Free riding and the provision of candy bars

Marco Haan, Peter Kooreman*

University of Groningen, Department of Economics, P.O. Box 800, 9700 AV Groningen, The Netherlands

Abstract

A wealth of experimental literature studies the effect of repetition and group size on the extent of free riding in the provision of public goods. In this paper, we use data from honor systems for candy bars in 166 firms to test whether such effects can be found outside the laboratory. We find that free riding increases with repetition, and weak evidence that free riding decreases with group size.   © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Free riding; Public good experiments; Honor system

1. Introduction

Standard economic theory assumes that people selfishly maximize their economic well-being, without taking the effect on the well-being of others into account. As a consequence, they are not inclined to contribute voluntarily to the provision of a public good. It is a dominant strategy not to contribute but rather to free ride on the contributions of others. Experiments, however, suggest that as a rule people do contribute to public goods (for a survey, see Ledyard, 1995). Typically, contributions amount to 40–60% of the Pareto efficient level. Contributions decline with repetition. The effect of the number of participants is ambiguous. This

*Corresponding author. Tel.: +31-50-363-7327; fax: +31-50-363-7337.

E-mail address: p.kooreman@eco.rug.nl (P. Kooreman).

evidence suggests that people tend to free ride to some extent, and that free riding increases when they face the same situation more often.

One disadvantage of such experimental evidence is that it is, by its very nature, artificial. The behavior of economic agents in experimental surroundings may give an indication of their behavior in real-world economic settings, but this is not necessarily the case. Unfortunately, it is hard to find data from real-world situations which allow us to systemically investigate the effect of repetition and the number of participants on the extent of free riding. In this paper, however, we study a data set which allows us to do just that. For several years, a substantial number of firms in the east of the Netherlands were supplied with candy bars, using an honor system. Employees who consumed a candy bar were expected to put one guilder in an (open) box. Whether or not they did so, was not actively monitored. Hence, employees could choose to pay one guilder, or, alternatively, to free ride on the contributions of their colleagues. Money was collected and stocks refilled biweekly. Using these data, we find evidence of free riding. Also, the extent of free riding increases in the number of repetitions, that is, the number of weeks the honor system has been present in a given firm. However, we do not find a significant effect of group size on the extent of free riding. These findings are consistent with those of the experimental literature.

The paper proceeds as follows. In the next section, we review the relevant literature. Section 3 describes the data. Empirical results are described in Section 4. Section 5 concludes.

2. The literature

In a typical public good experiment (see for example Isaac et al., 1984), each of a group of N subjects receive Z tokens. Each individual can choose to invest these in a group account or in an individual account. Tokens invested in an individual account yield 1 dollar to the investor, those invested in a group account yield G/N to every member of the group, with $1 < G < N$. Hence, it is Pareto optimal for all members of the group to invest all of their tokens in the group account. In a one-shot game, however, it is a dominant strategy to invest all tokens in the individual account.[1] The problem facing these subjects is equivalent to the one in

---

[1] Note that when the experiment is repeated and consists of a sequence of trials, the dominant strategy in each trial is not necessarily to contribute zero tokens. If the number of repetitions is unknown to the participants, equilibria can be sustained with a positive amount of contributions in each period, by invoking the folk theorem (see e.g. Fudenberg and Maskin, 1986). Yet, this theory cannot explain contributions that decrease with the number of repetitions. A theory that can explain this is provided by Kreps et al. (1982). They show that, in the context of a finitely repeated game, when subjects believe that there is a slight probability that the other subjects are not rational, the equilibrium typically involves contributions that are initially high, but decrease with the number of repetitions. Yet, this theory cannot explain positive contributions in the final period, which are typically observed. Also, positive contributions in one-shot games are consistent with neither theory.

which they have to decide whether or not to contribute voluntarily to a public good. Contributing yields a payoff for the group as a whole that is higher than the individual cost of contributing. Not contributing amounts to free riding. A subject that does not contribute enjoys the benefit of the public good provided by the other subjects, without bearing any of the costs.

The major findings in these types of experiments are the following (Ledyard, 1995, p. 121 or Dawes and Thaler, 1988). First, in one-shot trials and in the initial stages of repeated trials, subjects provide contributions roughly halfway between the Pareto-efficient level and the free riding level. Second, with repetition, contributions decline towards the Nash equilibrium of full free riding.[2] Third, face to face communication improves the rate of contribution. Fourth, the effect of group size is ambiguous.

