Mislabelling in Collective Labels: an experimental analysis

Olivier Bonroy (INRA, Université Grenoble Alpes), Alexis Garapin (Université Grenoble Alpes), Stephen F. Hamilton (California Polytechnic State University) and Diogo M. Souza Monteiro (Newcastle University)

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-2

Mislabelling in Collective Labels: an experimental analysis

Olivier Bonroy[1,2], Alexis Garapin[2,1], Stephen F. Hamilton[3], and Diogo M. Souza Monteiro[4]

[1]INRA, UMR 1215 GAEL, F-38000 Grenoble, France.
[2]Université Grenoble Alpes, UMR 1215 GAEL, F-38000 Grenoble, France.
[3]Department of Economics, California Polytechnic State University, San Luis Obispo, CA 93407, USA
[4]School of Agriculture, Food and Rural Development, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK

May 26, 2015

Abstract

In food markets the use of labels with collective reputations is widespread. Examples are organic or fair trade labels. These labels signal credence attributes of products which cannot be verified by consumers. Producers may mis-label their products to collect the premium associated to such collective reputations. We examine the propensity for fraud to emerge in a multi-player, collective action game. An important feature of our framework is that the premium depends on the proportion of mis-labeling in the market. We experimentally test the impact of market size, alternative monitoring and subsidy on the propensity to mis-label. In line with our theoretical predictions we find that more stringent monitoring and enforcement schemes and the presence of a subsidy reduce the propensity to mislabel, while an increase in market size elevates it.

Keywords: label, public policy, collective action game, experimental economics, agricultural economics.

JEL classification: C72, C92, Q18

1 Introduction

For many goods, consumers rely on product labels to inform them on the presence of valuable attributes that would otherwise be unobserved. The economic role of product labeling to signal attributes that are imperceptible in consumption has been extensively analyzed from theoretical, empirical and policy perspectives (see e.g. Bonroy and Constantatos, 2015 for a survey). In many cases, consumers products containing informative labels are produced in supply chains that lack traceability back to individual suppliers, for instance free range beef, sustainable timber, organic fruits and vegetables, and "fair trade" coffee. Yet, the potential for untruthful reporting under circumstances of collective labeling, and the consequences for market outcomes, is a subject that has not received much empirical attention to date. As demonstrated by Hamilton and Zilberman (2006), the incentive of an individual producer to mis-label products with collective reputations depends on the cost of producing the attribute relative to the cost of disguising it and successfully sell it in the market. This interpretation of collective label as a public good subject to free-riding behavior suggests that the equilibrium level of mis-labeling in a market should increase with the number of producers sharing the label, a finding that has important implications for the design of labeling programs. For example, labeling behavior may be more truthful in settings with decentralized, local "region of origin" labels than in larger, centralized programs. In this paper, we study incentives to truthfully report product attributes using an experiment design that allows us to examine labeling incentives in settings with collective reputation as the number of sellers in the market changes.

In principle, it is possible to curtail or prevent fraud in markets through the introduction of monitoring and enforcement schemes. However, these activities are costly, consequently it is important to determine to what extent market participants sharing a collective label can self-regulate mis-labeling.

We examine the propensity to mis-label in a multi-player, collective action game. This class of game analyses players' incentive to contribute to a collective action under circumstances where free riding is possible (i.e. not to contribute to the collective action but still benefit from it). We frame our experiment around producers' incentive to contribute or not to collective labels. A particular feature of our game is that producers have randomized payoffs from free-riding behavior. We find only weak incentives to free-ride, and hence limited need to invest in monitoring and enforcement activities, when the number of players in the game is "small". However, when a larger number of players is involved, we show that a coordination problem arises in a multi-player assurance game. In this case two Nash equilibria emerge, one in which all players use the label properly and one in which all players free-ride. We empirically investigate these results in a controlled, experimental study that focuses on how the randomization of payoffs from free riding behavior affects players' ability to coordinate actions.

Our experimental results show statistically significant support for an increase of the propensity to fraud when the group size increases from 4 players to 8 players. We also find that for our groups of 8 players, there is statistically significant evidence that raising the monitoring and enforcement levels reduces the propensity to free-ride.

Our findings have important implications for markets with labeling fraud. In an early paper on fraudulent labeling, Baksi and Bose (2007) show that third-party labeling may not be socially optimal in an environment where a government inspects products. The reason is that third party labeling raises the cost of providing high-quality attributes, leading to higher prices for certified attributes in the market and larger incentives for fraudulent labeling. Third-party labeling can thus raise monitoring and enforcement costs by forcing regulators to incur greater costs of inspection. Hamilton and Zilberman (2006) analyze the fraud strategy of producers that adopt a certified label. Firm adopting the certified label can disguise their product as containing the certified attribute even if these products do not in fact meet the certification standard. Under such circumstances, a positive per-unit certification cost can serve to reduce the degree of fraud in the market by narrowing the relative cost of producing truthful products instead of fraudulent products. Contrary to Baksi and Bose (2007), the cheating producer also pays the certification cost in Hamilton and Zilberman (2006), which reduces the marginal return from cheating.

Our experimental framework includes features of a public good game (Ledyard, 1997 and Chaudhuri, 2011, for surveys), but also involves coordination between players (Camerer, 2003 for a survey). Thus, our experimental setting is more general as it encompasses several classes of multi-player collective action games depending on the level of randomization of returns to free-riding. Specifically, we have a public good game when payoffs are certain and a multi-player assurance game when payoffs are randomized. Early experiments on public goods reviewed by Ledyard (1997) typically show that in one-shot versions of the public goods game, there is much more contribution than predicted in the Nash equilibrium of the game; however, there is wide variation in the level of individual contributions. When players interact repeatedly over a number of rounds, contributions tend to start near the social optimum and decline steadily over time as more and more players choose to "free ride". The higher the number of players interacting, the higher the number of free-riders. As far as we know, only Engel and Zhurakhovska (2014) randomizes the player's own payoff in a public good game. In comparison to a baseline treatment where payoffs to contribution are certain, Engel and Zhurakhovska (2014) report significantly less cooperation when such payoffs are randomized, but only when the probability is low enough.[1]. Yet, our approach departs from theirs as we randomized payoffs related to free-riding actions rather than to contribution actions.

To sum up, our contribution to the experimental literature is threefold. First, to our knowledge, no other study has experimentally investigated the impact of randomized payoffs on free riding behavior in a collective action game. Randomizing payoffs that result from free-riding behavior changes the nature of the game from a public good game to a multi-player assurance game for some probability value, which offers a more general experimental setting. The coordination problem generated by randomizing payoffs, by itself, provides for a novel experimental design. Second, while most experiments investigate sanction mechanisms by pairs under conditions of perfect observability or else randomize payoffs to contribute, our "monitoring" mechanism leads to *randomization in payoffs to free-ride*. This design is appropriate in settings where gains arise through free-riding on collective reputation. Third, while the experimental economics literature has devoted much attention to elicit consumers' willingness to pay for labeled products (Cason and Gangadharan, 2002; Michaud *et al.*, 2013; Kiesel and Villas-Boas, 2013), there is so far no experimental study of the labeling strategies of the firms.

