Working Paper Series
No. 30

Social norms and social choice

Anabela Botelho
Glenn W. Harrison
Lígia Pinto
Elisabet E. Rutström

November 2005

Núcleo de Investigação em Microeconomia Aplicada
Universidade do Minho

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

Social Norms and Social Choice

by

Anabela Botelho, Glenn W. Harrison, Lígia M. Costa Pinto & Elisabet E. Rutström †

August 2005

ABSTRACT

Experiments can provide rich information on behavior conditional on the institutional rules of the game being imposed by the experimenter. We consider what happens when the subjects are allowed to choose the institution through a simple social choice procedure. Our case study is a setting in which sanctions may or may not be allowed to encourage "righteous behavior." Laboratory experiments show that some subjects in public goods environments employ costly sanctions against other subjects in order to enforce what appears to be a social norm of contribution. We show that this artificial society is not an attractive place to live, by any of the standard social choice criteria. If it came about because of evolutionary forces, as speculated, then The Blind Watchmaker was having one of his many bad days at the workbench. In fact, none of our laboratory societies with perfect strangers matching ever chose to live in such a world. Our findings suggest that the conditions under which a group or a society would choose a constitution that is based on voluntary costly sanctions are very special.

† Department of Economics, University of Minho and NIMA (Botelho and Pinto) and Department of Economics, College of Business, University of Central Florida (Harrison and Rutström). E-mail: botelho@eeg.uminho.pt, gharrison@research.bus.ucf.edu, pintol@eeg.uminho.pt, and erutstrom@bus.ucf.edu. Rutström thanks the U.S. National Science Foundation for research support under grants NSF/IIS 9817518 and NSF/POWRE 9973669. Botelho and Pinto thank the Fundação para a Ciência e Tecnologia for sabbatical scholarships SFRH/BSAB/489/2005 and SFRH/BSAB/491/2005, respectively. We are grateful to Ryan Brosette, Linnéa Harrison, James Monogan and Bob Potter for research assistance, and to Andreas Ortmann for helpful comments. All data, instructions, and statistical code is available at the *ExLab* Digital Library at http://exlab.bus.ucf.edu.

One of the fundamental methodological attractions of experimental economics is the ability to undertake controlled comparisons of behavior under different institutions. For example, subjects can be recruited at random into experiments that differ only in the way that they implement an auction, and the results directly compared. What then? Behavior in the two experimental institutions might differ, and on the basis of those differences one might conclude that certain agents would prefer one institution over the other, if given a choice. For example, if a seller only cares about the revenue obtained from an auction, and the seller gets to pick the form of the auction, the metric is a simple one. Given the choice between the two institutions, and assuming that the experimental comparisons are reflective of behavioral differences expected in the field application, the seller's choice is obvious.

However, many institutional comparisons are not so simple. There are often several dimensions to the behavioral outcome: at the very least, the profit earned by each agent needs to be examined if the "participation constraint" of mechanism design theory is to be satisfied.[1] In some settings there might also be concerns about the equity or efficiency of the overall outcome, and each of these concepts are open to different interpretations. In short: inferring preferences from the outcome of play under different institutions is a difficult, if not impossible, task.

The solution to this problem is to *expand the experimental design to allow subjects to directly choose which institution they would prefer to operate under*. There are many possible ways in which subjects could choose an institution. There could be a direct referendum vote, some committee or public choice process, a bargaining process, or migration to locales that are differentiated by institutions.

We illustrate the importance of this type of extension in the context of the debate over the value to social groups of having some "punishment" option when others deviate from some explicit

[1] This constraint holds that the expected income of the proposed institution must exceed or equal the expected income under any alternative, including the "outside option" of not participating at all. It implicitly assumes that agents cannot be coerced into participating in the institution. Of course, governments do have some powers of coercion. The "participation constraint" is discussed in most modern game theory texts, such as Fudenberg and Tirole [1991; p.247ff.] or Binmore [1992; p.533ff.].

or implicit norm. Fehr and Gächter [2000] [2002] (FG) devised a laboratory environment in which individuals chose to impose costly sanctions on others in order to preserve a social norm. They report that voluntary, costly sanctions result in significant improvements in cooperative public goods contributions. Their experimental design controlled for many of the strategic reasons for self-serving sanctions to be imposed. They hypothesize [2002; p. 139] that their evidence of the use of such endogenous sanctions "... has profound implications for the evolutionary study of human behavior."

We consider some other hypotheses. If the sanctions that they observed in their experimental societies were indeed the product of evolutionary forces, then The Blind Watchmaker was having one of his many bad days.[2] In almost every period, net returns to the majority of participants were lower with sanctions than without. We should then rejoice that the technology of sanctions and norms *might* instead be guided in modern society by explicit social choice. We hypothesize that a deliberate, conscious voting process would not lead to such societies. Put another way: *who would want to live in a society with this norm, if given the choice?* Perhaps, if the public choice process of social institutions is indeed driven by evolutionary forces, institutions with costly sanctions would eventually be driven out. In any case, we expect the preferences over institutions to vary greatly across subjects, depending on how the presence of sanctions affect their earnings.

Our experiments directly address the question of whether institutions with such costly sanctions would be preferred over institutions where no such opportunities were present. If one assumes that earnings are the motivating factor behind institutional choice, their data do not unambiguously suggest that this should be the case. We report an experiment that extends their design with an explicit public choice phase where preferences over the institutions can be observed directly.

---

[2] Dawkins [1986] famously introduced the metaphor of a blind watchmaker to make the point that complex objects could be produced by an evolutionary process that had no conscious intention of producing the object. That process relies on some criteria of fitness and survival to be applied to weed out the many mistakes that random deviation tosses out as part of such a mindless process. In fact, it is critical to this evolutionary argument that there be lots of such mistakes.

In section 1 we briefly review the results of the behavioral comparisons undertaken by FG, and argue that they provide a particularly fertile case study for asking if anyone would choose to live in the world where voluntary, costly sanctions are available. In section 2 we undertake the extended experimental design in which we first impose institutions on subjects following FG, and then allow those subjects to vote on the "constitution" that they will subsequently live under. We compare treatments where subjects have some chance to anonymously encounter each other more than once, called Random Strangers, with treatments where it is clear and obvious to them that this can never happen (Perfect Strangers).

We find that *none of our laboratory societies under Perfect Strangers conditions chose to live in a world with sanctions*. This result was robust to the order in which they experienced the alternatives, and parametric variations in the opportunity cost of free-riding. The vote was not even close, and in one case it was unanimous. When we relax our re-encounter conditions to that of Random Strangers, we find one case in which the majority votes for the world with sanctions: this case requires that the rewards to contributing be relatively high, and that the different regimes be experienced in a certain order. Our findings suggest that the conditions under which a group or a society would choose a constitution that is based on voluntary costly sanctions are very special.

In our experiments the driving force behind this reluctance to adopt the sanctions scheme appears to have been the reduction and uncertainty of profits that it caused. There is a strong negative correlation between the vote for the sanctions institution and the loss in profits that it caused. Other motivations, such as fairness, may also have played a role, but profits are a key candidate for what motivated actions.[3]

Beyond the question of social norms and social choice, our experiments identify an

[3] Consider a world of sanctions in which a majority of subjects made more profit on average than they would in a non-sanctions world, but a minority of subjects earned virtually nothing. Average profit is greater with sanctions, and for a majority, but one could easily imagine that some in the majority might not want to live in such an inequitable world, and would vote against the world with sanctions. Our results can be explained more easily, but such inequities are rife in laboratory societies such as these, and we expect that factors other than average profit would play a role in social choice more generally.

important methodological point about the use of inferences from experiments in the design of policy. If subjects in the field have mechanisms by which they can avoid, lobby or self-select into or out of institutions, we must consider the effects of those margins of choice before drawing conclusions about which institutions are best. Another way to express this is to consider if the laboratory environment that takes a particular institution as fixed is correctly modeling the naturally-occurring environment in it's salient features, if that environment includes ways in which subjects can endogenously opt out of that institution.[4]

1. The Value of a Social Norm Enforced by Punishment

What is the net benefit to society of the norm considered by FG? More precisely, what is the net value of allowing the punishment technology that endogenously generates the norm? The top panels of Figure 1 illustrate the manner in which FG [2000] address this issue.[5] The top panels show the value of the economy with and without the norm in place. Consistent with their discussion [2000; p.90] of the payoff to subjects in their experiments, aggregate earnings in the pooled Stranger treatments were greater with the norm after period 8. Our analysis breaks out the experiments in which the norm was available *before* experiencing an environment in which it was not available (the "First Series"), and experiments in which it was available *after* (the "Second Series").[6] In one case the

---

[4] This point arises consistently when making inferences about treatment effects from social experiments (Heckman and Smith [1995] and Philipson and Hedges [1998]), as well as when one compares the methodological strengths of laboratory and field experiments (Harrison and List [2004]). The reason is that the inability to impose all aspects of an environment in field experiments and social experiments often obliges one to worry about such endogenity as an essential component of evaluating treatments or institutions. Of course, we have already noted the role of explicit "participation constraints" in mechanism design theory.

[5] We only consider the "Strangers" design in FG [2000], since it controls for the possible role of strategic self-interest in employing the sanctions. In their "Partners" design the same subjects played against each other for ten periods, and in their "Strangers" design individuals were randomly assigned to groups after each period. They only consider Strangers designs in FG [2002]. We are grateful to Simon Gächter for providing the data from their experiments.

[6] All experiments lasted for 20 periods. In the First Series, the initial 10 periods allowed punishment and the second 10 periods did not. In the Second Series, the initial 10 periods did not allow punishment and the second 10 periods did. Thus the subjects in the Second Series could be viewed as having lived through the

aggregate profits with the norm exceeded the aggregate profits without the norm in periods 9 and 10, and in the other case only in period 10.

The results in the top panels of Figure 1 suggest that there did appear to be convergence to a cooperative outcome in period 9 or 10. But the accumulated cost of the convergence path in periods 1 through 8 more than offsets the incremental gains in periods 9 and 10. The aggregate loss is 12½% in the First Series, and 17% in the Second Series.[7] Thus, to conclude that the norm is valuable for these economies, on balance, rests entirely on a speculative extrapolation beyond the life of the experiment. Looking solely at the top panels of Figure 1, some observers *might* be willing to make that extrapolation, but there is much more to the story.

Figure 2 reports comparable calculations for the experiments in FG [2002]. Here the design was similar to the Strangers treatments in FG [2000], although the punishment cost schedule was linear rather than convex in punishment points. Each treatment consisted of 6 periods. The results in Figure 2 are similar to those in Figure 1, but even more striking in terms of the persistent costliness of the norm. In each Series the aggregate loss in value from the norm is roughly 15%. Moreover, there does not appear to be a firm basis for extrapolating beyond the horizon of the experiments as the basis of a conclusion that the norm will eventually become valuable.

More to the point, would the participants in any of these experiments want to have this technology available if they were to make a social choice after their experience? The results in the bottom panels of Figures 1 and 2 consider this question, using two possible voting rules[8] for social choice:

- Majority Rule Referenda – would the median voter opt for the social technology?

"nasty, brutish and short" Hobbesian world in which contribution norms could not be punished.

[7] For example, average profits in the Second Series were 22.73 currency units and 18.85 currency units, respectively, for a difference of 3.88 currency units or 17.1%.

[8] The results in the top panels imply what would happen if a Classical Utilitarian social choice rule was used, in which aggregate benefits were compared to aggregate costs. Over the life of the experiment the norm would not be approved. However, it would be approved if one were to just use the results of the last period or two to calculate benefits or costs in the AER experiments (Figure 1).

- Super-Majority Rule Referenda – would 67% of the population vote for the social technology as a "constitutional" matter?

One particularly nice feature of the FG [2000][2002] design is that it allows in-sample comparisons of the value of the norm to each individual subject. Each subject participated in each condition, so it is a simple matter to calculate the earnings for each subject with and without the norm. From the distribution of net profits, so calculated, one can calculate the period-wise median and 33rd percentile. These are shown in the bottom panels.

The implication from Figure 1 and 2 is that, with two exceptions, *the social norm would not be adopted under either of these social choice criteria*.[9] Furthermore, it is much harder to argue *a priori* that simple extrapolation beyond the life of the experiment provides any basis for predicting that the norm would be socially acceptable under these criteria.