Early studies often find a negative effect of group size on contributions. Chamberlin (1974) finds a negative effect on contributions as N increases. Bagnoli and McKee (1991) also find a negative effect, particularly in earlier periods. Marwell and Ames (1979) find no significant effect from varying the number of subjects. Isaac and Walker (1988) do not find a pure numbers effect, which is defined as an effect caused by an increase in N while holding constant G/N constant. Their results suggest that free riding decreases rather than increases when group size increases from 4 to 10. Isaac et al. (1994) find the same result when group size is increased to 40 or 100.

In all these experiments, the amount of public good provided is a continuous function of the total amount of contributions. Some experiments incorporate a threshold or provision point, where the public good is provided if and only if total contributions are above some threshold level.[3] Qualitatively, these threshold experiments yield the same outcomes as experiments without such a threshold. Marwell and Ames (1980) explicitly test for this, and do not find a significant difference between the two designs.

It is hard to find evidence from everyday transactions which allows the same kind of systematic study of the influence of different treatment variables on the extent of free riding. A small literature studies the relationship between congregational size and giving in churches. Sullivan (1985); Stonebraker (1993); and Zaleski and Zech (1994) all report a negative relationship between congregational size and per-member rates of annual giving.[4] This suggests free riding that is increasing in group size.[5] Yet these results are hard to interpret. Tullock (1996) argues that the goods churches supply are mainly of a private nature. Iannaccone

---

[2]Falkinger et al. (2000) show that, also when the environment is such that the Nash equilibrium predicts that all participants contribute a strictly positive amount, contributions still have a tendency to fall to the Nash equilibrium level.

[3]See e.g. Isaac et al. (1989); Dawes et al. (1986) and Marwell and Ames (1979, 1980) for slightly differing versions of such a framework.

[4]Reported in Iannaccone (1998).

[5]Lipford (1995) finds no evidence of an effect of membership on giving, but Zaleski and Zech (1996) argue that this is due to a misspecification of his model.

(1998) argues that congregation size may be endogenous. Church members may also feel that the services their church offers are of lower quality as the number of members increases. Hence, a negative relationship between giving and membership is not necessarily the result of more free riding.

Experimental evidence thus suggests that voluntary contributions to public goods decrease with repetition. Empirical and experimental evidence for a group size effect is mixed.

3. The data

For several years, the firm Candyman supplied a box containing candy bars, crisps, etc.[6] to several companies in the eastern part of the Netherlands. These were small firms whose scale did not justify a vending machine or full-service canteen. Sales were made through an open box containing roughly 100 candy bars. Where consumption was expected to be higher, more boxes were put in the same place. When taking something from the box, an employee was expected to place an amount of 1 guilder in a separate container. This container was open, to allow consumers to take their change.

As a rule, once every 2 weeks a Candyman representative came to collect the money and refill the box. By counting the number of items left, and the amount of money collected, Candyman could evaluate the average payment. When this average was too low, employees of the firm were notified by either publishing the average payment or the shortfall of money next to the box. In some cases, a written or oral statement was delivered, in which employees were warned that the box would be withdrawn if payment behavior would not improve. When the average payment was consistently too low, Candyman discontinued its service. In most cases, the installation of the box involved a Candyman representative explaining the procedure to the employees involved.

Note that the box has a public good character. When the average payment is lower than 1, Candyman will not withdraw the box immediately. It will only do so when the average payment drops below a certain threshold, unknown to the employees. This gives them an incentive to free ride, by not paying for their candy bar. When all the other employees are paying, using this strategy will not cause the public good to disappear. Therefore, the set-up is virtually identical to that of a provision point public good experiment.

Candyman was run by business students. Their objective was to gain experience in running a business, and to make some money to support their studies. The owners expected the employees in the firms they served to be sufficiently well-behaved to make their venture worthwhile. Yet, total earnings were modest.

---

[6]For simplicity, we will refer to this merchandise as candy bars.