The remainder of the paper is structured as follows. In section 2 we present the basic set up of our game. In section 3, we describe the experimental design and procedure. In section 4 we present the experimental results. Finally, we conclude.

2 The basic setup: a multi-player collective-action game

We consider a game with N players. Each player simultaneously chooses between two actions, A and B, which have corresponding cost c_A and c_B, respectively. We assume

[1]As mentioned above, we designed this randomized payoff scheme to capture to some extent a "monitoring" mechanism. Experimenters investigated monitoring mechanisms in public good settings, but when monitoring is *enforced by the subjects themselves*, when they can observe others' shirking (Fehr and Gachter, 2000; Sefton *et al.*, 2007; Ertan *et al.*, 2009). When there is only some probability that the monitoring is enforced, which is a mechanism that is closer to ours, experimental results show that the higher the probability, the higher the contribution levels (Walker and Halloran, 2004; Sousa, 2010)

$c_A > c_B = 0$. Choosing action A generates a fixed revenue $p \geq c_A$ and an additional payoff $\gamma = \rho(n)\, a$, where $a > c_A$ and $\rho(n)$ is a premium arising from collective action and $\frac{d\rho(n)}{dn} > 0$. Specifically, we formulate our collective action premium using the function $\rho(n) = n/N$, where n denotes the number of players choosing action A. The payoff to choosing B is the same as for A but no costs are incurred ($c_B = 0$).

The incentive structure of the game is as follows. As more players choose action A, their payoff increases through the premium γ. That is, each player choosing A has the following return:

$$C(n) = p + \rho(n)\, a - c_A \tag{1}$$

Each time a player chooses action B, the payoff is

$$F(n) = p + \rho(n)\, a > C(n). \tag{2}$$

As more players engage in action B, the smaller the payoff from action A, as the $N - n$ players opting for action B reduce the collective premium, $\rho(n)$. Because the market cannot distinguish between individuals taking actions A and B, players choosing B receive a higher payoff than players choosing action A by avoiding the additional cost, c_A.

For each player, the decision to choose A or B depends on the decisions of the other $(N-1)$ players, specifically on the number of players in the game who choose action A. To see this, consider a "late-arriving" player who can decide the action to take after observing the action selected by every other player and suppose the late-arriving player observes m other players that select action A. If the late arriving player chooses A, the number of players that coordinate on action A becomes $m+1$ and she receives $C(m+1)$. Conversely if she chooses action B, her payoff is given by $F(m)$. Therefore, the final decision of a late-arriving player would be driven by the comparison between these two payoffs. When $C(m+1) > F(m)$, a payoff-maximizing player would opt for action A and otherwise choose B. Note that action A satisfies the convention properties of a pure public good, as the benefits are both non-excludable and non-rival (Dixit *et al.*, 2009).

Our game thus encompasses the coordination problem in collective action games. The welfare of the collective is higher when its members choose A, but this action may not be in the best interest of each individual. In other words, the socially optimal outcome may not be the Nash equilibrium of the game. Specifically, when the number of players is sufficiently small, $N < \frac{a}{c_A}$, the choice of A is the dominant strategy regardless of what the other $(N-1)$ players are doing: $C(m+1) > F(m)$, $\forall m \in [0, N-1]$. In this case, the Nash equilibrium coincides with the social optimum.[2] However, when $N > \frac{a}{c_A}$ opting for B becomes the dominant strategy: $C(m+1) < F(m)$, $\forall m \in [0, N-1]$. In this case, the game becomes a prisoners' dilemma game that results in a suboptimal Nash equilibrium. Thus, our framework nests two alternative classes of games depending on the number of players.

Consider now the case in which the player's payoff from action B is random such that it is 0 with a probability $(1-r)$. In this way the condition of non-excludability is relaxed such that action A is no longer a pure public good. The final decision of a player is driven by the comparison of the following payoffs:

$$\begin{cases} C(m+1) = p + \rho(m+1)\, a - c_A & \text{when the player chooses } A \\ F(m) = r(p + \rho(m)\, a) & \text{when the player chooses } B. \end{cases} \tag{3}$$

As in the previous case, when $N < \frac{a}{c_A}$ then $C(m+1) > F(m)$ whatever the feasible values of m and r, action A is the dominant strategy and the social optimum is achieved. When $N > \frac{a}{c_A}$, opting for A is still the dominant strategy, but only when r is sufficiently low. That is, A is the dominant strategy when $r < r_1(N) = \frac{a+(p-c_A)N}{pN}$ and the social optimum is still feasible. However, when r is sufficiently high (i.e. superior to $r_2(N) = \frac{(a+p-c_A)N}{(a+p)N-a}$), the game becomes a prisoners' dilemma game. Action B is then the dominant strategy, and the social optimum is not achieved. Interestingly, randomizing the player's return to choosing B introduces a new and third class of game. The reason for this is that for $N > \frac{a}{c_A}$ and $r \in [r_1(N), r_2(N)]$ the game has two equilibria: the socially optimal outcome and the prisoner's dilemma outcome.[3] This third class of game is a multiplayer assurance game with two pure-strategy equilibria: the dominant strategy is B for $n \in [0, \tilde{n}(N)]$ and A for $n \in [\tilde{n}(N), N-1]$, with $\tilde{n}(N) \equiv \frac{N(c_A - p(1-r)) - a}{a(1-r)}$ (see Dixit *et al.*, 2009).[4]

[2] Note that this requires $a > c_A$.

[3] $0 < r1(N) < r2(N) < 1$ for $N > \frac{a}{c_A}$.

[4] In this way for any value of n superior to $\tilde{n}(N)$, each player will want to contribute and will choose action A: there will be a Nash equilibrium at $n = N$ where all players contribute. Conversely, for any value of n inferior to $\tilde{n}(N)$, each player will want to act as a free rider and will choose action B: there will be a second Nash equilibrium at $n = 0$ where no player contributes. Technically, there is also a third Nash equilibria for $n = \tilde{n}$ in wich some players contribute and some act as free riders. This situation could be an equilibrium only if $\tilde{n}$ is exactly right but is strongly unstable, such as only two stable Nash equilibria hold

The following Lemma, summarizes the previous discussion:

Lemma 1. *Whenever* $N < \frac{a}{c_A}$ *and* $r \in [0,1]$ *all players choose action* A *at equilibrium. Whenever* $N > \frac{a}{c_A}$ *three equilibrium configurations may hold: i)* $\forall r \in [0, r1(N)]$ *only one Nash equilibrium where all players choose action* A *occurs, ii)* $\forall r \in]r1(N), r2(N)[$ *two Nash equilibria where all players take either action* A *or action* B *occur, and iii)* $\forall r \in [r2(N), 1]$ *only one Nash equilibrium where all players choose action* B *occurs.*

Let us now focus on the multiplayer assurance game. We found that for all $N > \frac{a}{c_A}$ then $\frac{\partial r1(N)}{\partial p} = \frac{Nc_A - a}{Np^2} > 0$ and $\frac{\partial r2(N)}{\partial p} = \frac{N(Nc_A - a)}{(-a+Na+Np)^2} > 0$, with $\frac{\partial r1(N)}{\partial p} > \frac{\partial r2(N)}{\partial p}$. We state:

Lemma 2. *Whenever* $N > \frac{a}{c_A}$ *increasing* p *enables i) to increase the region of* r *where the social optimum is achieved, ii) to reduce the region fo* r *where the game is a a prisoners' dilemma game, and iii) to reduce the region of* r *where the game is a multiplayer assurance game.*

3 Experimental design and procedure

To empirically test our Lemmas we resorted to experimental methods. This section starts with a description of our experimental treatments and parameterization of our theoretical model. Then we provide the equilibrium predictions that will be tested in the experiment. Finally we describe in detail the experimental procedure.