Of course, this conclusion rests on an *ex post* evaluation of the performance of each institution. One could argue that an *ex ante* evaluation would arrive at a different choice, since the "trends" are more likely to indicate future outcomes than the "integral" of past outcomes. This is a speculative claim given the trends shown in Figures 1 and 2.[10]

---

[9] These exceptions are period 8 of the First Series in Figure 1, and period 5 of the Second Series of experiments in Figure 2, where the median voter would just vote *for* the norm. The norm would not survive a constitutional referendum using a super-majority rule in these periods.

[10] One could argue that the punishment norm does not require costly punishment in order for it to work it's magic. Masclet, Noussair, Tucker and Villeval [2003] (MNTV) modify the basic FG design to consider a treatment in which subjects can send "sanctioning points" that do not reduce the profit of the recipient or cost the sender anything. The points do send a signal of disapproval that would not have been possible without the experimenter allowing this technology, but there is no direct cost to anyone in terms of profit reductions. They find that this treatment increases group contributions, suggesting that the social choice we consider is moot. Unfortunately, their design confounds the Partners and Strangers design kept separate by FG. In every one of the experiments of MNTV the first 10 periods used a Partners design in which the subjects played against the same players in each round. The 3 sessions that used a Strangers design only employed it after round 10, which is the same round that the sanctions could be employed in their design. They find that average group contributions rise when non-monetary sanctions are used, but only in the design in which they used the Partners treatment at the same time as allowed the non-monetary sanctions. When they used the Strangers treatment, they found that average contributions continued to decline from round to round when non-monetary sanctions were allowed (apart from a one-period "restart" effect, which is standard in these public goods experiments). The powerful interaction effect of the Strangers and Partners treatments can be seen from the experiments of FG [2000]. Conditional on the use of a Partners design, costly sanctions

2. Voting for a Social Norm Enforced by Punishment

A. Basic Experimental Design

We design a simple experiment to test whether subjects would choose to live in a world with costly sanctions. In the first part of the experiment we replicate the design of FG by providing subjects with experience in public goods contribution games in which there is a punishment norm as well as controls in which there are no such norms. We examine order effects as they do by running one set of subjects through experiments in which the punishment option comes first, and then the no-punishment option is experienced, and then running a separate set of subjects through the same game but with the reverse order. In each case we allow 10 periods in each setting.

To ensure that there are no confounding reputational factors, and to provide the cleanest possible test, we use a Perfect Strangers design in which no subject ever meets the same subject more than once. Virtually all previous public goods experiments use an ordinary Strangers design in which subjects are randomly re-assigned every period.[11] Although this reduces the chance that the subject will meet the same person to extremely low levels, and is coupled with anonymity, the critical behavioral issue is whether the subjects behave as if they believe that there is no reputational effect of their choices in a given round. In our experiments subjects participate in groups of 2 in each round.[12] We explain in great detail how we ensure that there is no chance that they will meet the same person in any other round. Since this is a departure from previous experimental practice, albeit

---

have a strong effect on average profits as early as rounds 4 or 5. The question posed by FG is whether costly sanctions would have an effect in the Strangers environment that controls for other reputational sanctions that might be present in a Partners environment.

[11] Andreoni [2005] reviews the literature on public goods contributions with Partners and Strangers. FG [2000; fn.3] report that the results of a Perfect Strangers replication of their design generated essentially the same results as their ordinary Strangers experiments. However, they only considered one sequence of regimes (Punishment followed by Non-punishment), and did not maintain the Perfect Strangers treatment after the first regime of 6 periods. Rather than debate if such comparisons are conclusive, we prefer to ensure the control against any reputational effects afforded by a Perfect Strangers design.

[12] Most public goods experiments use four subjects per group, although the effect of larger group sizes has been studied by Isaac and Walker [1989] and others. Harrison and Hirshleifer [1988] and Goeree, Holt and Laury [2002] employed groups of 2 in their public goods experiments.

one that is motivated by taking seriously the same game-theoretic rationale for ordinary Strangers designs over Partners designs, we undertake control experiments to see the effect of using an ordinary Strangers design instead of a Perfect Strangers design.

After round 20 we ask subjects to vote on the environment they would like to participate in for one "Final Jeopardy!" round.[13] The instructions they received in one of the treatments are as follows:

> We are now ready for your final task. This will consist of only one period. The task will be a repetition of one of the two tasks you have just completed. Which task this will be will be determined by a common vote in a moment. In this one period the stakes will be increased so that each token is now worth 50 cents, not just 5 cents. This is therefore 10 times the value that a token has had in each of the earlier periods.
>
> Before you play out this one period, you will be asked which environment you would like to participate in. You may choose either the one where you can reduce other participants' earnings and they can reduce yours (**environment B**) or you may choose the environment in which there is no such opportunity (**environment A**). Everyone will be asked to vote for the environment that they prefer, and **we will implement the environment that a majority of the participants in this room vote for.** Thus, we will implement the same environment on all matched pairs.
>
> In the event of a tied vote we will roll a ten-sided die for you all to see. If the die comes up 0-4 we will implement environment A, where earnings reductions are not available, if it comes up 5-9 we will implement environment B, where earnings reductions are available.
>
> Before you are asked to vote you will be shown a screen with a review of your earnings across the periods in both of the environments.

Once a decision is made, all subjects play the chosen environment for 1 period. In order to enhance the *relative* saliency of the voting decision, which is the main focus of our design, we tell subjects that their earnings in this period will be ten times those of each of the first 20 periods.

Table 1 summarizes the experimental design. Eight sessions were conducted. The first four used Perfect Strangers designs, and the last four used ordinary Random Strangers designs. The

---

[13] In the popular TV game show *Jeopardy!* there are three rounds of play: "Jeopardy!," "Double Jeopardy!," and "Final Jeopardy!" The first two consist of three categories of three questions each, but "Double Jeopardy!" has doubled dollar values. There is only one question in "Final Jeopardy!," and subjects can wager their accumulated earnings in that round.

return to the public good is discussed below, as are the votes.

B. Parameters and Treatments

Parameters must be chosen carefully. Table 2 spells out the arithmetic confronting the subjects in FG [2000], translated into equivalent U.S. dollars. Each subject could choose to impose a punishment defined in terms of points, which would cost a certain amount of Lab-Guilders (the unit of currency in the lab). These Lab-Guilders were converted to Swiss Francs at a fixed exchange rate of 0.2 Francs per Lab-Guilder. Since a Swiss Franc was worth $0.84 on average in January and February 1996, this converted to the Cost in U.S. dollars shown in the third column. To put these costs in perspective, average profits *before* punishment costs in this experiment were $23.67, as shown in the fourth column. Hence the cost of punishment can be calculated as a percentage of average profits, and the punishment effected as a percentage of average profit. The punishment effected was 10% of the recipient's profit times the number of points. Of course, to punish N players the cost to the punishing player would be N times the costs shown here.

Table 2 shows that the punishment technology is really quite Draconian. Relatively low cost punishments effect relatively large punishment impacts, measured in dollar terms of percentage reductions. This suggests a potentially interesting extension of our approach: instead of deciding if there should be this punishment technology or none, they could decide how "nasty" the technology should be. This could be effected by appropriate choice of parameters in a schedule such as shown in Table 2, or some cap on the impact of punishments. In fact, FG implemented a ceiling punishment of 100% of the recipient's profits in that period, so that if the poor subject received more than 10 points in punishment the impact would be no profit rather than a loss.

FG were very careful to ensure that subjects did not make a loss, or even face the prospect of making a loss unless they worked hard at doing so. In FG [2000] they gave each subject the equivalent of $12.60 for showing up, and then effectively endowed them with $1.05 per period with which to pay for punishments.[14] As Table 2 shows, this allowed *each* subject in *one* period to buy up to 9 punishment points without incurring a loss in that period (and before factoring in any profit from production of the public good or the private good).

We used the same punishment schedule in our experiments. Each point allocated to punish the other player implied a 10% reduction in the other player's earnings in that round. The cost to the subject inflicting the punishment was converted to our lab tokens at the same rate as their points were converted to their Lab-Guilders.[15] Each subject received an endowment of 20 tokens at the outset of each round, and each token was worth 5 cents. Additionally subjects received a one time endowment of 25 tokens to cover possible losses.

In one treatment we used a relatively low return on contributions to the public good, and in another treatment we used a relatively high return. The low return was 0.6 of token: hence every token contributed to the public good by one subject would decrease their private endowment by 1 token and return 0.6 of a token for themself. Of course, it would also generate 0.6 of a token for the other player, so the social return was 1.2 tokens for every 1 token invested. In the high return treatment we changed the public good return from 0.6 to 0.8, thereby increasing the social return from 20% to 60%. The objective of this treatment was to see the effects of making the environment more rewarding to anything that would increase contributions to the public good. Table 1 shows that the low return was used in sessions 1 and 2, and the high return in all other sessions.

We used a linear payoff schedule which was constant for all contributions, so the dominant strategy is simple: a subject that only seeks to maximize individual earnings in a single period should contribute nothing to the public good.[16]

---

[14] The endowment was actually paid out once for the experiment, rather than on a per-period basis.

[15] For example, punishments of 1, 2 and 3 points cost the punisher 1, 2 or 4 tokens, respectively.

[16] Alternative assumptions about the factors motivating subjects to contribute in public goods experiments have long been studied. See, in particular, Palfrey and Prisbrey [1996][1997] and Goeree, Holt and Laury [2002].

C. Procedures

We recruited 142 subjects from the University of Central Florida (UCF) in 2005.[17] Subjects were randomly assigned to each session, with no prior knowledge of the parameters or treatments. The sessions were all conducted at the Behavioral Research Lab of the College of Business Administration of UCF. This facility is a standard, computerized laboratory: each station has a "sunken" monitor, and we employed personal "cubicle-style" screens to ensure even more privacy. Instructions were provided in written form and orally, and the experiment was implemented using version 2.1.4 of the *z-Tree* software developed by Fischbacher [1999].[18] The same experimenter (Rutström) delivered the oral instructions for all sessions, to ensure comparability.[19] The oral instructions also utilized a large-screen display that could be easily seen by all subjects, to ensure that certain information was common knowledge. Training rounds were included prior to each regime, to ensure that subjects understood the task.

Average earnings in these experiments were $38, including a standard $5 show-up fee. No session lasted more than 2 hours, and most were at least 1½ hours in length.

D. Results

Table 1 shows the vote in each session, which is our "bottom line" result: when there was a zero chance of ever meeting any other person again, as in the Perfect Strangers design, no cohort voted for the punishment regime.[20] Overall only 18% of participants in the Perfect Strangers treatment voted for the punishment regime. The vote was close in one of the four Perfect Strangers

[17] UCF is located in Orlando, Florida. It has a large student body, with Fall 2004 enrollment of 42,837. The entering class in 2004 had an average SAT of 1,186. The student body is also ethnically divers: in 2004 8.5% stated that they were Black and Non-Hispanic; 70% stated that they were White and Non-Hispanic; 5.0% stated that they were Asian; and 12.2% stated that they were Hispanic.

[18] All instructions, scripts, and software are available at http://exlab.bus.ucf.edu. The latest version of the *z-Tree* software and documentation is available at http://www.iew.unizh.ch/ztree/index.php.

[19] A digital recording of the oral instructions in one typical session is available at the ExLab archive.

[20] We use the term "cohort" for the group of subjects who together make a public choice. In the Random Strangers design, two cohorts were present in each session.

sessions, but there was little doubt in the other three. In fact, in one session, all 26 subjects agreed unanimously to implement the no-punishment regime. This results was robust to the use of high or low returns to the public good, and the history that the subjects experienced prior to the vote.

We do find a significant difference in public goods contributions under Perfect Strangers and Random Strangers conditions, and this difference is also reflected in differences in voting behavior. Specifically, we find that *the assumption that subjects treat Random Strangers designs as if they were one-shot experiments is false.* One shot games are best modeled using a Perfect Strangers matching protocol since there is no repeated interaction there, not even a probabilistic one. Our subjects behave in a systematically different manner in the Perfect Strangers design. In fact, Botelho, Harrison, Pinto and Rutström [2005] show that the Perfect Strangers design is associated with more subjects adopting a strict free-riding behavior consistent with the one-shot theory, rather than with subjects simply providing smaller contributions conditional on making some contribution. Thus the use of the Perfect Strangers design seems to *encourage a qualitative change in the way subjects view the game.* We show below that this effect interacts with the number of potential opponents in the Random Strangers design, as one would expect (since a large enough N would make Random Strangers virtually Perfect Strangers, and a small enough N would make Random Strangers virtually a Partners design, in which the same subjects are matched up in every round).

