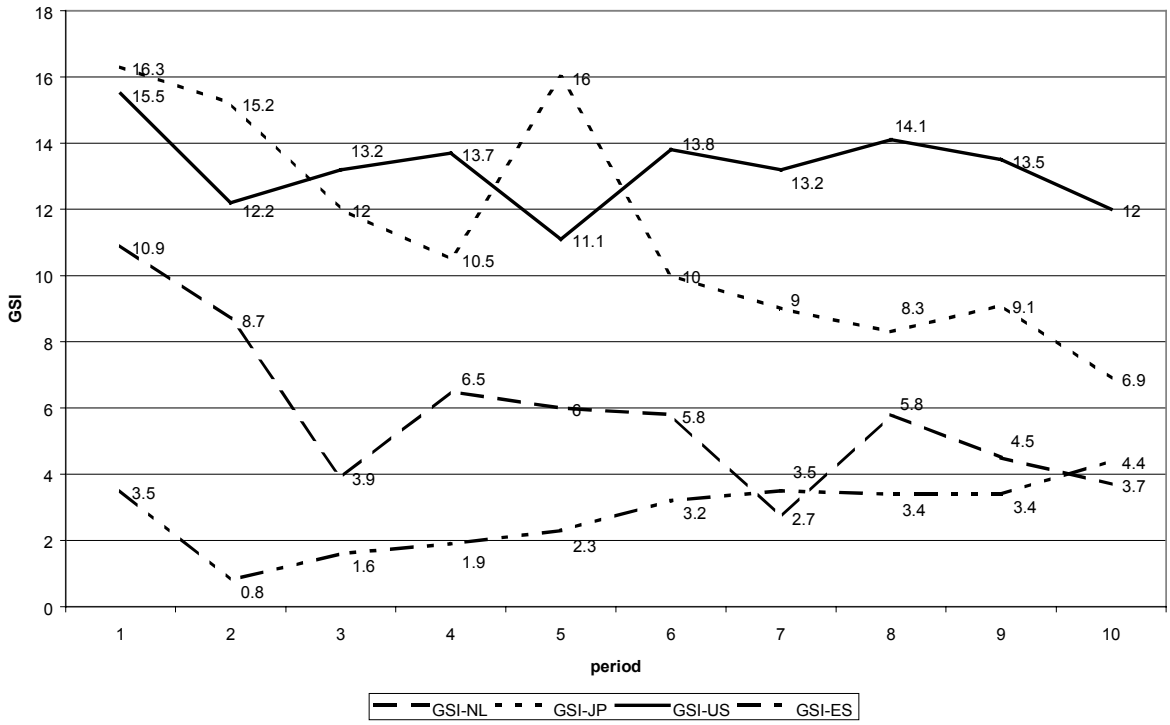

Figure 3c: GCI per country over time; S-sessions

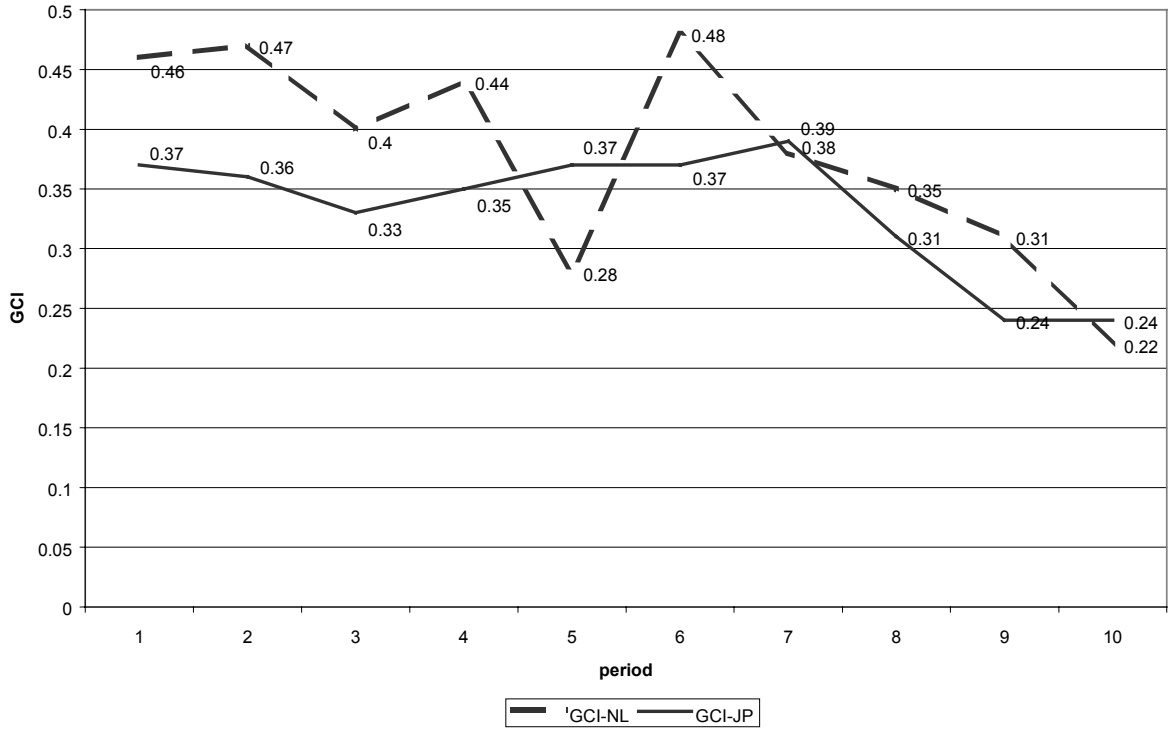