3.1 Experimental treatments and parameters

Our experiment reproduces in the lab the setting of the collective action game presented above. Subjects were told that they had to choose between "*kappa*" (*action A* in the model) or "*phi*" (*action B* in the model) [5] and that all the members of their group had to make this choice simultaneously. Subjects were told how many subjects there where in their group but they didn't know who they were nor were allowed to communicate with other subjects during the sessions. Each subject played the game during 20 periods (the number of periods was common information in the experiment). As the subject's risk

(see Dixit *et al.*, 2009)

[5] To be coherent with the frame of the model, we will go on referring in the paper to *action A* and *action B*, even if they were labeled *Kappa* and *Phi* in the experiment

preference is a variable that might affect the decision to free ride in the game, we controlled subjects' attitudes towards risk in the experiment: the subjects performed the Holt and Laury (2002)'s risk test before and after the collective action game to control possible order effects.[6]

Subjects were given a payoff table and encourage to use them before making their choices. This table showed the payoff for the period when choosing A or B, according to the number other members in their group deciding to play A or B. To build the payoff tables of the experiment, we used the following parameter values: $p = c_A = 2$, $c_B = 0$, $a = 10$ such that:

* for $N = 4$ and p=2, only one Nash equilibrium where all players adopt action A occurs

* for $N = 8$ and p=2, the partition of equilibrium is defined by $r1 = 0.625$ and $r2 = 0.930$.

Thus, to test the effects or r we select the following values of r: 0.55, 0.75, 0.95 and 1. Note that with these values we cover all the partitions of equilibrium considered in the game. Since Lemma 1 suggests that the number of subject and the effects of r affect the optimal strategy, using a between subjects procedure we designed 8 experimental treatments according a 2x4 factorial design that crosses the size N of the group of subjects ($N = [4; 8]$) and the probability r of retaining the payoff to fraud ($r = [1; 0.95; 0.75; 0.55]$). For instance, the treatment [4;0.75] corresponds to experimental sessions conducted with groups of 4 subjects with a probability of 0.75 of getting a positive payoff (and 0.25 of getting a null payoff) for B. Treatments [4;1] and [8;1] can be considered as "benchmark" treatments. Since Lemma 2 suggests that increasing

The payoff tables (see examples in the appendix section) displayed all the possible payoffs for every combination of self's and others' choices[7] In treatments in which the

[6]This test is a menu of 10 paired lottery choices (sets of two options, one of which has to be chosen by the subject), designed to make inferences about risk preferences under various payment conditions. The subjects can choose the safe option when the probability of obtaining the higher payoff is small, and then cross over to the risky option without ever going back to the safe one. The number of safe choices made by the subjects (before the switch to the risky one) determines their risk attitude. If subjects choose four safe options (i.e. the switch occurs at the fifth set of choices), they are risk neutral; fewer than four signals risk attraction, and more than four risk aversion.

[7]The real payoffs were displayed only when the play of the game started. During the reading of the instructions, the payoffs in the tables were symbolized by letters.

payoff to *action* B was randomized, the payoff table displayed that the choice of B could lead to a positive amount with r probability, or a null payoff with $(1-r)$ probability. For instance, in treatment [8;0.55], subjects were told that if they chose B they had a 0.55 chance of getting some positive payoff in Ecus and a 0.45 chance of getting 0 Ecu (there was no wording such as "probability of being monitored" or "probability of destruction"). The resulting expected payoff for the choice of B was not calculated and displayed in the payoff table, to avoid biases in subjects' risk preferences. The subject's payoff was her cumulative payoff over all the periods. Each subject was credited each period with the amount of Ecus corresponding to the number of A and/or B choices in her group (including her own choice). After each period, a subject was informed of her decision, the repartition of A and/or B choices in her group (including her own choice), her own payoff and her cumulated playoff since the beginning of the experiment. Each subject could see the history of all previous periods with such information on choices and payoffs throughout the experiment.

3.2 Predictions

Based on both Lemmas 1 and 2 the equilibria predictions of the game played are:

Prediction 1. ***The effect of N***
Increasing number of players increases the free-riding. In this way i) For $N = 4$ all players will choose action A whatever the level of r. ii) For $N = 8$ all players will choose action A (respectively action B) whenever $r \leq r1 = 0.625$ (respectively $r \geq r2 = 0.930$). Otherwise both previous equilibrium configurations are possible.

In accordance with this prediction, when 4 subjects play together they should choose action A whatever the level of r. But when there are 8 subjects in the group they should not choose action A for high levels of return to fraud (i.e., $r = 0.95$ and $r = 1$). Thus, we postulate that increasing the number of players reduces the incentive to participate to the action A.

Prediction 2. ***The effect of r***
For intermediate values of r, players are confronted to a coordination problem on the equilibrium choice that modify the incentive to free-ride. In this way for $N = 8$, $r = 0.75$ and $p = 2$ the players' choice to free-ride is not the same as when there is no coordination problem (i.e., for $r \leq r1 = 0.625$ or $r \geq r2 = 0.930$.)

This prediction postulates that for intermediate values of r, the players cannot coordinate on an equilibrium such as the free-riding is not the same as only one equilibrium holds. Lemma 2 establishes that increasing p increases the region of r where the social optimum is achieved. In order to test this result a treatment [8;0.75] has been conducted by considering $p = 2c_A = 4$. For such a value of p the partition of equilibrium is defined by $r1 = 0.813$ and $r2 = 0.941$. Thus we have:

Prediction 3. ***The effect of p***
Increasing the fixed revenue p related to the action A may reduce the free-riding. In this way, for $N = 8$, $r = 0.75$ and $p = 4$ all players participate to the Action A conversely to the treatment where $N = 8$, $r = 0.75$ and $p = 2$.

This prediction postulates that increasing the fixed revenue p related to the action A may enable to avoid the coordination problem on the equilibrium choice and thus to reduce the incentive to free-ride.

3.3 Experimental procedure

The experiment, which was entirely computerized, was conducted at the experimental laboratory of the GAEL research center in Grenoble. When they arrived in the laboratory, subjects received a personal code both to preserve their anonymity and to log into the software dedicated to the experiment. They were randomly assigned places in the room. Each session of the experiment corresponded to a treatment. Our laboratory has 16 working stations, thus in each session, subjects were randomly and anonymously matched in four groups of 4 players or two groups of 8 players. As we chose a between subjects design, subjects participated in just one treatment and were told that the group composition would remain fixed throughout the session and that no interaction between the groups would be possible.