Figure 3 provides detailed results for session 1, to illustrate the outcomes. The top panel shows average token contributions in each period, and the bottom panel shows average profits in each period. Since there was a punishment regime in periods 11 through 20, we show pre-punishment profits as well as post-punishment profits. Of course, the latter were the "take home" profits to subjects, and the ones that they are assumed to be motivated by. In terms of contributions, we observe a now-standard pattern in voluntary contribution experiments: subjects start out making some contributions, and then free riding sets in. This particular session collapsed almost to complete free riding, which is more extreme than our other sessions, but the decline was general. After round 10 there is a "re-start" effect, which is also a common effect, although not a universal one.

In terms of profits, the striking outcome is that of periods 11-20 for the punishment regime. The pre-punishment profits of subjects was roughly comparable to the profits earned in periods 1-10, but the post-punishment profits were much lower. This reduction is particularly evident in the first 4 periods of the punishment regime, with many subjects exercising their ability to punish others. If one compares the average profit in periods 1-10 with the average post-punishment profit in periods 11-20, it is not hard to see why every subject voted for the no-punishment regime.

These results are extreme, but illustrate the factors that went into each vote. One could argue that the vote was stacked against the punishment regime by it being second, when the standard decay in contributions had set in. But a counter-argument is that it is precisely in such a setting where the punishment regime might be of value, since nobody needs a punishment regime if everyone is contributing heavily. And, of course, we test for such order effects from the sequencing of the two regimes. One could also argue that the vote was stacked against the punishment regime by the return to the public good being low, but again a counter-argument would be that this is precisely when one needs some external device to get people to contribute, since the intrinsic returns are not high. We also considered higher returns to the public good in sessions 3 though 8.

For completeness, we also show in Figure 3 the average contributions after the vote, in period 21. The profits for this period were ten times the profits for each of the prior rounds, to increase the salience of the vote, but we display scaled-down levels of profits for comparability. An appendix displays similarly detailed outcomes for each of the other sessions.

Figures 4 though 7 show the average "take home profits" in each period and session, along with the percentage vote for the punishment regime. For comparability, each has the same vertical scale.

Figure 4 shows the results for sessions 1 and 3, which shared the same NP-P history and the Perfect Strangers design, but differed in terms of the return to the public good being low or high. The unanimity of session 1 has been noted, but here we also see that only 8% of the subjects in the high return session 3 voted for the punishment regime. In this case the contributions to the public good were relatively high in periods 1-10, were still around 7 or 8 tokens by period 10, and declined very slowly in periods 11-20. This is exactly what one would expect from the change from low returns to high returns to the public good, which is the only difference between the two sessions. The use of punishment in periods 10-20 of session 3 was relatively sparing. FG noted that some subjects also engaged in so-called "spiteful punishment." Such punishment is said to occur when someone who was a free rider punishes a contributor, and is extremely costly for the cohort and the evolution of a social norm. In these session we found very little "spiteful punishment" occurring. However, the punishment that did occur, along with the continued slow decay in contributions over time, resulted in take home profits for session 3 that were systematically lower than those in the no-punishment regime.

Figure 5 shows the results for sessions 2 and 4, also Perfect Strangers sessions, which shared the same P-NP history and again differed in terms of low or high returns to the public good. We again see the marked difference in contributions with the change in the return to the public good, across both regimes. Round 1 deserves comment, since we see a dramatic reduction in take home earnings in both sessions, due to extravagant use of the punishment option. We conjecture that this is due to some subjects learning about the nature of the punishment technology "the hard way." In one session we had one subject privately ask the experimenter, "if I punish the other person, do I get their earnings?" Of course, this had been explained in the instructions, but as every experimenter knows there are always some subjects that gloss the written and oral instructions, or do not trust them, and use the actual session to try things out. It should also be noted that we included two periods of non-paid training prior to each session. Nonetheless, the behavior in period 1 in sessions 2 and 4 (and sessions 5 and 6, discussed below) is consistent with this conjecture. The fact that the reduction stopped being so dramatic after round 1 is consistent with the subjects learning the rules of the game, as distinct from experimenting with the right dose of punishment (as one observed in periods 11-14 of session 1, shown in the bottom panel of Figure 3).

Nonetheless, these two sessions provided a stronger vote in favor of the punishment regime than the other two Perfect Strangers sessions. Compared to sessions 1 and 3 (Figure 4), the major change is the sequence of the regimes, with the punishment regime being experienced first. In session 2 average take home profits under the punishment regime were consistently around 85 cents or 90 cents after the bloodbath of period 1, but they were steadily just above $1.00 for the no-punishment rounds 11-20. Thus only 21% of the subjects voted for the punishment regime. We undertake a formal statistical analysis of individual votes below, to see if the personal history of the subject influenced the vote. That is, even if average profits were lower for all subjects under the punishment regime in session 2 compared to the no-punishment regime, maybe they were higher for those 21% that voted for the punishment regime.

Session 4 was a voting cliff-hanger, of the kind that one only finds in Florida! Contributions started out relatively high, and apart from another period 1 bloodbath of punishment, the punishment was relatively efficient and non-spiteful. Average contributions actually increased from around 10 tokens in period 1 to 11 or 12 in periods 4-10, with take home profits around $1.25 after period 2. The happy bubble crashed in period 11, with a dramatic fall in contributions. However, free riding did not take over completely, and subjects continued to contribute around 5 tokens per period on average, and profits averaged about $1.16 in periods 11-20. When the vote came, 42% voted for the punishment regime.

Figure 6 shows the results for sessions 5 and 6. These are both Random Strangers sessions, which share the same P-NP history, and high returns to the public good. They differ in terms of the number of subjects that were in each cohort. In session 5 we had random draws from N=10, and in session 6 we had random draws from N=16. Subjects were made aware of the size of their cohort. This difference provides a nice bridge between the complete absence of re-encounters in the Perfect Strangers design, and the perfect rematching in the Partners design. With N=10 there is a higher chance of meeting the same people in later rounds than with N=16.[21]

[21] Sessions 5 and 6 were conducted in the same physical session, so there were 26 subjects in the room. Computer stations were previously logged on to two different servers running two different sessions.

Session 6 provided results that matched those in sessions 2 and 4, consistent with subjects being aware that the larger cohort size implied a smaller chance of a rematch with the same person. Contributions started out around 7 tokens per period, and decayed very slowly. They were around 4.5 tokens by period 10, and declined slowly through period 20. Punishment in periods 1-10 was costly, even after the period 1 bloodbath: average profits were lower by over 20 cents in each period because of the punishment. As Figure 6 shows, average profits were systematically higher, and less variable, in periods 11-20 of session 6, so it was no surprise that only 8% voted for the punishment regime.

Session 5 was a "poster boy" for the story that FG proposed. Contributions started high, around 10.5 tokens, and generally kept at that level with the use of sporadic, efficient punishment. But this was a setting in which the mere threat of the use of sanctions seemed to have the desired effect: nobody needed to "pull the trigger" since contributions were generally high and profits robust, and there were no vandals engaging in spiteful punishment that would destroy any evolving social norm. With only 10 subjects in the cohort, it is likely that the subjects perceived the higher rematching probability as resulting in reputation effects. Although the now-customary period 1 bloodbath occurred, and might have weighed against the vote for the punishment regime, 60% of the subjects presumably viewed that as an outlier from the promise of things to come if they had another, final period of the punishment regime. They were right: in period 21 of session 5 average contributions jumped from close to 0 in period 20 to over 5. This is in itself an interesting finding since the one-shot nature of the final round could easily have caused the social norm of cooperation under the threat of punishment to break down, but it did not.

Sessions 7 and 8, both Random Strangers sharing an NP-P history and a high return to the public good, have fewer cohorts than sessions 5 and 6. The differ from these earlier sessions also in the order of the sanctioning regimes. Although the patterns here are more erratic than in session 5,

---

Upon entering the lab subjects chose their seats. After seating subjects were handed cards showing the number of subjects in the cohort. The same procedure was used for sessions 7 and 8.

for example, the punishment regime successfully increases profits over time at a relatively small cost, but not sufficient to result in a higher average profit than the no-punishment regime immediately prior. Despite the smaller size of cohorts, compared to sessions 5 and 6 with the opposite order, we see a fairly strong preference for the no-punishment regime.

In summary, we only find one session in which there is majority support for the punishment regime. This is a Random Strangers session with a small cohort, where subjects experience the punishment regime first, and where the return to contributions is high. In a similar session, but where subjects experience the non-punishment regime first, we find almost majority support, but in all other sessions the majority of the subjects prefer to live in a world without punishment regimes.

Table 3 displays the results of a statistical analysis of individual votes. The dependent variable is the vote for the no-punishment regime. Explanatory variables include individual demographics and treatment effects. Binary dummy variables are included for the Perfect Strangers designs (Pstrangers), the size of the cohort conditional on the use of a Random Strangers design (Gsize),[22] the history during periods 1-10 (np_p), whether the subjects received a high rate of return for contributions to the public good (high), the ratio of take home profits in the NP regime to the P regime (ProfitRatio), and the ratio of the standard deviation of profits in the P regime to the standard deviation of profits in the NP regime (ProfitSDRatio). Demographics include a measure of age in years (Age), binary indicators for sex (Male), race (Black, Asian, Hispanic or White), academic major (Business), class standing (PreSenior), cumulative GPA below 3¼ (GPAlow), cumulative GP above 3¾ (GPAhigh), number of people in the subject's household (Hhsize), and a binary indicator of those that work part-time or full-time (Work). An appendix lists descriptive statistics for these variables.

The top panel of Table 3 shows estimates for the complete sample, and the bottom panel

[22] This variable takes on the value 0 for the Perfect Strangers treatment and the size of the cohort (the "N in session" column from Table 1) for the Random Strangers treatments. Thus it can be viewed as an interaction between the Perfect Strangers treatment and cohort size. Cohort size here is not the number of players in each particular public good game, which is always 2, but the number of people from which the pairings were selected.

then reduces the sample to those sessions with high returns to the public good. Consider the results for the complete sample. The effects from absence of re-encounters in the Perfect Strangers are striking, and associated with an increase in the probability of voting for NP of 68.3 percentage points (*p*-value = 0.006). Related to this effect, the size of the cohort of potential opponents in the Random Strangers environment also has a significant effect: every extra cohort member is associated in this environment with a 3.6 percentage point increase in the probability of voting for NP (*p*-value = 0.016). Panel A also provides evidence that the history experienced by the subjects significantly affects their propensity to vote for the NP regime: just moving from the P/NP sequence to the NP/P sequence increases the probability of voting for NP by 15.7 percentage points (*p*-value = 0.018). As expected from our prior discussion of results, the use of high rewards encourages outcomes in which participants are less likely to vote for the NP environment, since the return to encouraging cooperation by the *efficient use of punishment* is higher: the effect of this treatment is to lower the probability of voting by 12.5 percentage points on average (*p*-value = 0.033). Another powerful factor in the vote is the ratio of take home profits in the two regimes. The average value for this measure was 1.07 over all sessions, with a standard deviation of 0.304, so a single standard deviation change in the ratio would be associated with the subject being 15.0 (= 0.304 × 49.3) percentage points more likely to vote for the NP regime.[23] Turning to individual demographics, we find that Blacks and Asians are more likely than others to vote for the NP regime, with average increases in the probability of voting for NP of 8 percentage points and 11 percentage points (with *p*-values of 0.05 and 0.002, respectively). Those with higher cumulative GPAs are roughly 6 percentage points more likely to vote for the NP regime, although the effect is not as statistically

[23] The ratio of take-home profits, as well as the standard deviation in those profits, might be viewed as an endogenous variable in this specification of voting behavior. We checked for this using an efficient instrumental variables estimation procedure proposed by Newey [1987]. The most natural identifying assumption for this procedure is that the experimental treatments determine the profits, but only affect the vote through their impact on profits. Under this assumption there is evidence of endogeneity, but the conclusions about the effect of average profits in the two regimes remains the same. A stronger identifying assumption, that the experimental treatments *and individual demographic characteristics* only affect the vote by determining profits, leads to the different conclusion that there is no endogeneity, but again leads to the same conclusion about the effects of average profits on voting.

significant as the others (p-value = 0.17).