Figure 3d: GSI per country over time; S-sessions

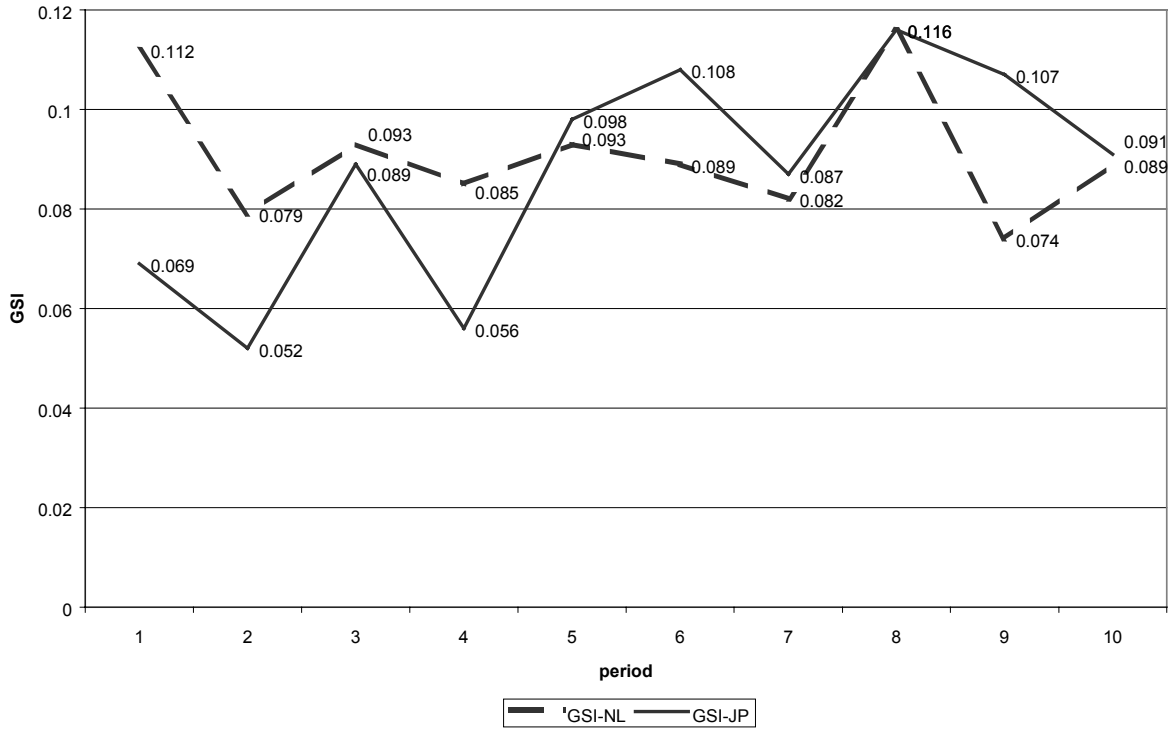

Figure 4a: Aggregate behavior; AS sessions