Once subjects reached their place, they found an envelope on their table, containing a show up fee of 5 euros. Before the actual experiment started, the experimenter read the instructions aloud to the subjects. In addition, they were able to read these instructions on their individual screen. It was made clear that the instructions were identical for all the participants. To ensure complete understanding, subjects were given a questionnaire on the meaning of the variables, profit calculations, etc. The questionnaires were corrected with the experimenter before the experiment started. Participants had complete information on their own and others' payoffs. The payoff tables corresponded to the ones depicted in previous section. Recall that the subjects repeated the game 20 times and they were told so in the instructions.

In addition to the attendance fee and in order to make the decisions non-hypothetical, the subjects were informed at the beginning of each session that at the end they would anonymously be paid an amount in cash depending on their decisions and the decisions of others. Subjects could earn between 8.4 and 28.7 Euros (including their show up fee, their payoff for the collective action game and their payoff for the risk elicitation task).[8] The currency was the Ecu during the experiment and the exchange rate was 0.80 Euro=1000 Ecus.

At the end of the experiment, subjects filled in a small questionnaire asking them basic profiling information (male/female, university, diploma prepared). Then subjects could see their total payoff for the experiment. Finally, they were called one by one in a separate room to receive privately their money in cash and were free to leave the lab. The payoff tables are available in the appendix section. Each session lasted approximately 75 minutes, including time devoted to the subjects' payment.

To test predictions 1 (effect of N) and 2 (effect of r), we conducted 16 experimental sessions from February 2014 to May 2014, with a total of 244 subjects (Female: 155, Male: 89). To test the prediction 3 (effect of p, with $p = 2c_A = 4$), which concerns only groups of N=8 subjects, we conducted in March-April 2015 an additional set of 8 experimental sessions with a total of 128 subjects (Female: 89, Male: 39). In this new set of experiments, we performed treatments with r=0.75 but also with r=0.55.[9] Subjects were undergraduates from different universities (arts, sciences, social sciences, engineering schools) with no background in game theory. The data about each session and the number of observations are given in Table 1. Here, as we chose a fixed partner matching, each group can be considered as a statistically independent observation. The experiment was

[8] In order to get their payoff for the Holt and Laury's test, a computer program enabled subjects to throw a 10-sided die twice: the first time to determine the relevant lottery, and the second time to determine the payoff for the chosen action. This procedure of payment was carried out at the end of the experiment, to ensure that the subjects' behaviors in the game were not influenced by their earnings in the risk test. Payoffs to lotteries were labeled in Euros and were identical to those of Holt and Laury (2002).

[9] The prediction 3, in its general formulation, states that an increase in the premium p should reduce the propensity to free-ride. Thus, even it is especially interesting to assess this effect for intermediate levels of monitoring (here, r=0.75) because it may solve the coordination problem raised by such levels, it also makes sense to run experiments with r=0.55, eventually to control whether such effect can be observed even for high rates of monitoring.

conducted to get at least 4 observations per treatment.

4 Results

We start this section with an overview of the data and then we present the results from our panel data econometric estimations with which we further assess the predictions formulated in section 3.2.

4.1 Descriptive results

Overview. Our model and predictions distinguish two situations associated to the number of subjects in the group. To start, we visualize the average choice of B, across periods and for the sessions where there were 4 and 8 players. Figure 1 below shows these results. There are a few points worth noticing in this picture: first, consistently with our theoretical predictions, it can be seen that the size of the group leads to two distinct outcomes: first subjects do free ride more when involved in groups of 8. Second, there seems to be more variation on the average choice of B when N=8, than when there are fewer subjects in the game. In the latter case, note that the average choice of B becomes quite stable across treatments. A third point is that only a minority of subjects play B when there are 4 players in the game, whereas the proportion is close to 50% when there are 8 players. This provides preliminary evidence that the randomization of payoffs hinders the ability of subjects to coordinate their actions.

Along with the size of the group, a key element of our model is the probability $1 - r$ of loosing the payoff when choosing B. In *Prediction 1*, we suggest that when there are four players in the game, players should play A and r shouldn't matter. Figure 2 seems to support this claim at least in the final periods when the confidence intervals do seem to overlap when r=1, r=0.95 or r=0.75. What is also worth noticing, is that the decision to play B does decrease with the raise of the probability of payoff loss. Continuing in this vein, Figure 3 shows the effect of r for the case where there are eight players. There it is quite clear that with higher levels of r, especially when it is 0.55, the frequency of playing B decreases, thus providing initial support for *Prediction 1*, that states that subjects should play A whenever $r \leq r1 = 0.625$.

Treatment effects. To further our analysis, we conducted non-parametric tests on our different treatments. Specifically, we started examining whether there are significant differences between the group size and random payoff treatments. We first run the Kruskall-Wallis equality of population rank test for the whole sample and to compare the choice of B across the number of subjects and the different levels of r. The test on the number of subjects indicates there are statistically significant differences between the population according the group size (Chi square (1) = 14.545 and *p-value*=0.0001), while for the levels of r the difference is statistically significant at the 1% level, but the test suggests a tie between, at least, two groups (Chi square (3) with ties = 14.451 and *p-value*=0.0024). Then we resort to the Mann-Whitney test and focus on comparisons across the r treatments. We consider for the tests the average frequency of "free riding" (choosing *action* B) per group in each group as independent variables. When there are four players involved in a group, the Mann-Whitney test cannot reject the null hypothesis that there is no difference in the frequency of free riding between the control treatment and the treatments where payoffs to B are randomized (*p-value*=0.132, [4, 1] *vs* [4, 0.95]; *p-value*=0.774, [4, 1] *vs* [4, 0.75]; *p-value*=0;773, [4, 1] *vs* [4, 0.55]).[10] Again a large majority of subjects play A when there are 4 players in a group, which is consistent with *Prediction 1.i)*: when they are involved in small size groups, subjects do not change their actions significantly whatever the level of r. Now, if we move to the treatments with groups of eight players, behavioral patterns differ consistently. Without surprise, when there is only a small probability of getting a null payoff when choosing the free-riding action, behaviors do not change significantly (*p-value*=0.771, [8, 1] *vs* [8, 0.95]). Overall, pooling the groups, we observe in these two treatments a majority of free-riding behaviors: the average frequency of choosing B is 54.8% in treatment [8, 1] and 51.8% in treatment [8, 0.95]. Though not all the players free-ride, this result gives evidence supporting *Prediction 1.ii)*. On the other hand, the null hypothesis is strongly rejected when we conduct other pairwise Mann-Whitney tests (*p-value*=0.014, [8, 1] *vs* [8, 0.75]; *p-value*=0.009, [8, 1] *vs* [8, 0.55]; *p-value*=0.013, [8, 0.95] *vs* [8, 0.75]; *p-value*=0.009, [8, 0.95] *vs* [8, 0.75]). The average frequency of choosing B is 36.25% in treatment [8, 0.75] and 27.08% in treatment [8, 0.55]. These results confirm that for higher levels of r the incentive for subjects to free-ride decreases significantly. The null hypothesis remains rejected for the latter treatments (*p-value*=0.09, [8, 0.75] *vs* [8, 0.55]), yet more weakly. These results suggest with the highest probability level (0.55) a large majority of subjects tend to play action A, yet not all of them. This provides some evidence to *Prediction 2* such as when a coordination problem on the equilibrium choice

[10]The null hypothesis cannot be rejected either when we run the test between the other pairwise treatments.

holds, the players' incentive to free-ride is not the same as when this one is absent.