Panel B examines the same statistical model estimated on the sample with high returns to investment in the public good. Very similar results are obtained: the absence of re-encounters of Perfect Strangers is very important, as is the history experienced by the subject. The profit ratio in the two environments becomes more important. But essentially the same qualitative story emerges from this restricted sample.

3. Related Literature

Ehrhart and Keser [1999] examine the effects of allowing "Tiebout mobility" in a basic public goods contribution game. Their idea is to allow subjects to "vote with their feet" and decide which group they would like to be in, so that individuals that have a taste for the public good could associate with like individuals. Their experiments implemented this option in a simple manner, with 9 subjects in each session being able to choose which group they wanted to participate in at the outset of each of 29 rounds after the first. In the first round subjects were randomly assigned to groups of 3. Thereafter, groups could consist of 1 through 9 subjects, depending on subject choices. Subjects knew the size, total and average contribution in each group prior to deciding whether to migrate. Migration was costly: 50% of the endowment each period.

The results were disappointing, in the sense that endogenous migration did not generate the homogeneous groups one might expect. Basically, free-riding individuals behaved as if seeking out cooperating individuals. That would not be so bad for public good provision if they changed their self-interested ways, but after joining the group they exploited it and the process cycled. Overall, average contributions to the public decayed steadily. Unfortunately, there is no control experiments with the same sample, so one cannot say if this decay was less than one would otherwise see.

Page, Putterman and Unel [2002] extend this idea in several ways. In each session 16 subjects participate in a voluntary public goods contribution game in groups of 4 for 20 rounds. After round 3, they are allowed to individually rank the other 15 individuals, who have anonymous labels. The information available after each round is the *average* contribution of the other individual over the experiment up to the previous round. A subject could assigned multiple ranks to several individuals, so one could identify that 4 subjects were ranked #1, for example. Each top rank, or set of top ranks, cost the subject who made the ranking 25 experimental cents. Since the initial endowment in each round was 10 experimental dollars, this was not a major cost in comparison. Each subsequent rank cost the subject 5 experimental cents. For perspective, each experimental dollar was converted to 7 cents at the end of the experiment; so subjects were effectively endowed with 70 cents, had a cost of 1¾ cents to make a top rank, and had a cost slightly greater than ⅓ cent to make a lower-level rank.

After the rankings were submitted, an algorithm assigned subjects to groups of four. The algorithm first found the group of 4 subjects who had the smallest summed ranks of each other, then did the same for the next group of 4 subjects, and so on. Although alternative objective functions could be employed to this voting problem, the basic logic of this matching routine was relatively easy to explain to subjects.

Four environments were examined. One was a *baseline* in which there was no punishment option or ranking. The second was a *punishment* environment, akin to the one studied by FG. The third was a *regrouping* environment in which subjects were placed into groups in rounds 4-20 based on the rankings submitted. The fourth was a *combination* of the punishment and regrouping treatments. They found no significant pairwise differences in contributions or earnings between the last three environments. However, they did find a statistically significant increase in contributions and earnings when the baseline and regrouping environments were compared, and when the baseline and combined environments were compared.

4. Conclusions

The use of controlled treatments is a fundamental feature of most experimental designs. A baseline treatment is defined, some different treatment imposed, and the subjects randomly assigned to one or the other. This use of imposed treatments may not be a control, however, if the behavior of interest involves the subjects themselves making a decision as to which "treatment" to participate in. Although related to the sample selection and sample attrition problems, the issue can be better framed by asking if the experimental control removes the very margin of choice that it is supposed to help explain. We illustrate the natural methodological extension to experiments in which subjects can explicitly choose the environment in which they are to participate.

In the specific setting examined here the simple maximization of expected profit appears to explain the choices made by subjects when allowed to vote on the environment. One could anticipate more complex settings, in which the dispersion and inequity of profits might also factor into the voting choices of individuals. Furthermore, there are many ways in which subjects can rationally respond to environments they find unattractive, and we have only examined one such route of social choice.

Figure 1: Value of the Social Norm in AER Experiments

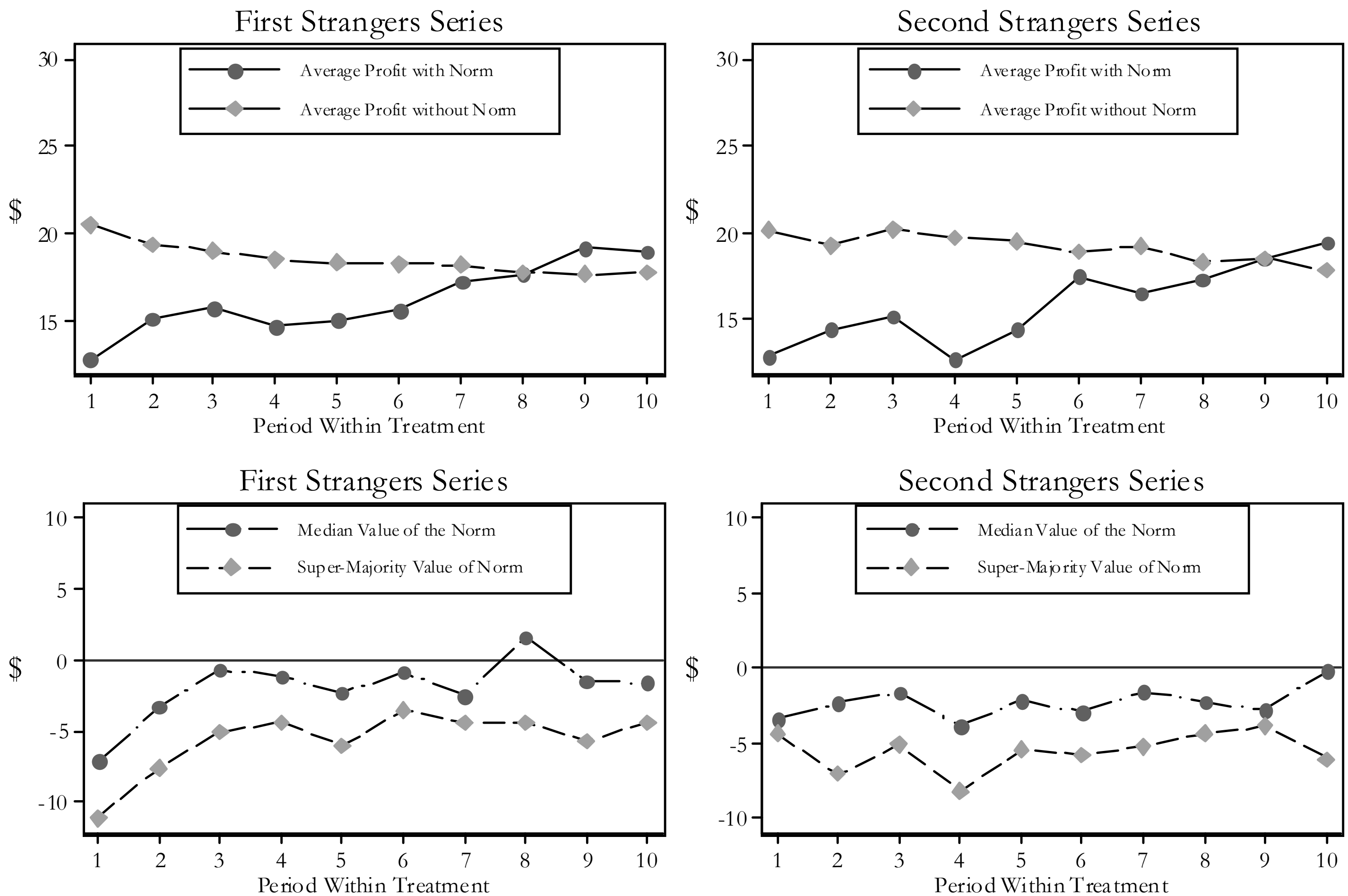


Figure 2: Value of the Social Norm in Nature Experiments

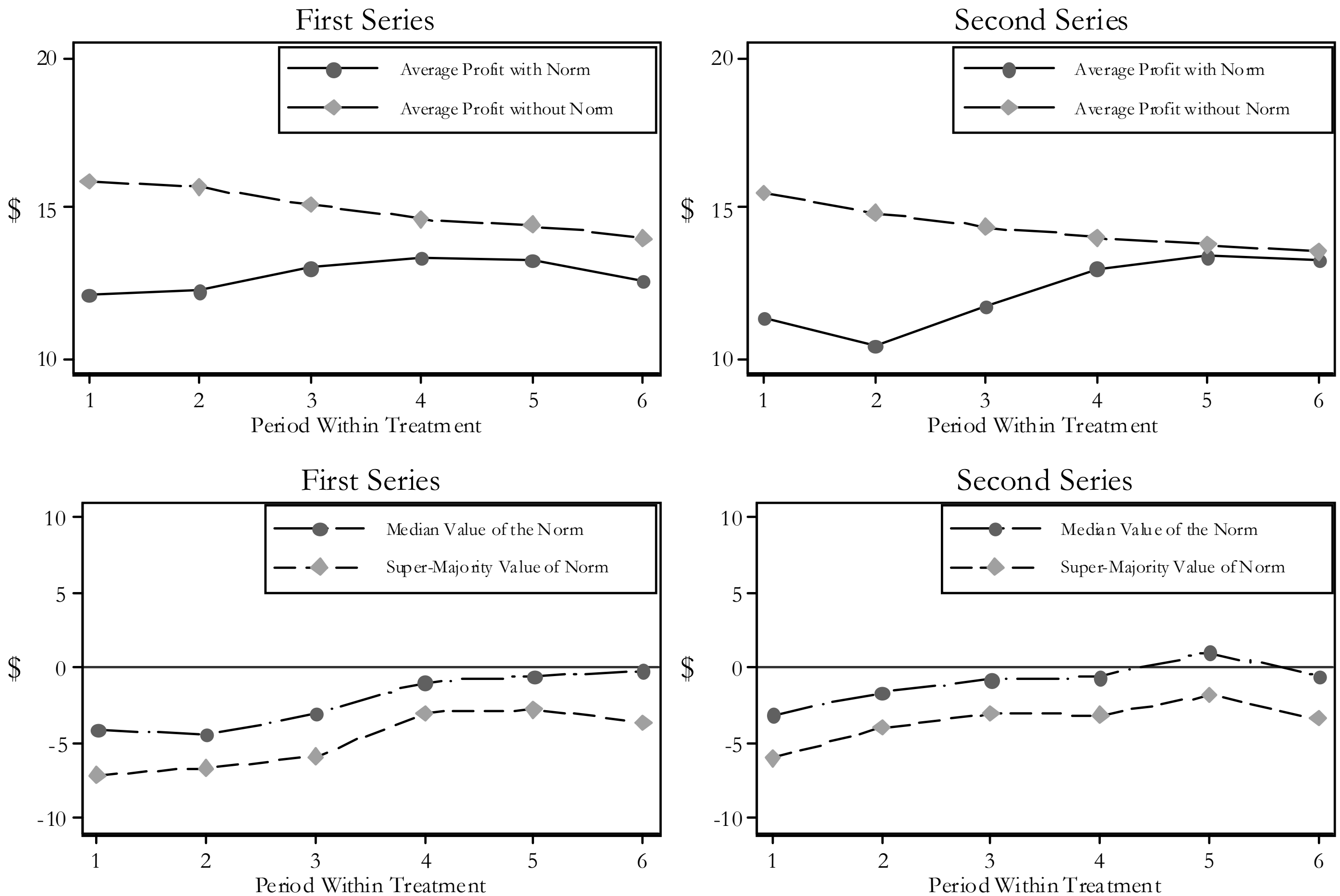


Table 1: Experimental Design

Each experiment had 10 rounds of one regime, followed by 10 rounds of the other regime
After round 20, all subjects voted on the regime for round 21
Round 21 had 10 times the payoffs of each of rounds 1-20