The average cost of a candy bar for Candyman was around 55 cents. Given the overall average payment of 81 cents and the total amount of candy bars sold of 387,585 (see Table 1), this leaves the firm with a total gross profit, before allowing for transportation costs and taxes, of less than 100,000 guilders — for an operation that ran for two and a half years and involved several people.[7] To find new customers, Candyman tended to volunteer its services to firms in which they expected their scheme to be profitable. Usually, these were smaller firms without

Table 1
Sample statistics[a]

		Mean	Std. Dev.	Min.	Max.
Observation #		21.18	7.7	2	63
Week		23.90	15.3	1	51
Year		1993.7	0.783	1992	1995
Consumption	C_{it}	81.84	51.9	1	384
Payment	p_{it}	66.31	44.4	0	494.1
Avg. payment	b_{it}	0.831	0.219	0	6.125
Summer		0.086			
Christmas		0.056			
Temperature		12.85	6.11	0.4	28.7
Apeldoorn		0.462			
Bank		0.014			
Car		0.161			
Transport		0.081			
Printing		0.045			
Software		0.018			
$T \leq 10$		0.301			
$11 \leq T \leq 20$		0.224			
$21 \leq T \leq 30$		0.187			
$31 \leq T \leq 40$		0.140			
$41 \leq T \leq 50$		0.110			
$T > 50$		0.038			

Total # observations	4736	Total consumption	387,585
# of firms	166	Total payment	314,025
Obs. per firm	28.53	Overall avg. payment	0.810

  [a] Fractions in table refer to the fraction of observations. As fraction of the number of firms, we have: Bank 0.012; Car 0.157; Transport 0.072; Printing 0.042; Software 0.018; Apeldoorn 0.349. The variable T refers to the number of periods the box has been in a particular firm. Thus the observation for $T \leq 10$ implies that out of all observations, 30.1% concern observation that are among the first 10 for that particular period.

---

  [7] During the relevant period, the average exchange rate for the Dutch guilder was USD 0.56.

other vending facilities, such as canteens or vending machines. As a rule, boxes were put in a place where the employees had easy access to it, but other people, such as customers of that particular firm had not. Obviously, Candyman preferred the firms it serviced to be located close to each other, to save on transportation costs.

There is a total number of 166 firms for which sufficient data (i.e. more than one observation) are available. Every observation consists of the exact week in which the payment is collected, the total number of candy bars sold, and the total amount of money collected. Sample statistics are given in Table 1. The total number of observations is 4736. A total of 387,585 candy bars were sold throughout the sample period. For these sales, 314,025 guilders were collected, which implies an average payment of 81%.[8] This is much higher than what is usually found in public good experiments. The difference is probably due to the fact that, in such experiments, subjects are to a large extent free to choose their contribution. In the data considered here, there is a strong moral obligation to pay 1 guilder when taking a candy bar. Paying less amounts to stealing, whereas not contributing in a public good experiment is merely being selfish. Also, subjects within one firm are able to communicate. As noted, experimental evidence suggests that communication increases total contributions.

Observations in our data set differ widely with respect to the number of candy bars consumed. The average consumption per firm per two-week period equals 82. On average, 66 candy bars were paid for. In the table, 'Summer' is a dummy that equals 1 for observations in weeks 27 through 34, and zero otherwise. 'Christmas' is a dummy that equals 1 for weeks 51, 52, 1 and 2, and zero otherwise. This variable will be included in the empirical analysis to pick up a possible 'good intentions' effect of Christmas and New Year on payment behavior. 'Temperature' gives the average daily temperature in the east of the Netherlands, in degrees Celsius. The dummy 'Apeldoorn' equals one if the firm considered is located in Apeldoorn, which is by far the largest city in our sample.[9] We include this dummy to capture the possible effect of city size on payment behavior.

The 166 firms are active in very different sectors and industries. We were able to classify some clusters, yet in a lot of cases this was impossible, since many firms were not active in one unique sector, or we were not able to determine the sector(s) in which they were active. The clusters that could be identified are car dealerships and car maintenance (26 firms), transportation (12), printing (7), computer/software (3), and bank/insurance (2). From the table, it can be seen that 30.1% of our observations have $T \leq 10$, which means that the number of periods the box has been at that particular firm at the time of the observation is less that 10 two-week

---

[8] Note that this is lower than the average value of b_{it}, which equals 0.83. The latter value gives the unweighted average of all average payments, whereas in the main text we refer to the overall average payment.

[9] In 1993, the number of inhabitants in Apeldoorn was 149,504, which is at least a factor 6 larger than all other municipalities considered in this study.

periods. In 3.8% of the observations, the box has been in a firm for more than 50 periods.