To end this analysis, we turn to the results for *Prediction 3*. In treatment [8, 0.75], the Mann-Whitney test rejects the null hypothesis that there is no difference in the frequency of playing B when the fixed revenue to action A is p=2 and when it is p=4 (p-*value*=0.03). Pooling the groups, the average frequency of choosing B is 36.25% when p=2, whereas it is 26.4% when p=4. This result provides evidence to this prediction, which expected that an increase in p would reduce the propensity to free-ride. We also conducted experimental sessions for the treatment [8, 0.55]. Again, the Mann-Whitney test rejects the null hypothesis that there is no difference in the frequency of playing B when the fixed revenue to action A is p=2 and when it is p=4, yet more weakly (p-*value*=0.07). Pooling the groups, the average frequency of choosing B is 27.08% when p=2, whereas it is 17.9% when p=4. This latter result suggests that it may be useful to subsidy actions to contribute to the collective label, even when monitoring is quite strongly enforced.

4.2 Econometric results

This section presents our econometric analysis results. Making use of the panel structure of our experimental data, we use standard panel data techniques to test our hypothesis and to take into account the concerns over independence of observations and session dynamic effects when using experimental data (Houser, 2008; Frechette, 2012). Our dependent variable is dichotomous, as in each period the subjects of our experiment were asked between a decision to "cooperate" or free ride, therefore we resort to limited dependent variable approaches to analyse our data. Our problem and approach is similar to Rojas (2012), as he investigates how the propensity to collude is affected by different levels of information and monitoring. Thus, following his strategy we start exploring our data with a probit model and then resort to panel data methods. Specifically for the later we use Random Effects Probit that we specify as:

$$Pr(y_{it} = 1 \mid x_{it}) = \Phi(x_{it}\beta + \upsilon_i) \tag{4}$$

Where y_{it} is our dependent variable, coded 1 when the subject selects the choice B. In our setting, this means that the subject decided to free ride or to fraud. So our dependent variable can be interpreted as the propensity to free ride or fraud. x_{it} is a vector of independent variables and υ_i is our error term that we assume to be i.i.d. and $N(0, \sigma_\upsilon^2)$. Φ is the standard normal cumulative distribution. Variables in x_{it} include dummies accounting for the different treatments, a subject specific risk measurement variable[11] and variables accounting for time. The variable N accounts for the group size effects and is code as 1 when there were 8 subjects in the group. The monitoring treatments dummies were named $R1$, $R95$, $R75$ and $R55$, respectively denoting zero, 5, 25 and 45% probabilities of loosing payoff when the subjects chose to free-ride. We also created time dummies for each of the periods of the game.

As described in the previous section 372 subjects participated in our experiment and were placed in groups of 4 or 8 subjects. Recall these subjects were randomly assigned to 3 different treatments: market size effect, monitoring effect and fixed revenue effect. In each session the subjects made 20 choices, thus we end up with a panel of 7440 observations. For ease of exposition we start presenting a standard probit model to help us get a feel for the data. Also we first show the results for the pooled data as it allows us to investigate the group size effects and then we focus on a sub-sample containing the observations where we have 8 subjects in each group.

Table 1 reports the results of 4 models. In the table N, is a dummy variable accounting for our group size (1 if there were 8 subjects in the group), while R accounts for the monitoring levels. To avoid the dummy variable trap and consistently with our experimental protocol, we did not include the dummy for r=1 (that is $R1$) in our estimation, which is then the benchmark variable.

First, consider Models 1 and 2 which are Probit models. These models don't take into account the structure of our data but give us an initial insight into the data. Model 1, evaluates the impact of the treatment variables on the propensity to free ride. As it can be seen an increase in the number of subjects in the group statistically significantly increases the propensity, while more stringent monitoring and enforcement statistically significantly reduces the propensity to free ride. Model 2 adds the effect a variable measuring the attitude to risk as explained above. The results for the treatment variable are identical to those in model 1 and while the risk variable is statistically significant and has the expect sign, its magnitude is rather low. Models 3 and 4 use a random effects probit estimation. We observe that the value of ρ is significant across the models which supports for our random effects specification. Model 3 includes both the effects of N and R. The treatment effects are, with exception of R95, all statistically significant. So, as number of players increases to 8 so does the propensity to free ride, while the lower the level of R (i.e., the

[11] The risk variable is a quantitative variable that corresponds to the number of safe choices made by the subjects in the Holt and Laury (2002)'s multiple prize list test.

higher the randomness of payoffs to choosing B), the lower is propensity to free ride. This result is line with what we saw on for models 1 and 2 and supports *Prediction 1).* Model 4 includes the attitude to risk which has the expected sign but, consistently with findings in experiments conducted in strategic settings (Eckel and Wilson, 2004; Houser *et al.*, 2010) as a very small magnitude and is only significant at the 1% level of significance [12].

Turning to the case of the groups of 8, Table 2 reports the results of 3 models. The first shows the effect of the monitoring treatments on the propensity to free ride, the second takes also into account the risk attitude, finally the third adds the effect of the subsidy. In line with the results for the whole sample the treatment variables are strongly significant (except for $R95$, which as we have seen in the descriptive section above is not statistically significantly different from $R1$). What the results suggest is that an increase in the level of monitoring statistically significantly lessens the propensity to free ride. This offers support for point *ii*. of *Prediction 1.* Turning to model 4, we can see that while there is considerable more uncertainty in this treatment, the risk variable is only statistically significant at the 10% level and although the sign is as anticipate, the magnitude of the coefficient suggests very limited impact on the propensity to free ride. Finally, the last model shows that adding a subsidy statistically significantly reduces the propensity to free ride. This evidence supports our *Prediction 3* which suggest that an increase in the fixed revenue facilitates the coordination and therefore reduces the propensity to free-ride.

In short our analysis supports the first prediction from our theoretical model, in that an increase in the number of players in the market significantly increases the level of free ride. Also, in the case of groups of 8, an increase in the level of monitoring significantly decreases the propensity to free ride, which supports at least partly the second prediction. Finally, we show that the introduction of a subsidy that increases the fixed revenue reduces the propensity to free-ride.

[12] Consistently with other studies that used this same test, our subjects as categorized by the Holt and Laury's risk elicitation task are overwhelmingly risk averse (73.7%). Only 11% of the subjects are risk neutral and 5% are risk prone. The other subjects could not be classified because of inconsistent behaviors. The absence of correlation between risk attitudes as categorized by this test and the decision to free-ride does not mean necessarily that risky decisions are unimportant in our game. Actually, though the Holt and Laury's test remains widely used in experimental economics, it is more and more discussed in its ability to elicit risk attitudes relevantly in strategic settings.