Session	Return to Public Good	Matching	N in Session	History	Average Profit per Period			Vote for Punishment
					NP	P	NP	
1	Low	Perfect	26	NP-P	$1.01	$0.96		0%
2	Low	Perfect	24	P-NP		$0.89	$1.01	21%
3	High	Perfect	26	NP-P	$1.25	$1.12		8%
4	High	Perfect	26	P-NP		$1.22	$1.16	42%
5	High	Random	10	P-NP		$1.26	$1.05	60%
6	High	Random	16	P-NP		$0.98	$1.08	19%
7	High	Random	8	NP-P	$1.34	$1.34		25%
8	High	Random	6	NP-P	$1.30	$1.25		50%

Table 2: Punishment Schedules

Schedules used in Fehr and Gächter [2000] *AER* Experiments

Points	Cost in Lab-Guilders	Cost in $USD	Average Profit in $USD	Punishment Impact in $USD	Cost as % of Profit	Punishment as % of Profit
1	1	$0.04	$23.67	$2.37	0%	10%
2	2	$0.08	$23.67	$4.73	0%	20%
3	4	$0.17	$23.67	$7.10	1%	30%
4	6	$0.25	$23.67	$9.47	1%	40%
5	9	$0.38	$23.67	$11.84	2%	50%
6	12	$0.50	$23.67	$14.20	2%	60%
7	16	$0.67	$23.67	$16.57	3%	70%
8	20	$0.84	$23.67	$18.94	4%	80%
9	25	$1.05	$23.67	$21.30	4%	90%
10	30	$1.26	$23.67	$23.67	5%	100%