The number of observations per firm varies from 2 to 63, with an average of 29. There are several reasons for this wide range of values. First, the time at which a particular firm entered the dataset differs. Initially (in week 39, 1992), there were 12 firms in the sample. As Candyman became more successful, it attracted more clients. In the final week for which we have data[10] (week 17, 1995), there were 109. Given the total of 166 firms in the data set, this implies that 57 firms had left the panel before the end of our sample period. In some cases the service was discontinued for several weeks, either in the hope that this would improve behavior, or for other reasons.[11]

There are two possible reasons for a firm to leave the sample. First, a firm may itself choose to no longer use Candyman's services. There were cases in which the scheme was so successful that employees decided to run it themselves. Also, a firm may relocate or go out of business. In one case, management refused to allow the service if payment would be any less than 100%. A second reason for a firm to leave the sample is that Candyman decides to discontinue its service because the average payment is too low. As a rough rule[12], it did so when the average payment in a given firm was consistently below some 65 cents. Of course, other factors also played a role, such as the distance to other firms, the firm's turnover, and past payment behavior. Using this rough threshold, we have divided the 57 'drop-outs' in two categories: 'voluntary' for the 28 firms that chose to leave the scheme and 'forced' for the 29 where payment fell short. The average payment in their final period was 0.88 for the voluntary drop-outs, and 0.41 for the forced drop-outs.[13]

Further information about the distribution of the 3 types of firms is given in Fig. 1. From this figure, it can be seen that out of all the observations with $T=1$, so out of all the cases in which the box has been in a firm for just 1 period, 66% of the observations concern firms that are still in the sample in the final week of our data, whereas 17% concern firms that are ultimately forced to drop out, and another 17% are firms that ultimately drop out voluntarily. While the fraction of surviving firms is stable over the first 10 periods, it increases steadily after that time. At $T=47$, we only have surviving firms left.[14]

---

[10]After this, the owners sold Candyman. The new owners did not collect data on a level detailed enough for our analyses.

[11]As an example, in one case there was a problem with mice.

[12]not known to its customers.

[13]For the penultimate period, these numbers are 0.82 and 0.55 respectively, for the period before that 0.78 and 0.58. Note therefore that, at the time of the drop-out, the average payment was increasing among voluntary drop-outs and decreasing among forced drop-outs.

[14]Further sample statistics of these 3 types are as follows. For survivors, the average number of observations is 34.66, for voluntary drop-outs it is 16.29, for forced drop-outs 17.38. Average per-period consumption is 96.7 among voluntary drop-outs, 92.1 among forced drop-outs, and 78.6 among survivors. Out of a total of 4736 observations, 470 are from forced, and 490 from voluntary drop-outs.
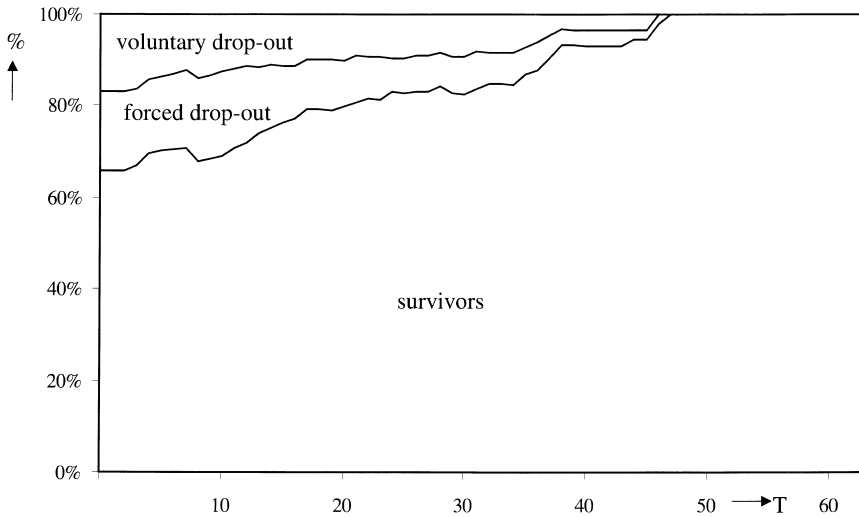
Fig. 1. Distribution of firms over time.