5 Final Remarks

Food labels signaling the presence of quality attributes or production processes are widespread in the food industry. The implications and impact of labels have great interest from both theoretical and policy perspectives (Bonroy and Constantatos, 2015). The labels communicate attributes that are typically credence that is those that the consumer cannot verify. Moreover often these labels have a collective reputation as they are shared by a group of producers. The limited verifiability and the collective reputation associated to these labels leads to opportunities for fraudulent behavior by producers, who may have an incentive to free ride on the collective reputation. In recent years the press and USDA uncovered some cases of fraudulent behavior in organic food labels (Charles, 2014; USDA-AMS, 2014). Whereas there is a vast theoretical and empirical literature investigating the demand for labels with credence attributes, surprisingly, few papers have investigated the economics of labelling fraud. An exception is Hamilton and Zilberman (2006) who showed that the higher the number of producers sharing a label with collective reputation the higher is the incentive to fraud. Here, we aim to investigate what incentive structures mitigate the propensity to free ride on markets with collective reputation labels. Specifically, our goal is to assess how market size and alternative monitoring and enforcement strategies assuage free riding in a multi-player game setting. We also investigate how a subsidy policy may deter free riding behavior.

Drawing on Hamilton and Zilberman (2006) we develop a multi-player collective action game that analyses the players' incentive to contribute to a collective action and or free ride. A key feature of our approach is that the premium to the collective reputation is inversely related to the number of free riders. Moreover, the payoff to free ride is randomized under alternative monitoring and enforcement schemes. These features provides with a very rich and encompassing frameworks, as depending on the number of players and the uncertainty of payoffs to free ride, the nature of our game shifts from a public good to an assurance game and to a prisoners' dilemma game. The games have different equilibria and therefore allow us to explore alternative policies to obtain a social optimum outcome. We derive the following testable predictions: 1) when the number of players is small we have a pure public good game where the optimal outcome is to offer the collective reputation good; 2) when there is a large number of players and no monitoring and enforcement, the game becomes a prisoners dilemma where the optimal strategy is to free ride; 3) with a large number of players and random payoffs to free ride, we have a multiplayer assurance game, where there are multiple equilibria. The higher the probability of loosing payoff to free ride (in other words the more stringent is the monitoring and enforcement scheme) the closer is the equilibrium to the social optimum of no free ride.

To test our hypothesis we resort to experimental methods. We develop a novel experimental protocol, which is based on a public good game when payoffs are certain and turns into a multi-player assurance game when there are randomized payoffs to fraud. Subjects were put in groups of 4 or 8 players played for 20 periods, in which they had to choose between cooperating or to fraud. All the players in the group had to make their choice simultaneously. The groups of 4 and 8 players were further divided into four treatments corresponding to different levels of monitoring and uncertainty of the payoff to their actions. We conducted 24 experimental sessions in Grenoble, France between February and May 2014 and then in March-April 2015. We had a total of 372 subjects, who were undergraduate students from different programs in three universities. The students did not have background in game theory. We chose a fixed partner matching, so each group can be considered as a statistically independent observation. The experiment was conducted to get at least 4 observations per treatment.

To analyse our data we used descriptive statistics, non-parametric test and, making use of the panel structure of our data, we estimated random effects probit models to test our predictions. The results of the non-parametric tests examined whether there are significant differences in the propensity to free ride across group size and monitoring treatments. Providing initial evidence for the validity of prediction 1, we find a statistically significant difference in the propensity to free ride between the group with 4 and 8 players. This result is confirmed by our econometric analysis where we find that an increase on the number of players statistically significantly increases in the propensity to fraud. Then, as we focus on the treatments with 8 players, the results from non-parametric pairwise tests show a statistically significant difference between higher and lower levels of monitoring. Again, the econometric analysis strengths this evidence as it shows that an increase in uncertainty of payoffs to free ride statistically significantly reduces this action. Thus, our results also provide evidence for our second prediction. Finally, we investigate the effect of a subsidy as an alternative to more stringent (and costly) monitoring. We also find that a subsidy statistically significantly decreases the propensity to fraud, consistently with our third prediction.

While we acknowledge that our methodology limits our ability to provide clear and definitive policy recommendation, we still think our research provide important insights and lessons. First our framework suggests a trade-off between policies promoting market competition markets and those assuring consumer protection. This is because reducing barriers to entry, increases the number of players in the market which hampers the ability to coordinate actions, creating opportunities for free riding. In other words, when consumers are willing to pay for credence attributes delivered through a collective reputation label, it may be more economical to let producers of such goods restrict access to market. Secondly, and perhaps more relevant, an important instrument to restrict free riding on products carrying collective reputation labels, is to have an institution that regularly communicates the level of free ride detected in the market place. That is a mechanism increasing the level of transparency in the market place will provide a disincentive for firms to free ride. This is because a rational consumer won't be prepared to pay such a high premium when they there is a good chance of being defrauded. Actually markets with collective reputation labels could easily create such transparency mechanism by mandating third party certifiers to publicly and regularly report the proportion of product that failed inspection and why. Third, there is a possible trade-off between subsidizing the prices of products marketed with collective reputation labels and monitoring. Our model show that the introduction of a subsidy increases the efficacy of lower levels of monitoring and enforcement in deterring fraud. Off course this may cause problems, as subsidies cause well know distortions in trade and our model does not consider the costs of monitoring nor the ones for the subsidy.

Our research can be extended in a number of ways. First, it may be possible to evaluate some of our proposition with real world data. For example, data from the USDA National Organic Program reporting fraud detection could be used to assess decrease of organic foods premiums after an fraud detection is announced. Our model could be extended to consider two separate issues related to monitoring an enforcement: first monitoring agencies may fail to detect the fraud or have a detection system with a high probability of incurring a type II error. This will lead to another level of uncertainty in the model, that may impact the nature and therefore equilibria of the game. Second, punishment for free riding may increase, for instance the player caught cheating may be temporarily or permanently excluded from the market. Finally, we could investigate alternative ways of communicating the level of fraud in the market and analyse how it might change the equilibria.

References

Baksi, S. and Bose, P. (2007). "Credence goods, efficient labelling policies, and regulatory enforcement". *Environmental and Resource Economics*, **37**(2):411–430.

Bonroy, O. and Constantatos, C. (2015). "On the economics of labels: How their introduction affects the functioning of markets and the welfare of all participants". *American Journal of Agricultural Economics*, **97**(1):239–259.

Camerer, C. (2003). *Behavioral Game Theory. Experiments in Strategic Interaction.* Princeton University Press.

Cason, T. N. and Gangadharan, L. (2002). "Environmental labeling and incomplete consumer information in laboratory markets". *Journal of Environmental Economics and Management*, **43**(1):113–134.

Charles, D. (2014). "Can you trust that organic label on imported food?" *NPR news online*.

Chaudhuri, A. (2011). "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature". *Experimental Economics*, **14**(1):47–83.

Dixit, A., Skeath, S., and Reiley, D. (2009). *Games of Strategy.* W.W. Norton.