Figure 3: Results in Session 1

N=26 Perfect Strangers in Groups of 2
Low Return to Public Good

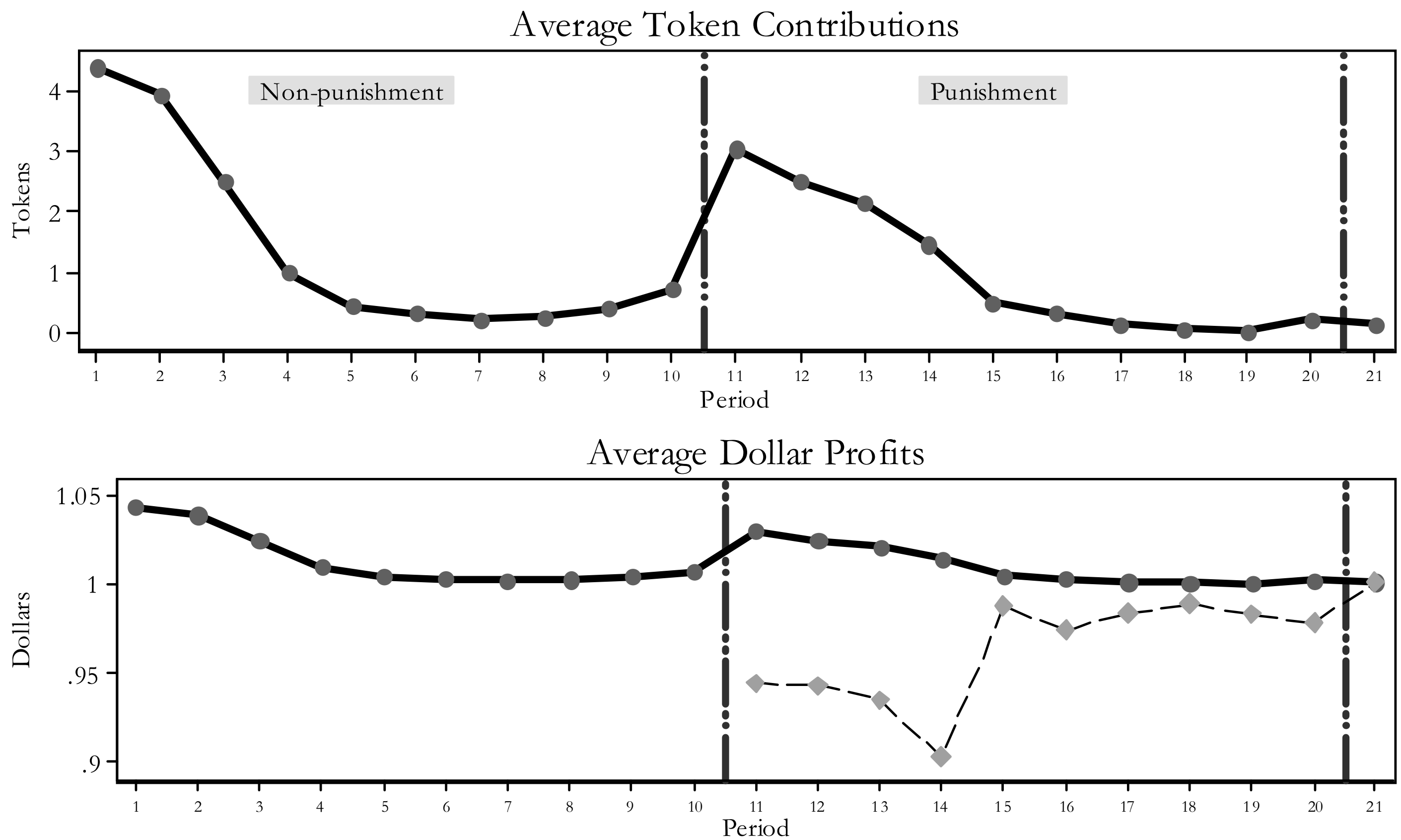


Figure 4: Average Profits With
Perfect Strangers and NP-P History

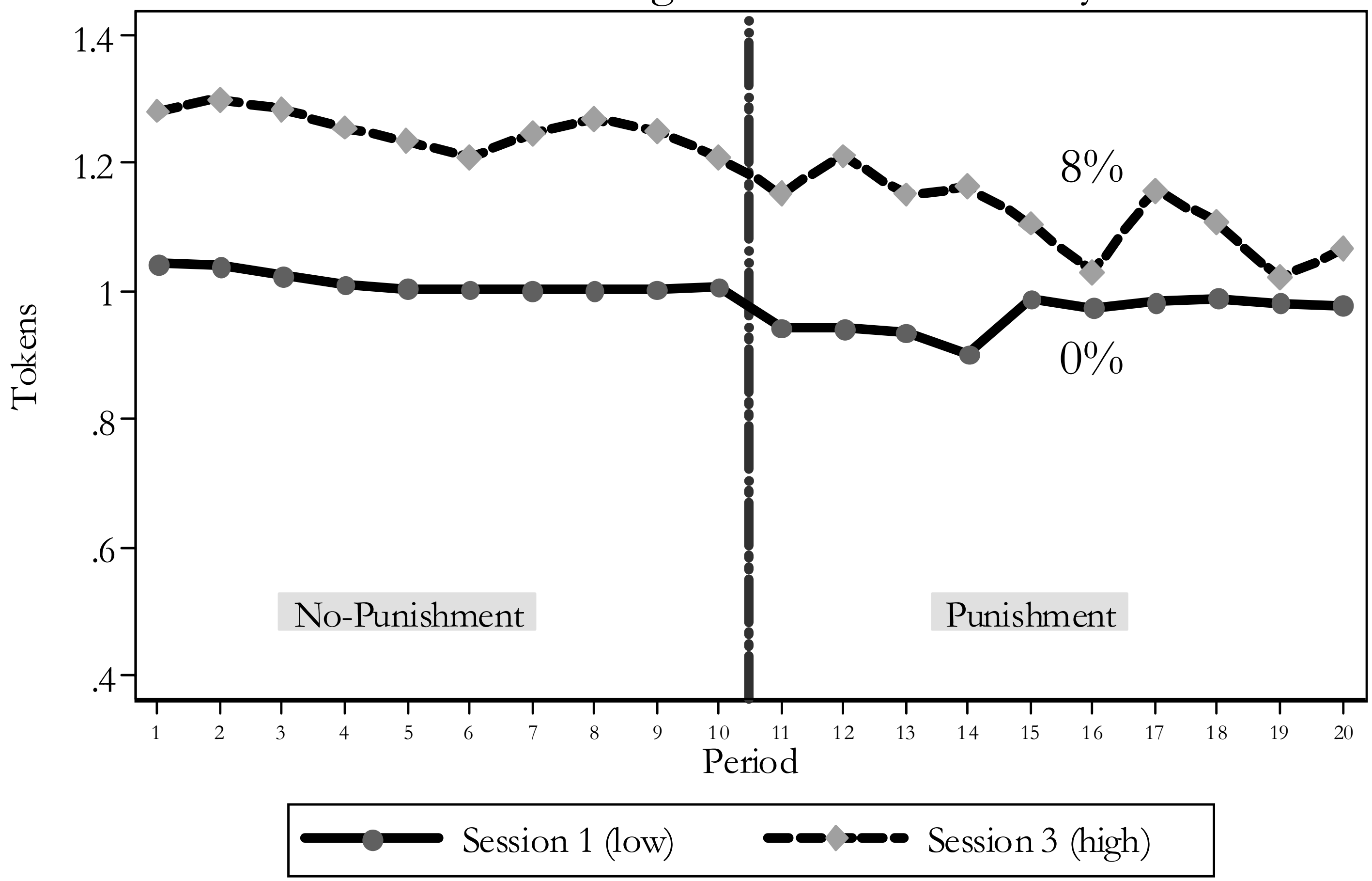


Figure 5: Average Profits With
Perfect Strangers and P-NP History

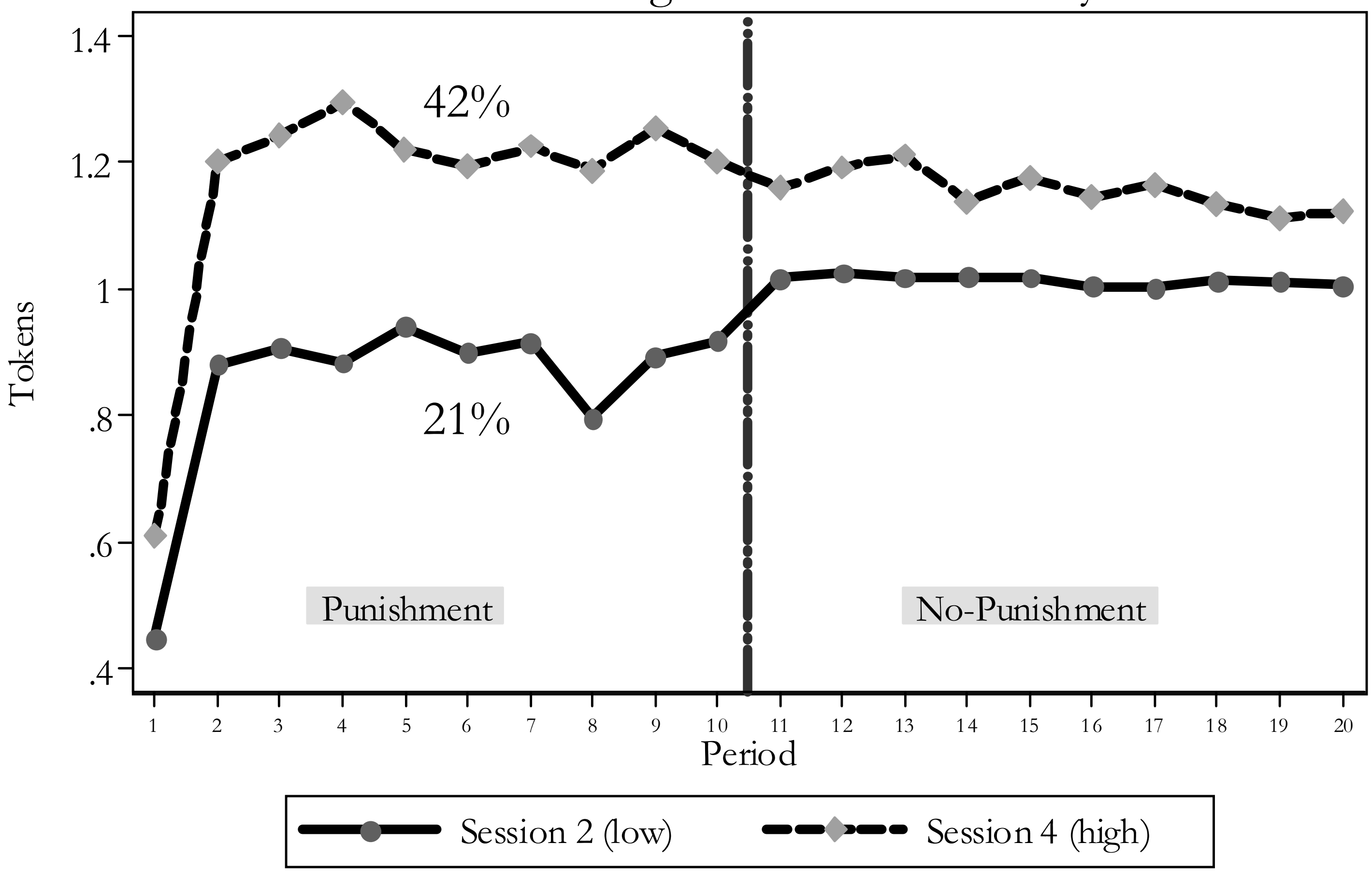


Figure 6: Average Profits With
Random Strangers and P-NP History

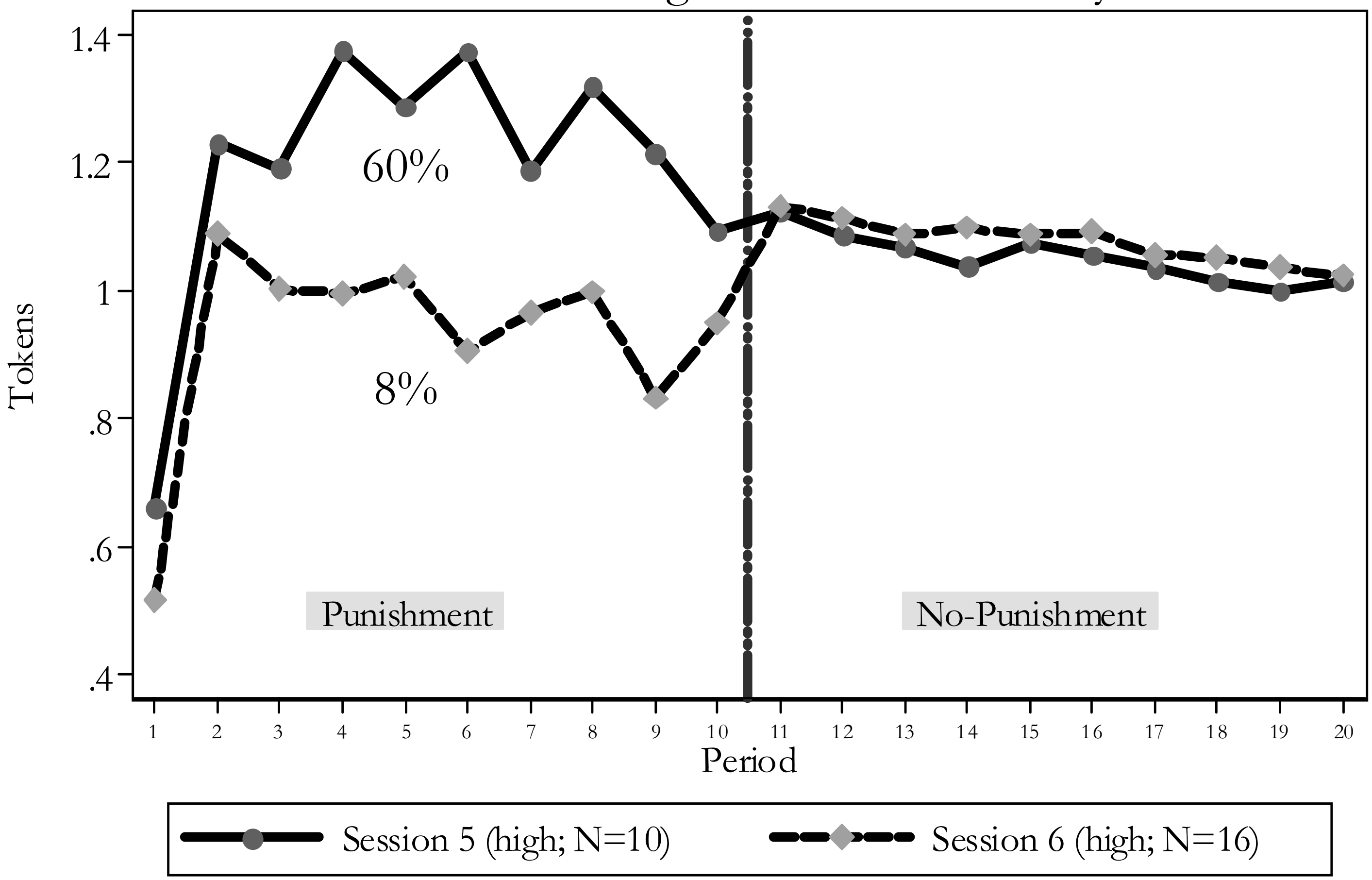


Figure 7: Average Profits With
Random Strangers and NP-P History

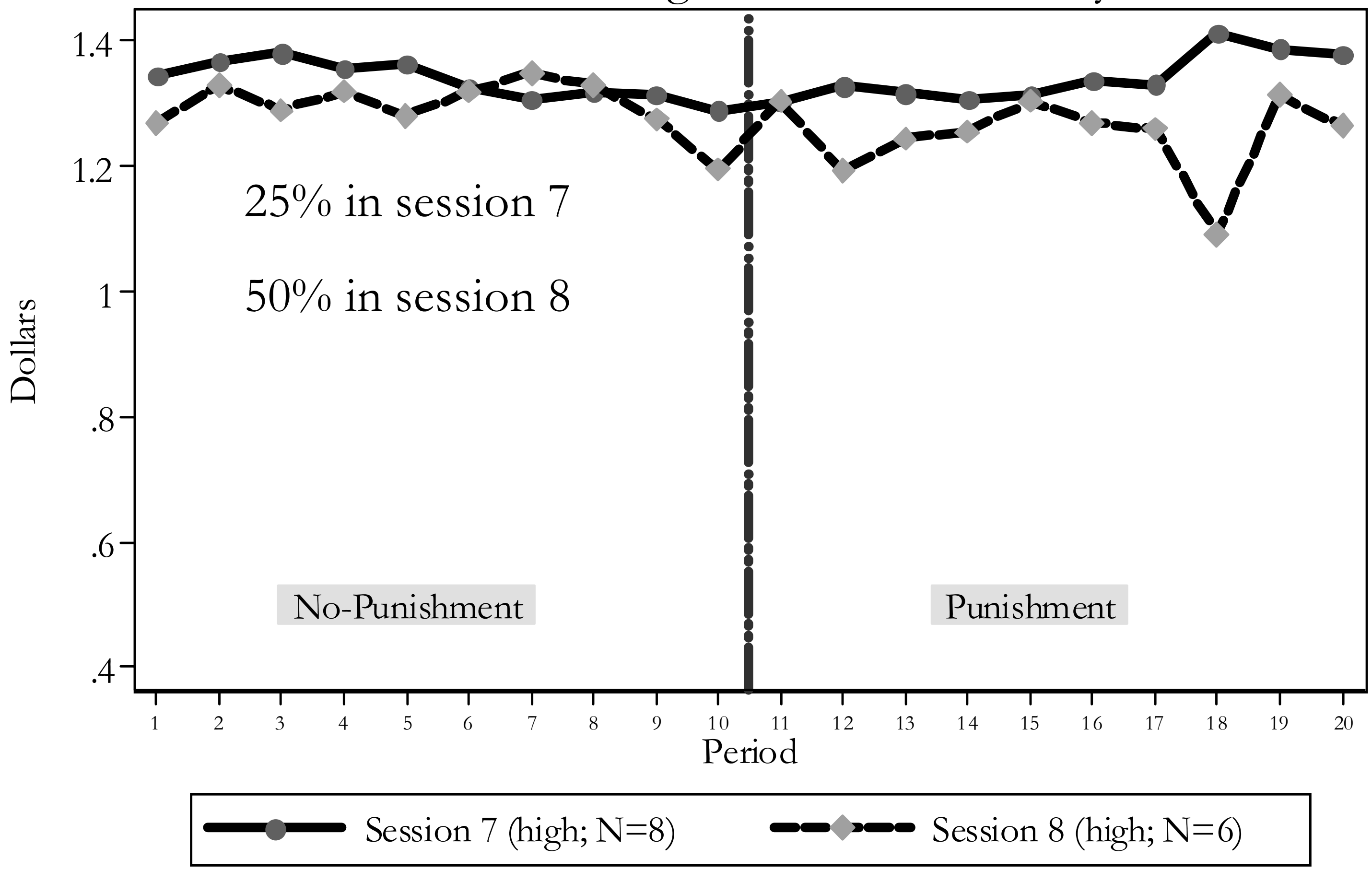


Table 3: Voting Model

Marginal effects from logit specification, with robust "Huber-White" standard errors

Variable	Estimate	Standard Error	*p*-value	95% Confidence Intervals		Variable Mean
A. Complete Sample (N=142; 79% vote for NP; Wald χ^2_{18} = 48.8; *p*-value = 0.0001)						
Pstrangers	0.683	0.246	0.006	0.200	1.166	0.718
Gsize	0.036	0.015	0.016	0.007	0.064	3.211
np_p	0.157	0.066	0.018	0.027	0.286	0.465
high	-0.125	0.058	0.033	-0.239	-0.010	0.648
ProfitRatio	0.493	0.206	0.017	0.089	0.898	1.069
ProfitSDratio	-0.054	0.023	0.018	-0.099	-0.009	1.694
Age	-0.009	0.008	0.260	-0.026	0.007	21.690
Male	-0.003	0.064	0.966	-0.129	0.123	0.669
Black	0.078	0.040	0.051	0.000	0.157	0.092
Asian	0.110	0.036	0.002	0.039	0.180	0.085
Hispanic	-0.114	0.102	0.262	-0.313	0.085	0.127
OtherRaces	-0.254	0.299	0.396	-0.839	0.332	0.049
Business	0.021	0.064	0.738	-0.104	0.146	0.430
PreSenior	0.029	0.052	0.579	-0.073	0.131	0.451
GPAlow	0.003	0.055	0.962	-0.105	0.110	0.521
GPAhigh	0.057	0.042	0.171	-0.025	0.138	0.148
HHsize	0.030	0.024	0.206	-0.017	0.076	1.620
Work	-0.011	0.047	0.820	-0.102	0.081	0.711
B. High Returns Sample (N=92; 71% vote for NP; Wald χ^2_{17} = 28.8; *p*-value = 0.036)						
Pstrangers	0.485	0.279	0.082	-0.062	1.032	0.565
Gsize	0.040	0.022	0.074	-0.004	0.083	4.957
np_p	0.128	0.112	0.256	-0.093	0.348	0.435
ProfitRatio	1.119	0.400	0.005	0.336	1.903	1.032
ProfitSDratio	-0.060	0.034	0.080	-0.126	0.007	1.459
Age	-0.002	0.015	0.869	-0.032	0.027	21.457
Male	-0.005	0.132	0.972	-0.263	0.254	0.663
Black	0.081	0.125	0.519	-0.164	0.325	0.098
Asian	0.176	0.055	0.001	0.069	0.283	0.087
Hispanic	-0.339	0.272	0.212	-0.872	0.193	0.109
OtherRaces	-0.377	0.371	0.309	-1.105	0.350	0.065
Business	-0.021	0.106	0.842	-0.228	0.186	0.467
PreSenior	0.087	0.105	0.405	-0.118	0.293	0.478
GPAlow	0.086	0.103	0.405	-0.116	0.288	0.500
GPAhigh	0.041	0.091	0.651	-0.137	0.219	0.141
HHsize	0.017	0.042	0.683	-0.065	0.100	1.587
Work	-0.041	0.072	0.569	-0.182	0.100	0.728

References

Andreoni, James, "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 37, 1988, 291-304.

Andreoni, James, and Croson, Rachel T.A., "Partners versus Strangers: Random Rematching in Public Goods Experiments," in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results* (North-Holland: Amsterdam, 2005).