   Clearly, using this data set entails both advantages and disadvantages compared to the use of experimental data. A major advantage is that the subjects are not placed in an artificial environment in which they have to make decisions which they are not used to make in everyday life. Every time they take a candy bar, subjects simply decide how much to pay for it, without ever knowing that they are part of an economic experiment. Also, we do not suffer from a selection bias regarding the selection of subjects. In experiments, the subjects are typically undergraduate students in economics or some other social science, who volunteer to participate in such an experiment. Our data involves a much broader range of people. For these reasons, it seems that the results from this approach can be generalized more easily to other cases in which free riding is an issue. Also, our approach allows easy access to much more data. The typical experimental analysis consists of some 20–30 experiments, each consisting of 10 repetitions, and involving a total number of subjects somewhere between 100 and 250.[15] In our data set, we effectively have 166 experiments, each consisting of 2–63 repetitions, with an average of 29. The data describe the behavior of several thousand people. Of course, the disadvantage of using empirical rather than experimental data is that it is harder to control circumstances. Also, we can only observe aggregate data,

---

[15]The outlier is Isaac et al. (1994) who use 87 experiments involving 1908 subjects. They use extra credit points rather than money to motivate subjects.
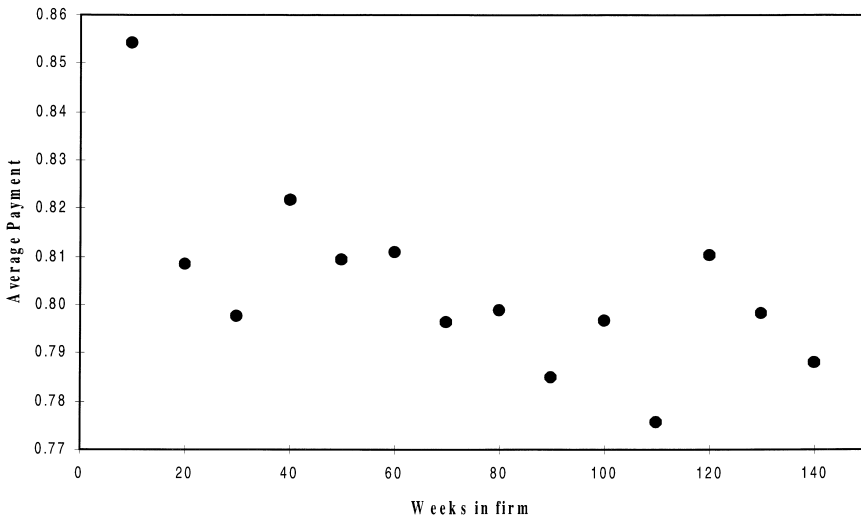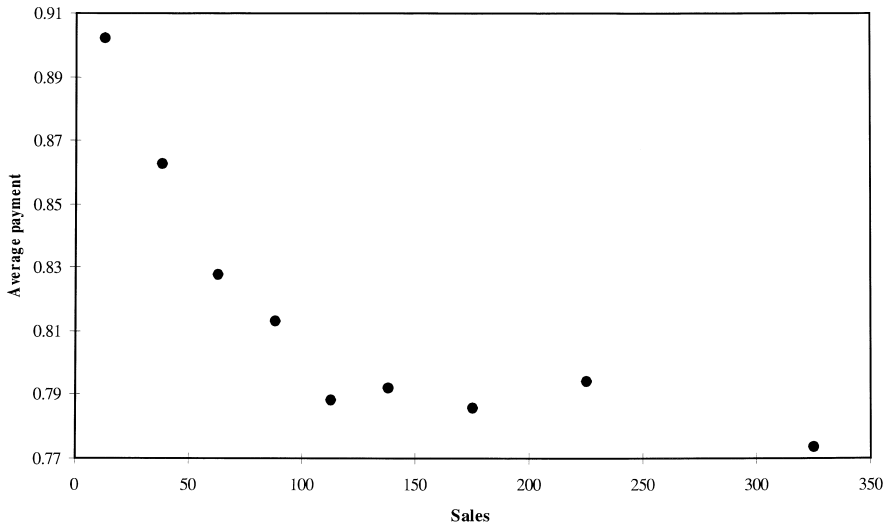
Note: Points in the figure give the average payment for all observations where the number of weeks the box has been in a firm is smaller than or equal to the value on the horizontal axis, but higher than the value associated with the previous point in the figure.

Fig. 2. Average payment over time.

and are not able to track the payment behavior of each individual over time, something that can be done with experimental data.

The average payment over all observations differs considerably. Figs. 2 and 3 give an impression of these differences. First, we have aggregated the data to calculate the average payment in the first 10 weeks that a box was in a certain firm, the average payment in weeks 11–20, etc. These aggregated data are plotted in Fig. 2. For example, the observation (10, 0.854) implies that of all the candy bars sold in any firm during the first 10 weeks the box was present in that particular firm, 85.4% was paid for. The figure suggests that the average payment is decreasing in the number of weeks the box was in a firm. Especially in the first 10 weeks, this effect is pronounced. Apparently, the average payment is high initially, then drops considerably, and slowly decreases for the rest of the period.