Eckel, C. C. and Wilson, R. K. (2004). "Is trust a risky decision?" *Journal of Economic Behavior and Organization*, **55**(4):447–465.

Engel, C. and Zhurakhovska, L. (2014). "Conditional cooperation with negative externalities: An experiment". *Journal of Economic Behavior and Organization, forthcoming.*

Ertan, A., Page, T., and Putterman, L. (2009). "Who to punish? individual decisions and majority rule in mitigating the free rider problem". *European Economic Review*, **53**(5):495–511.

Fehr, E. and Gachter, S. (2000). "Cooperation and punishment in public goods experiments". *American Economic Review*, **90**(4):980–994.

Frechette, G. (2012). "Session-effects in the laboratory". *Experimental Economics*, **15**(3):485–498.

Hamilton, S. F. and Zilberman, D. (2006). "Green markets, eco-certification, and equilibrium fraud". *Journal of Environmental Economics and Management*, **52**(3):627–644.

Holt, C. A. and Laury, S. K. (2002). "Risk aversion and incentive effects". *American Economic Review*, **92**(5):1644–1655.

Houser, D., Schunk, D., and Winter, J. (2010). "Distinguishing trust from risk: An anatomy of the investment game". *Journal of Economic Behavior and Organization*, **74**(1-2):72–81.

Houser, D. E. (2008). "experiments and econometrics". In S. N. Durlauf and L. E. Blume, editors, "The New Palgrave Dictionary of Economics", Palgrave Macmillan, Basingstoke.

Kiesel, K. and Villas-Boas, S. B. (2013). "Can information costs affect consumer choice? nutritional labels in a supermarket experiment". *International Journal of Industrial Organization*, **31**(2):153–163.

Ledyard, J. O. (1997). *Public Goods: A Survey of Experimental Research*. Princeton University Press.

Michaud, C., Llerena, D., and Joly, I. (2013). "Willingness to pay for environmental attributes of non-food agricultural products: a real choice experiment". *European Review of Agricultural Economics*, **40**(2):313–329.

Rojas, C. (2012). "The role of demand information and monitoring in tacit collusion". *RAND Journal of Economics*, **43**(1):78–109.

Sefton, M., Shupp, R., and Walker, J. M. (2007). "The effect of rewards and sanctions in provision of public goods". *Economic Inquiry*, **45**(4):671–690.

Sousa, S. (2010). "Cooperation and punishment under uncertain enforcement". *The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham, Discussion Papers*, (2010-06).

USDA-AMS (2014). *Fraudulent Organic Certificates*. USDA Publishing.

Walker, J. M. and Halloran, M. A. (2004). "Rewards and sanctions and the provision of public goods in one-shot settings". *Experimental Economics*, **7**(3):235–247.

APPENDIX

Dans mon groupe, joueurs ayant choisi		Mon gain pour:	Mon gain pour:
κ	φ	κ	φ
4	0	1000	
3	1	750	950
2	2	500	700
1	3	250	450
0	4		200

Veuillez choisir entre κ et φ

Je choisis κ   Je choisis φ

Matrix 1: A

Dans mon groupe, joueurs ayant choisi		Mon gain pour κ:	Mon gain pour φ:	
κ	φ	avec 100 / 100	avec 55 / 100	avec 45 / 100
4	0	1000		
3	1	750	950	0
2	2	500	700	0
1	3	250	450	0
0	4		200	0

Veuillez choisir entre κ et φ

Je choisis κ   Je choisis φ

Matrix 2: A

Dans mon groupe, joueurs ayant choisi		Mon gain pour:	Mon gain pour:
κ	φ	κ	φ
8	0	1000	
7	1	875	1075
6	2	750	950
5	3	625	825
4	4	500	700
3	5	375	575
2	6	250	450
1	7	125	325
0	8		200

Veuillez choisir entre κ et φ

Je choisis κ

Je choisis φ

Matrix 3: A

Dans mon groupe, joueurs ayant choisi		Mon gain pour κ:	Mon gain pour φ:	
κ	φ	avec 100 / 100	avec 55 / 100	avec 45 / 100
4	0	1000		
3	1	750	950	0
2	2	500	700	0
1	3	250	450	0
0	4		200	0

Veuillez choisir entre κ et φ

Je choisis κ   Je choisis φ

Matrix 4: A

Dates	Treatment	Nb of Sessions	Nb of subjects	Nb of observations
Feb. 16, 2014	[4;1]	1	16	4
Feb. 18, 2014	[8;1]	2	32	4
Feb. 18, 2014	[4;0.55]	1	16	4
Feb. 19, 2014	[8;0.55]	1	16	2
Feb. 20, 2014	[8;0.55]	1	16	2
Feb. 24, 2014	[4;0.95]	1	12	3
Feb. 24, 2014	[8;0.95]	1	16	2
Feb. 26, 2014	[8;0.95]	1	16	2
Feb. 27, 2014	[4;0.75]	1	16	4
Feb. 27, 2014	[8;0.75]	2	32	4
May 19, 2014	[4;1]	1	12	3
May 19, 2014	[8;0.55]	1	16	2
May 20, 2014	[4;0.95]	1	12	3
May 21, 2014	[8;0.75]	1	16	2
March 23, 2015	[8;0.55;p=4]	1	16	2
March 26, 2015	[8;0.75;p=4]	3	48	6
March 27, 2015	[8;0.75;p=4]	1	16	2
March 31, 2015	[8;0.55;p=4]	1	16	2
April 1, 2015	[8;0.55;p=4]	1	16	2
April 2, 2015	[8;0.55;p=4]	1	16	2
	Total	24	372	57

Table 1: Experimental sessions and observations

	Model 1	Model 2	Model 3	Model 4
	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef (Std. Err.)	Coef (Std. Err.)
N	1.046***	1.044***	1.327***	1.320***
	(0.047)	(0.048)	(0.131)	(0.130)
R95	0.060	0.064	0.152	0.157
	(0.056)	(0.056)	(0.168)	(0.167)
R75	−0.526***	−0.528***	−0.606***	−0.608***
	(0.049)	(0.049)	(0.146)	(0.145)
R55	−0.791***	−0.800***	−0.938***	0.948***
	(0.052)	(0.053)	(0.154)	(0.153)
Risk		−0.002***		−0.002**
		(0.000)		(0.001)
cons	−1.034***	−1.043***	−1.402***	−1.411***
	(0.050)	(0.050)	(0.143)	(0.142)
Log likelihood	-4032.046	-4021.091	-3529.1632	-3526.483
ρ			0.399***	0.394***
obs.	7440	7440	7440	7440

Table 2: Probit and Random effects Probit estimation for the choice of B for the pooled data

	Model 1	Model 2	Model 3
	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err)
R95	-0.092 (0.208)	-0.083 (0.206)	-0.092 (0.201)
R75	-0.856^{***} (0.166)	-0.855^{***} (0.165)	-0.575^{***} (0.175)
R55	-1.124^{***} (0.172)	-1.128^{***} (0.171)	-0.910^{***} (0.175)
Risk		-0.002^{*} (0.001)	-0.002^{*} (0.001)
Subsidy			-0.448^{***} (0.114)
cons	0.131 (0.147)	0.116 (0.146)	0.116 (0.142)
Log likelihood	-3047.32	-3045.51	-3037.93
ρ	0.374^{***}	0.372^{***}	0.358^{***}
obs.	5760	5760	5760