Botelho, Anabela; Harrison, Glenn W.; Pinto, Lígia M. Costa and Rutström, Elisabet E.,"Testing Static Game Theory with Dynamic Experiments: A Case Study of Public Goods," *Working Paper*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Binmore, Ken, *Fun and Games* (Lexington, MA: D.C. Heath and Company, 1992).

Croson, Rachel T.A., "Partners and Strangers Revisited," *Economics Letters*, 53, 1996, 25-32.

Dawkins, Richard, *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design* (New York: Norton, 1986).

Erhart, Karl-Martin, and Keser, Claudia, "Mobility and Cooperation: On the Run," *Working Paper 99s-24*, CIRANO, University of Montreal, June 1999.

Fehr, Ernst, and Gächter, Simon, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), September 2000, 980-994.

Fehr, Ernst, and Gächter, Simon, "Altruistic Punishment in Humans," *Nature*, 415, 10 January 2002, 137-140.

Fischbacher, Urs, "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual," *Working Paper Nr. 21*, Institute for Empirical Research in Economics, University of Zurich, 1999.

Fudenberg, Drew, and Tirole, Jean, *Game Theory* (Cambridge, MA: MIT Press, 1991).

Goeree, Jacob K.; Holt, Charles A., and Laury, Susan K., "Private costs and public benefits: unraveling the effects of altruism and noisy behavior," *Journal of Public Economics*, 83, 2002, 255–276.

Harrison, Glenn W., and Hirshleifer, Jack, "An Experimental Evaluation of Weakest-Link/ Best-Shot Models of Public Goods," *Journal of Political Economy*, 97, February 1989, 201-225.

Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1013-1059.

Heckman, James J., and Smith, Jeffrey A., "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9(2), Spring 1995, 85-110.

Isaac, R. Mark and Walker, James M., "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *Quarterly Journal of Economics*, 53, 1988, 179-200.

Masclet, David; Noussair, Charles; Tucker, Steven, and Villeval, Marie-Claire, "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review*, 93(1), March 2003, 366-380.

Newey, Whitney K., "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Journal of Econometrics*, 36, 1987, 231-250.

Ostrom, Elinor; Walker, James, and Gardner, Roy, "Covenants With and Without a Sword: Self-Governance Is Possible," *American Journal of Political Science*, 86(2), June 1992, 404-417.

Page, Talbot; Putterman, Louis, and Unel, Bulent, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Working Paper 2002-19*, Department of Economics, Brown University, 2002.

Palfrey, Thomas R., and Prisbrey, Jeffrey E., "Altruism, Reputation, and Noise in Linear Public Goods Experiments," *Journal of Public Economics*, 61, 1996, 409-427.

Palfrey, Thomas R., and Prisbrey, Jeffrey E., "Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?" *American Economic Review*, 87, 1997, 829–846.

Philipson, Tomas, and Hedges, Larry V., "Subject Evaluation in Social Experiments," *Econometrica*, 66(2), March 1998, 381-408.

Yamagishi, Toshio, "The Provision of a Sanctioning System as a Public Good," Journal of *Personality and Social Psychology*, 51(1), 1986, 110-116.

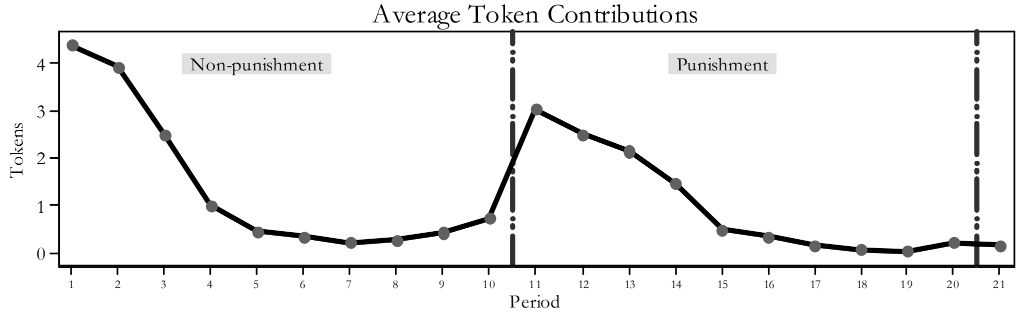
Yamagishi, Toshio, "Seriousness of Social Dilemmas and the Provision of a Sanctioning System," *Social Psychology Quarterly*, 51(1), 1988, 32-42.

Appendix: Detailed Results (NOT FOR PUBLICATION)

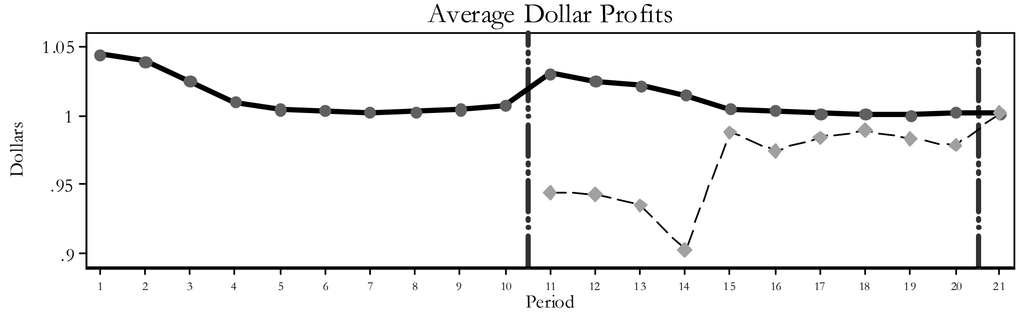
Results in Session 1 (N=26 Perfect Strangers)

Low Return to Public Good

Average Token Contributions



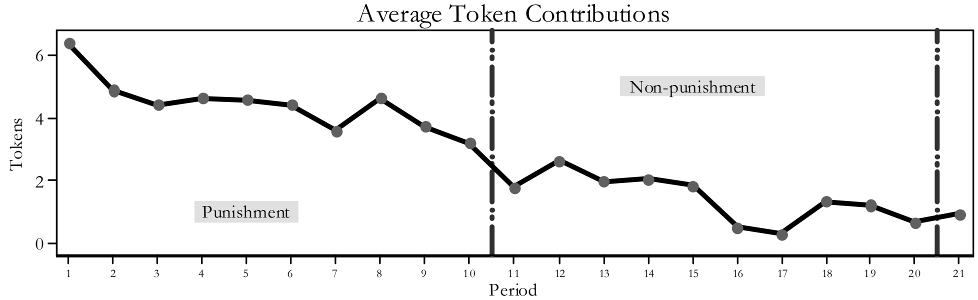
Average Dollar Profits



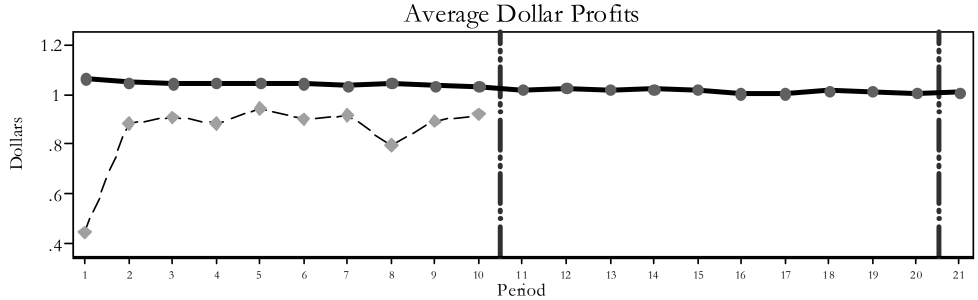
Results in Session 2 (N=24 Perfect Strangers)

Low Return to Public Good

Average Token Contributions

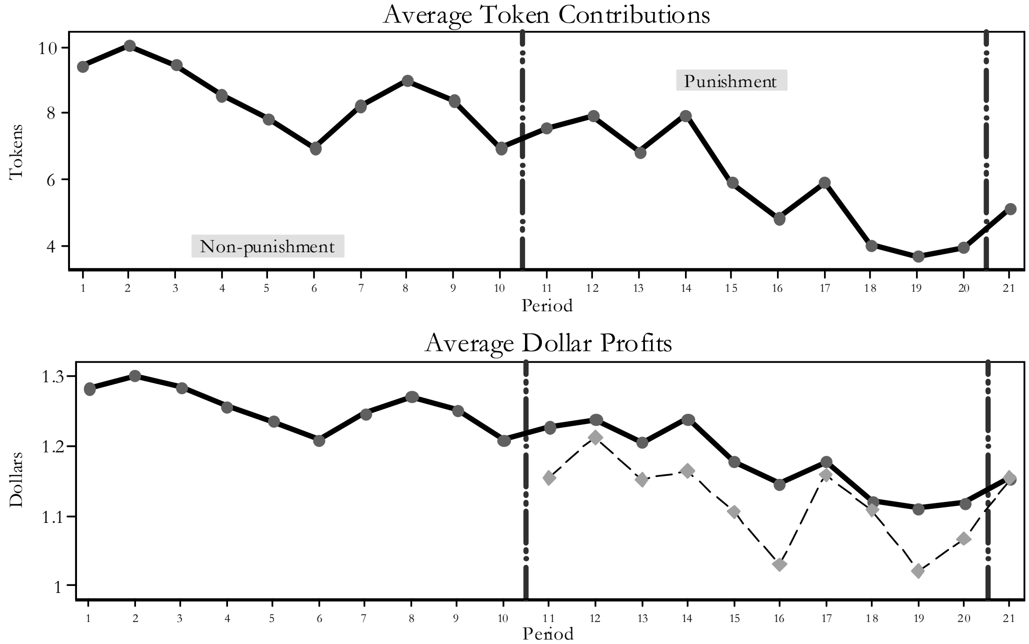


Average Dollar Profits



Results in Session 3 (N=26 Perfect Strangers)

High Return to Public Good

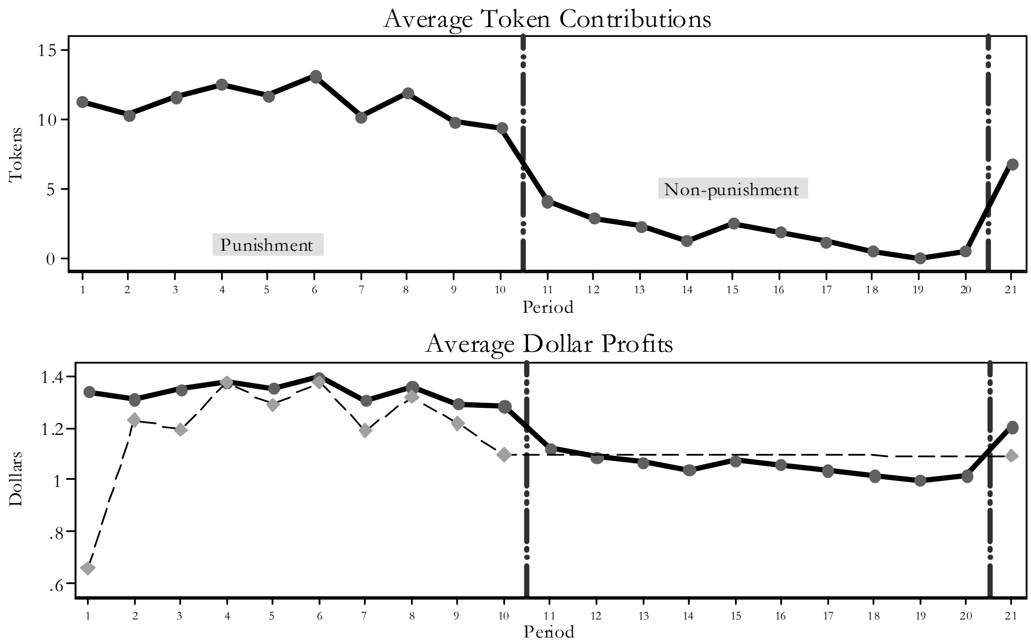


Results in Session 4 (N=26 Perfect Strangers)