In Fig. 3, we perform a similar exercise with total sales in a two-week period on the horizontal axis. Here, we use sales as a proxy for the number of people within a certain firm. For the first observation, for example, we consider all observations in which fewer than 26 candy bars were sold. Over all these cases, an average amount of 90.2% were paid for. Since there are very few observations in which total sales exceed 200, observations in this area are aggregated over a broader range of values. Fig. 2 therefore gives the average payment over a certain range of values for sales, plotted against the midpoints of the relevant intervals. This graph suggests a strong negative correlation between the average payment and total

Note: Points in the figure give the average payment for all observations where total sales are in a certain interval of which the midpoint is given by the value on the horizontal axis.

Fig. 3. Average payment relative to sales.

sales. When we take the latter as an approximation of the number of employees within a particular firm, this suggests that the amount of free riding increases with the number of employees.

Figs. 2 and 3 thus suggest that the extent of free riding is increasing in both group size and the number of repetitions, that is, the number of weeks the honor system has been present in a firm. In the next section we give a formal econometric test for these relations.

4. Empirical results

One problem in the empirical implementation of the model is the appropriate definition of group size. Information on the number of employees in the firm is not available in the data set. Even if it were, we would still not have any information on the number of employees actually using the box, which seems the relevant variable to determine the effect of group size on the amount of free-riding. For this reason, we focus on the relationship between average payment and consumption. If consumption patterns are roughly similar, then total consumption should be a good indicator of the number of people using the box. Let the subscripts t and i refer to time and the firm, respectively. We focus on the following equation:

$$\log b_{it} = a_0 + a_1 \log C_{it} + a_2 T_{it} + bX_i + cZ_t + \epsilon_{it} \tag{1}$$

Here b_{it} is average payment, C_{it} the number of candy bars sold, and T_{it} the number of two-week periods the box has been in the firm. X_i is a vector of time-invariant firm specific variables: five dummies for firm type (car repair, transportation, printing, computer/software, and bank/insurance), and a dummy for the municipality of Apeldoorn. Z_t is the 'Christmas' dummy.

To control for endogeneity of C_{it} Eq. (1) is estimated by two-stage least squares. As instruments, we use a temperature variable (average maximum daily temperature in the eastern part of The Netherlands, where all the firms are located), a dummy for summer vacation weeks, and the other exogenous variables. While both the temperature and the vacation variable will affect the total number of candy bars consumed in a firm, they are unlikely to have any direct effect on payment behavior, conditional on the other variables included.[16] The first stage estimation results are reported in Table 2.

The number of periods the box has been in the firm has a small but significant positive effect on total consumption: every 3 months total consumption increases by about one percent. During the summer weeks consumption is about 9 percent lower than usual, whereas in the first two and the last 2 weeks of the year consumption is about 7 percent lower. Both effects are probably related to a higher absence of employees. The outside temperature has a strong negative effect on consumption. On a day when the temperature is 30°C (86°F) consumption is twenty percent lower than on a day with a temperature of 0°C (32°F), other things equal. This may be due to substitution towards drinks on warmer days (the boxes do not contain beverages). Municipality and firm type dummies are all insignificant.

Column 2 in Table 2 reports the results for the average payment equation. The number of periods the box has been in the firm has a significant and sizeable negative effect on average payment. The dummy for firms in the printing industry is significant at the ten percent level and large: In these firms average payment is twenty percentage points higher than in other firms. All the other explanatory variables are not significant. In particular, we do not find an effect of total consumption on average payment, contrary to what Fig. 2 suggests. We have also added the square of log consumption to our estimation equation. The result is reported in the right-hand column of Table 2. Both consumption variables are now insignificant. Inclusion of this variable has virtually no effect on the estimated impact of the number of periods.

To analyze the effect of the number of periods in more detail, we also estimated an alternative specification in which the time effect is specified as $f(t) = a_1.\exp(-a_2.t)$, with b_{it} rather than log (b_{it}) as the left hand side variable. Note that $f(0) - f(\infty) = a_1$ (if $a_2 > 0$), so that the parameter a_1 measures the difference between the initial and the long run level of average payment; a_2 measures the

---

[16]All equations were estimated using a random effects estimator. Hausman specification tests did not reject the random effects specification against a fixed effects specification.