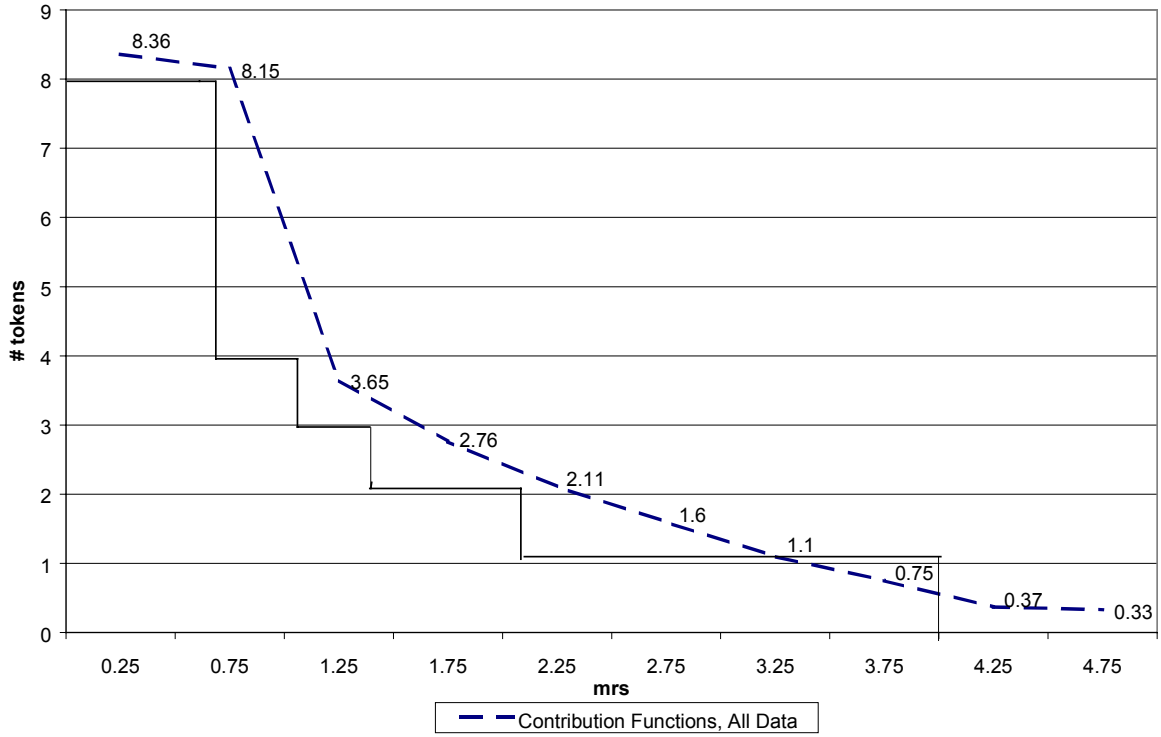
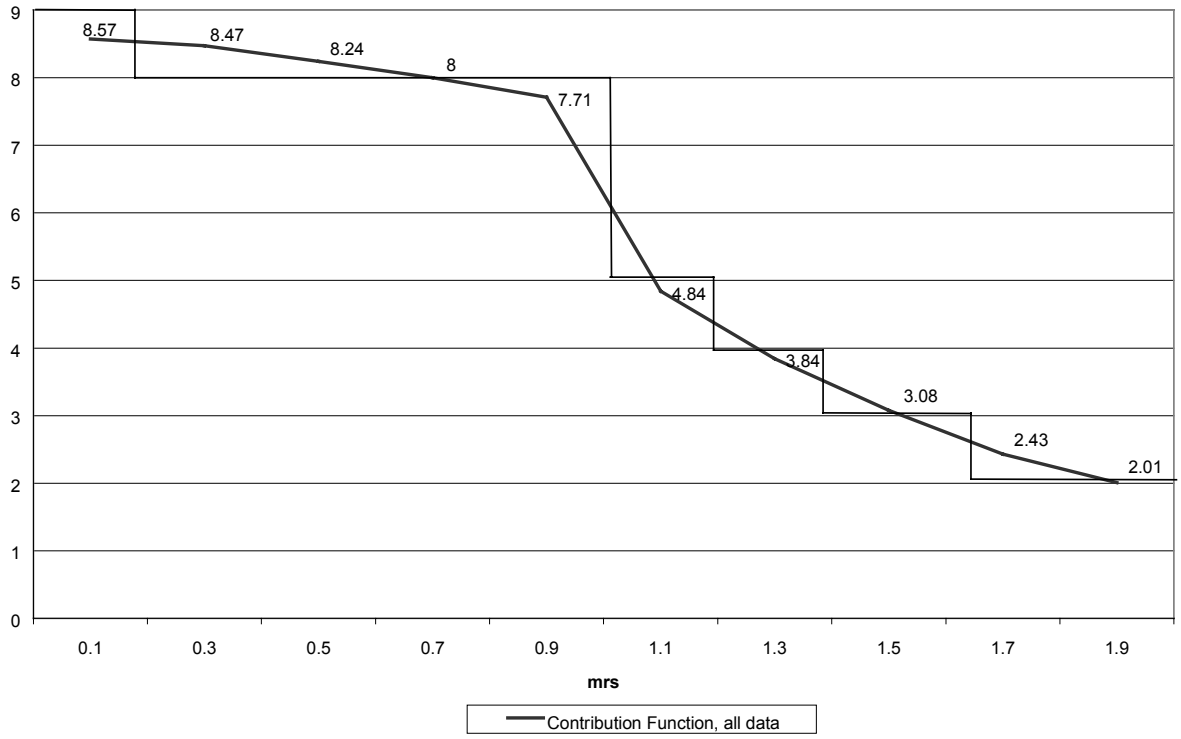
Figure 4b: Aggregate behavior; S sessions