High Return to Public Good

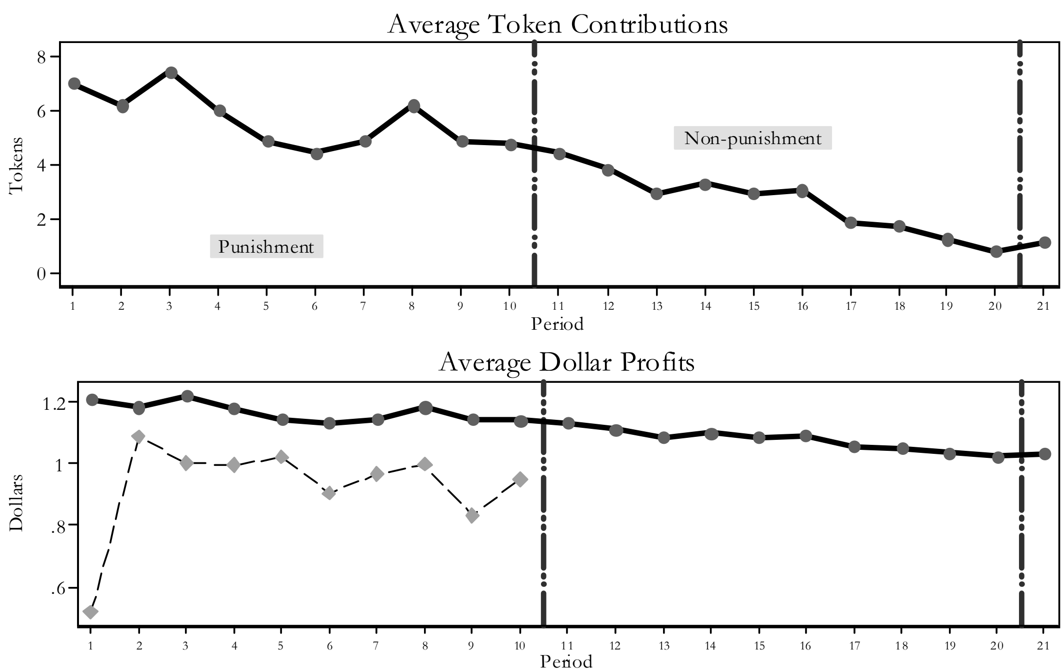
Results in Session 5 (N=10 Random Strangers)

High Return to Public Good



Results in Session 6 (N=16 Random Strangers)

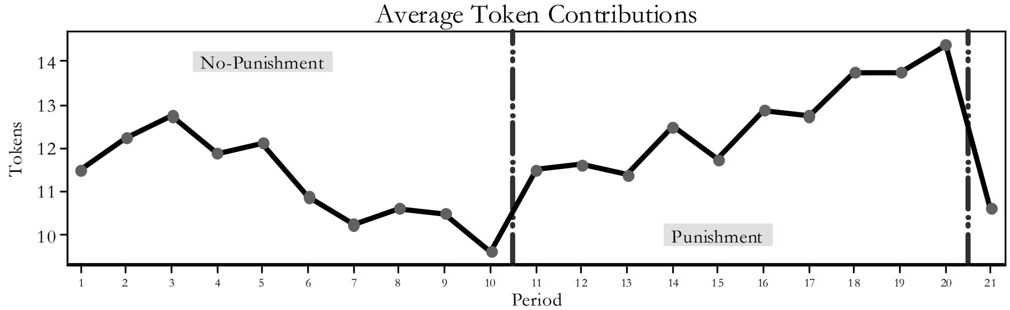
High Return to Public Good



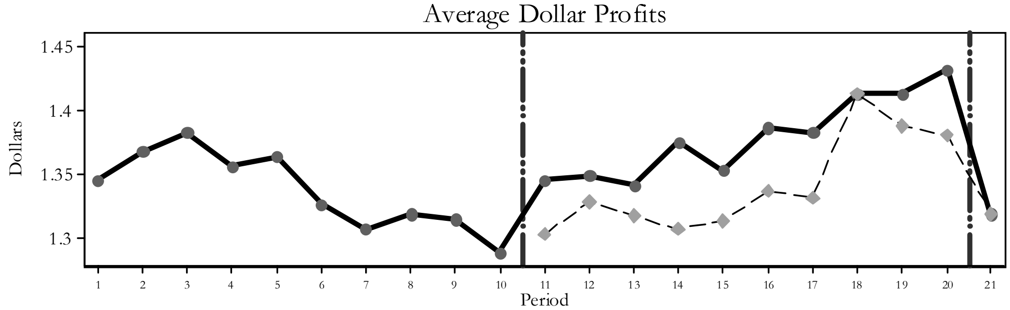
Results in Session 7 (N=8 Random Strangers)

High Return to Public Good

Average Token Contributions



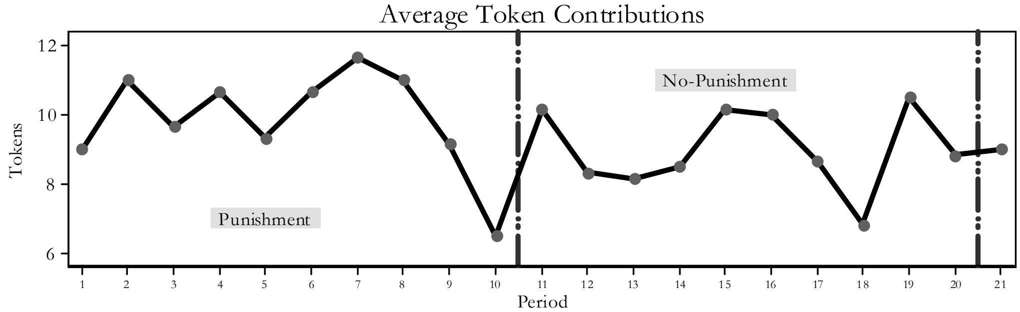
Average Dollar Profits



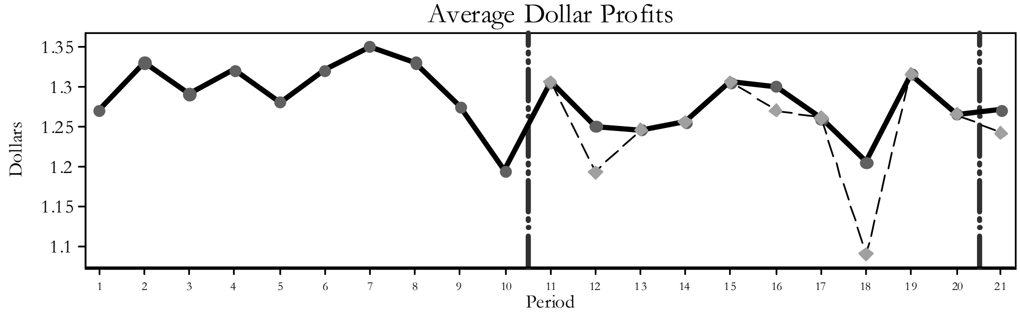
Results in Session 8 (N=6 Random Strangers)

High Return to Public Good

Average Token Contributions



Average Dollar Profits



Descriptive Statistics for Explanatory Variables in Voting Model

Variable	Obs	Mean	Std. Dev.	Min	Max
Pstrangers	142	.7183099	.4514154	0	1
np_p	142	.4647887	.5005241	0	1
high	142	.6478873	.4793196	0	1
ProfitRatio	142	1.06927	.2901981	.7128713	3.958333
ProfitSDRa~o	141	1.6939	1.889186	.0844262	18.41636
Age	142	21.69014	2.770624	18	36
Male	142	.6690141	.4722337	0	1
Black	142	.0915493	.2894095	0	1
Asian	142	.084507	.2791313	0	1
Hispanic	142	.1267606	.3338823	0	1
White	142	.6478873	.4793196	0	1
Business	142	.4295775	.4967681	0	1
PreSenior	142	.4507042	.4993253	0	1
GPAlow	142	.5211268	.5013218	0	1
GPAhigh	142	.1478873	.3562449	0	1
HHsize	142	1.619718	1.134453	1	6
Work	142	.7112676	.4547774	0	1

Working Papers - NIMA series

No.

1. Lígia Pinto , Glenn Harrison, *Multilateral negotiations over climate change policy*, May 2000

2. Paulo Guimarães, Douglas Woodward, Octávio Figueiredo, *A tractable approach to the firm location decision problem*, May 2000

3. Miguel Portela , *Measuring skill: a multi-dimensional index*, September 2000

4. Rosa Branca Esteves , Paulo Guimarães, *Price discrimination and targeted advertising: a welfare analysis*, November 2000

5. Anabela Botelho , Lígia Pinto , *Has Portugal gone wireless? Looking back, looking ahead*, December 2000

6. Pedro Barros, Clara Dismuke , *Hospital production in a national health service: the physician's dilemma*, December 2000

7. Anabela Botelho , Mark A. Hirsch, Elisabet E. Rutström, *Culture, nationality and demographics in ultimatum games*, December 2000

8. Miguel Portela , *The impact of segregation on wage inequality: a look at recruitment and pay policies at the firm level*, January 2001

9. Pedro Portugal, Ana Rute Cardoso, *Disentangling the minimum wage puzzle: an analysis of job accessions and separations from a longitudinal matched employer-employee data set*, April 2001

10. Ana Rute Cardoso, Priscila Ferreira , *The dynamics of job creation and destruction for University graduates: why a rising unemployment rate can be misleading*, May 2001

11. Octávio Figueiredo, Paulo Guimarães, Douglas Woodward, *Asymmetric information and location*, July 2001

12. Anabela Botelho , Lígia Pinto , *Hypothetical, real, and predicted real willingness to pay in open-ended surveys: experimental results*, September 2001

13. Anabela Botelho , Lígia Pinto , Miguel Portela , António Silva, *The determinants of success in university entrance*, September 2001

14. Anabela Botelho , *Strategic behavior at trial. The production, reporting, and evaluation of complex evidence*, September 2001

15. Paulo Guimarães, *The state of Portuguese research in economics: an analysis based on publications in international journals*, September 2001

16. Anabela Botelho , Glenn Harrison, Marc Hirsch, Elisabet E. Rutström, *Bargaining behavior, demographics and nationality: a reconsideration of the experimental evidence*, December 2001

17. João Cerejeira da Silva , *Identification of the Portuguese industrial districts*, February 2002

18. Octávio Figueiredo, Paulo Guimarães, Douglas Woodward, *Modeling industrial location decisions in U.S. Counties*, April 2002

19. Aslan Zorlu , Joop Hartog, *The effect of immigration on wages in three European countries*, October 2002

20. Elvira Lima , David K. Whynes, *Finance and performance of Portuguese hospitals*, February 2003

21. Aslan Zorlu , *Do ethnicity and sex matter in pay? Analyses of 8 ethnic groups in the Dutch labour market*, June 2003

22. Cécile Wetzels , Aslan Zorlu , *Wage effects of motherhood: a double selection approach*, June 2003

23. Natália Barbosa , *What drives new firms into an industry? An integrative model of entry*, October 2003

24. Elvira Lima , Teresa Josefina Lopes Esquerdo , *The economic costs of alcohol misuse in Portugal*, October 2003

25. Anabela Botelho , Lígia Pinto , Isabel Rodrigues , *How to comply with environmental regulations?The role of information*, October 2003

26. Natália Barbosa , Helen Louri, *Corporate performance: does ownership matter? A comparison of foreign - and domestic - owned firms in Greece and Portugal*, October 2003

27. Anabela Botelho , Lígia Pinto , *Students' expectations of the economic returns to college education. Results of controlled experiment*, December 2003

28. Paula Veiga , *Income-related health inequality in Portugal*, July 2005

29. Anabela Botelho , Glenn Harrison, Lígia Pinto , Elisabet E. Rutström, *Testing static game theory with dynamic experiments: a case study of public goods*, November 2005

30. Anabela Botelho , Glenn Harrison, Lígia Pinto , Elisabet E. Rutström, *Social norms and social choice*, November 2005

31. Anabela Botelho , Glenn Harrison, Lígia Pinto , Elisabet E. Rutström, Paula Veiga , *Discounting in developing countries: a pilot experiment in Timor-Leste*, November 2005

32. Paula Veiga , Ronald P. Wilder , *Maternal smoking during pregnancy and birthweight - A propensity score matching approach*, January 2006

The Working Papers of the Applied Microeconomics Research Unit (NIMA) can be downloaded in PDF format from http://nima.eeg.uminho.pt