Table 2
Estimation results[a] (*t*-values in parentheses)

Experimental variable	Dependent variable		
	ln (consumption)	ln (avg. payment)	ln (avg. payment)
# of periods	0.00158	−0.00243	−0.00241
	(4.3)	(−3.8)	(−3.8)
ln (consumption)	–	0.151	1.192
		(1.2)	(0.3)
$[\text{ln (consumption)}]^2$	–	–	−0.124
			(−0.3)
Firm Type			
car	−0.156	0.046	0.501
	(−1.3)	(0.9)	(0.8)
transport	0.119	0.072	0.767
	(0.7)	(1.0)	(1.0)
printing	−0.319	0.206	0.213
	(−1.4)	(1.9)	(1.9)
computers	0.550	−0.087	0.044
	(1.7)	(−0.6)	(−0.2)
bank	−0.389	0.230	0.243
	(−0.9)	(1.1)	(1.2)
Apeldoorn	−0.039	0.056	0.056
	(−0.3)	(1.4)	(1.4)
Christmas	−0.072	0.011	0.011
	(−3.3)	(0.3)	(0.3)
Summer	−0.0916	–	–
	(−4.7)		
Temperature	−0.00778	–	–
	(−8.4)		
Constant	4.341	−0.891	−3.069
	(69.2)	(−1.6)	(−0.4)

[a] Estimated with two-stage least squares, using a random effects estimator. 'Apeldoorn' is a dummy for firms located in Apeldoorn, 'Christmas' a dummy for observations in weeks 51, 52, 1, and 2, 'Summer' a dummy for observations in weeks 27 through 34. Estimates are for a panel of 166 firms, yielding a total of 4736 observations. Consumption refers to the total consumption in a particular firm during a two-week period, avg. payment to the average payment for a candy bar in that firm during that period, # of periods the number of such periods the box has been in firm.

speed of adjustment towards the long run level. Using non-linear estimation methods we find $\hat{a}_1 = 0.0836$ (*t*-value = 6.4) and $\hat{a}_2 = 0.10522$ (*t*-value = 2.8). Thus, the long run average payment is eight percentage points lower than the initial average payment. It takes 13 weeks to reach half of the adjustment; after one year 94 percent of the adjustment has taken place.

To obtain further insight into the possibly non-linear effect of the number of periods, we re-estimated Eq. (1) with dummies for different time intervals, rather than using *T* as a continuous variable. The results are reported in Table 3, where $T \leq 10$ serves as the reference category. Interestingly, the average payment first

Table 3
Estimation results[a] (*t*-values in parentheses)

Expl. variable	All firms ln(avg. payment)	Survivors ln(avg. payment)
$11 \leq T \leq 20$	−0.042	−0.041
	(−1.9)	(−1.6)
$21 \leq T \leq 30$	−0.114	−0.074
	(−4.5)	(−2.6)
$31 \leq T \leq 40$	−0.097	−0.085
	(−3.6)	(−3.0)
$41 \leq T \leq 50$	−0.092	−0.097
	(−2.9)	(−2.9)
$T > 50$	−0.086	−0.095
	(−1.8)	(−2.0)
ln (consumption)	0.178	0.301
	(1.3)	(1.9)
Firm Type		
car	0.052	0.082
	(0.9)	(1.8)
transport	0.066	0.012
	(0.9)	(0.2)
printing	0.214	0.169
	(2.0)	(2.5)
computers	−0.109	−0.224
	(−0.7)	(−1.4)
bank	0.237	0.206
	(1.1)	(1.7)
Apeldoorn	0.055	0.029
	(1.3)	(0.9)
Christmas	0.020	0.018
	(0.6)	(0.5)
Constant	−1.005	−1.439
	(−1.7)	(−2.2)

[a] Apeldoorn is a dummy for firms located in Apeldoorn, Christmas a dummy for observations in weeks 51, 52, 1, and 2, Summer a dummy for observations in weeks 27 through 34. Estimates are for a panel of 166 firms, a total of 4736 observations. Consumption refers to the total consumption in a firm during a two-week period, avg. payment to the average payment for a candy bar in that firm during that period, The variable T is the number of periods the box has been in a firm. Survivors are firms that are still serviced at the end of sample period.