Table 3: Random effects Probit estimation for the choice of B for groups of 8 players

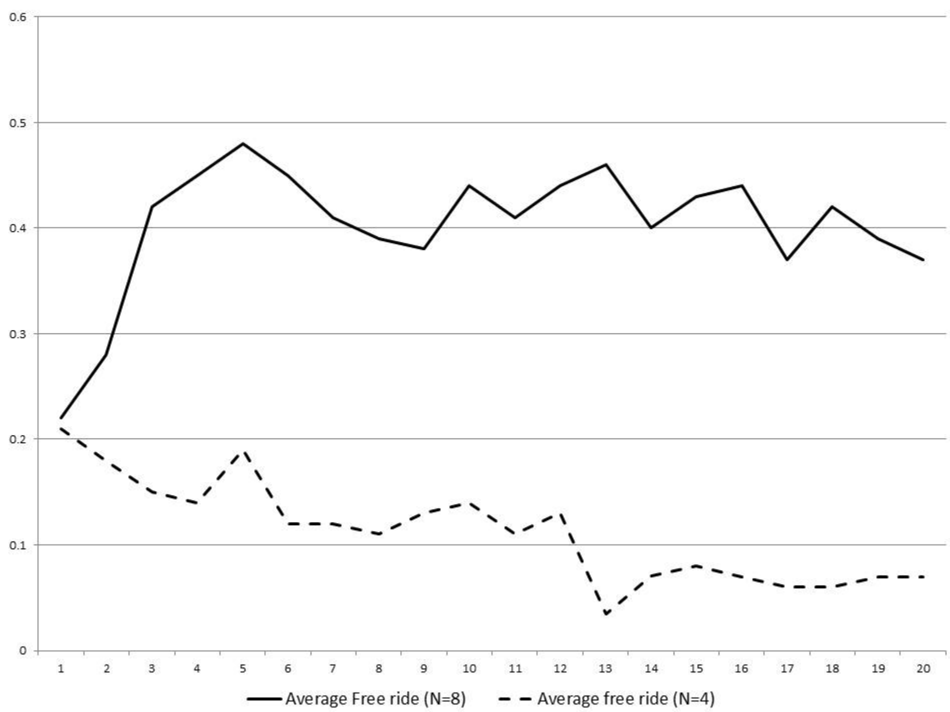

Figure 1: Average choice of B when N=4 and N=8

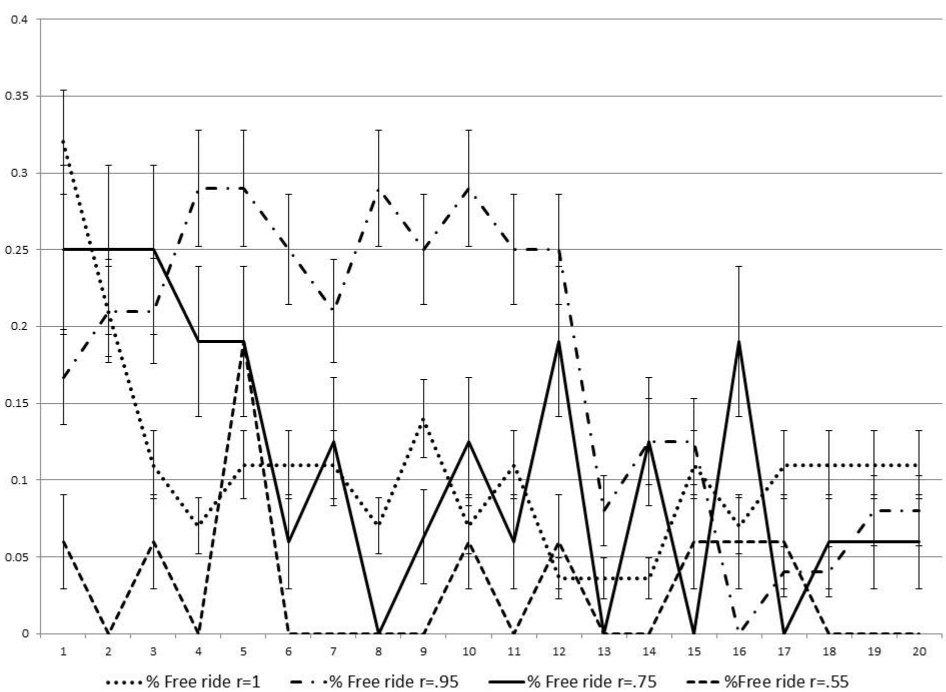

Figure 2: Average choice of B when N=4 according the treatments

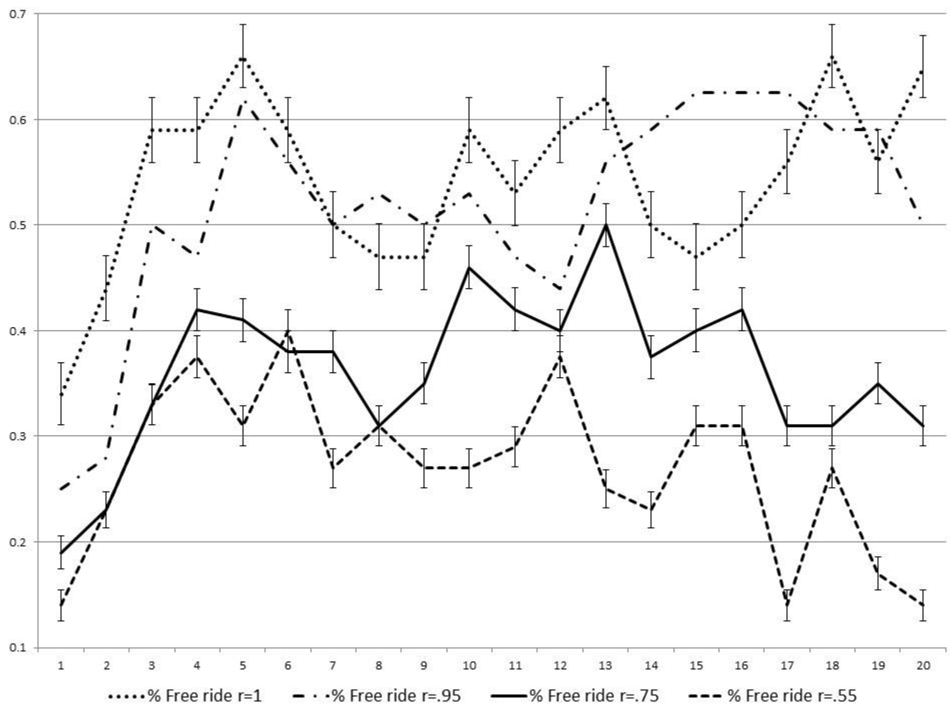

Figure 3: Average choice of B when N=8 according the treatments