drops 11.5% below that in the category $T \leq 10$, and then recovers to the long-run rate of −8.5%. Yet, an *F*-test shows that the coefficients of the categories $21 < = T < = 30$ through $T > 50$ do not differ significantly.[17] Table 3 also reports the result of the same regression on the subset of firms that do not drop out of the sample. The effect of a recovery after $T > 20$ now disappears, suggesting that it is

---

[17]The coefficient on $11 \leq T \leq 20$, however, *is* significantly different from the coefficients on $21 \leq T \leq 30$ through $T > 50$.

merely due to a selection effect. Interestingly, in the survivor sample, the effect of consumption on payment behavior is weakly significant (at a 10% level)[18] and positive. This suggests that among the surviving firms, an increase in group size actually decreases the extent of free-riding. One possible explanation is that, as consumption increases, more people are expected to be found at or near the box at any point in time. This may make consumers more reluctant to cheat by not paying for their candy bar.

5. Conclusion

In this paper, we showed that some stylized facts from the experimental literature on the voluntary provision of public goods carry over to a more real-life environment. We used data on the payment for candy bars in honor systems in 166 firms over, on average, 29 two-week periods. We find that the average payment decreases through time. This decrease is concentrated in the first 30 periods (60 weeks). After that, the average payment stabilizes. We find some weak evidence that an increase of group size decreases the extent of free-riding. This may be partly due to our specific set-up. As consumption increases, payment behavior is more likely to be implicitly monitored by other consumers.

Acknowledgements

References

Bagnoli, M., McKee, M., 1991. Voluntary contribution games: efficient private provision of public goods. Economic Inquiry 29 (2), 351–366.

Chamberlin, J., 1974. Provision of public goods as a function of group size. American Political Science Review 68 (2), 707–716.

Dawes, R.M., Orbell, J.M., Simmons, R.T., Van de Kragt, A.J., 1986. Organizing groups for collective action. American Political Science Review 80 (4), 1171–1185.

Dawes, R.M., Thaler, R.H., 1988. Cooperation. Journal of Economic Perspectives 2 (3), 187–197.

---

[18]The corresponding P-level equals $P = 0.061$.

Falkinger, J., Fehr, E., Gächter, S., Winter-Ebmer, R., 2000. A simple mechanism for the efficient provision of public goods: experimental evidence. American Economic Review 90 (1), 247–264.

Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting and with incomplete information. Econometrica 54 (3), 533–554.

Iannaccone, L.R., 1998. Introduction to the economics of religion. Journal of Economic Literature 36 (3), 1465–1496.

Isaac, R.M., Schmidtz, D., Walker, J.M., 1989. The assurance problem in a laboratory market. Public Choice 62 (3), 217–236.

Isaac, R.M., Walker, J.M., 1988. Group size effects in public goods provision: the voluntary contributions mechanism. Quarterly Journal of Economics 103 (1), 179–199.

Isaac, R.M., Walker, J.M., Thomas, S.H., 1984. Divergent evidence on free riding: An experimental examination of possible explanations. Public Choice 43 (2), 113–149.

Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group size and the voluntary provision of provision of public goods — Experimental evidence utilizing large groups. Journal of Public Economics 54 (1), 1–36.

Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory 27 (1), 245–252.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, New Jersey, pp. 111–195.

Lipford, J.W., 1995. Group size and the free-rider hypothesis: An examination of new evidence from churches. Public Choice 83 (3), 291–303.

Marwell, G., Ames, R.E., 1979. Experiments on the provision of public goods. I. Resources, interest, group size and the free-rider problem. American Journal of Sociology 84 (6), 1335–1360.

Marwell, G., Ames, R.E., 1980. Experiments on the provision of public goods. II. Provision points, stakes, experience and the free-rider problem. American Journal of Sociology 85 (4), 926–937.

Stonebraker, R., 1993. Optimal church size: the bigger the better? Journal for the Scientific Study of Religion 32 (3), 231–241.

Sullivan, D.H., 1985. Simultaneous determination of church contributions and church attendance. Economic Inquiry 23 (2), 309–320.

Tullock, G., 1996. On 'Group size and the free-rider hypothesis: Another hypothesis.' by Jody W. Lipford: Comment. Public Choice 87 (1), 185.

Zaleski, P.A., Zech, C.E., 1994. Economic and attitudinal factors in Catholic and Protestant religious giving. Review of Religious Research 36 (2), 58–167.

Zaleski, P.A., Zech, C.E., 1996. Group size and the free-rider hypothesis: A re-examination of old evidence from churches: comment. Public Choice 88 (3), 407–416.