

Incentives for Subjects in Internet Experiments*

Peter Duersch[§] Jörg Oechssler^{§,†} Burkhard C. Schipper[‡]

April 27, 2009

Abstract

Internet experiments are a new and convenient way for reaching a large subject pool. Yet, providing incentives to subjects can be a tricky design issue. One cost-effective and simple method is the publication of a *high score* (as in computer games). We test whether a high score provides adequate and non-distortionary incentives by comparing it to the usual performance based incentives. We find significant differences and conclude that high scores are not always appropriate as an incentive device. Performance based financial incentives seem to be required also in internet experiments.

JEL codes: C72; C91; C92; D43; L13.

Keywords: incentives; experiments; internet; high score.

*We thank one referee for very constructive comments. [§]Department of Economics, University of Heidelberg, email: oechssler@uni-hd.de; [‡]Department of Economics, University of California, Davis.

[†]Corresponding author.

1 Introduction

Experiments on the internet have become more and more frequent in recent years.¹ Using the internet has some obvious advantages. It is a relatively inexpensive way of reaching many subjects. It allows to broaden the subject pool beyond the standard undergraduate student population. In addition, decision making at one's own PC at home or in the office is arguably a more natural setting than that in the laboratory. It resembles an environment that is familiar to many people, e.g. from home banking.

However, experimenting on the internet also poses a number of new challenges. In this note, we will focus on one such problem, namely the way incentives are provided to subjects. In experimental economics, standard procedures require that subjects receive adequate incentives. Although usually those incentives are provided through cash payments, other incentives are being used (e.g. grades, lottery tickets for non-cash prizes, etc.). For obvious reasons, the distribution of cash is difficult and/or expensive in large-scale internet experiments. Furthermore, often an experimenter wants to collect data anonymously, which can nicely be done via the internet. Yet, to preserve anonymity one does not want pay subjects directly because such payment requires knowing the subject's identity (e.g. name plus postal address or the bank account number). It would thus be desirable if other incentives schemes could be found that would provide adequate incentives without distorting results.

We consider here one such incentive scheme that is well known from computer games (or for the older of us, from pinball machines), the high score. The high score is simply a list of top performers (i.e. usually their initials or nicknames) with their associated score or payoff. The idea is that subjects are inherently motivated to achieve a top placement on this list. If this were true, the high score would be a very cost-effective and simple way

¹For experiments that have been conducted over the internet, see e.g., Forsythe et al. (1992, 1999), Lucking-Reiley (1999), Shavit et al. (2001), Bosch-Domenech et al. (2002), Güth et al. (2003), Bossaerts and Plott (2004) and Drehmann et al. (2005, 2007). Greiner et al. (2002), Anderhub et al. (2001), and Charness et al. (2007) contain detailed discussions of methodological issues.

of providing incentives to subjects in internet experiments.

We test the hypothesis that a high score provides adequate and non-distortionary incentives by comparing subjects' behavior in an internet experiment with behavior of subjects in a laboratory where we provide various forms of incentives. Incentives in the lab range from the usual performance pay over fixed payments to no payment at all. All incentive schemes in the lab are supplemented through a high score.

Our main findings are that results obtained with only a high score differ significantly from those obtained with the usual cash incentives. This holds independently of whether the experiment is run on the internet or in the laboratory.

2 Experimental design

More than 700 subjects participated in a simple experiment with the structure of a Cournot duopoly. Table 1 provides a summary of the four experimental treatments.² The bulk of the experiment was conducted as an internet experiment (setting *net*). In *net*, subjects played on the internet, in a location of their own choice (home, office etc.), and at their own pace. The only incentives for subjects was the chance to be listed on a *high score* table, which publicly displayed on our webpage the score and the chosen nickname of each player.³

Additionally, there were three different laboratory experiments with various forms of monetary incentives, which were designed to bridge the gap between *net* and a standard laboratory experiment. Closest to *net* was setting *lab-np*, in which subjects played in the laboratory but without any monetary incentives.⁴ As in *net*, the only incentive was the high score table. Thus, the only difference between the two treatments was the environment, that is, laboratory versus subjects' homes or offices. Next in line was setting

²See Duersch et al. (2009) for more details about the experimental design, instructions, and screen shots.

³For at least one subject the incentive was so great that he or she invested sufficient time to (unsuccessfully) hack our system, and tried to manipulate the high score table.

⁴Subjects in *lab-np* were first year students taking part in a campus tour.

lab-f, in which additional to the high score table, subjects received a *fixed* payment of 10 euros as soon as they entered the lab.⁵ Finally, setting *lab* was the usual laboratory experiment in which subjects were paid according to the sum of their profits. A high score table was displayed in addition. The instructions for all settings were the same up to the incentive structure. At the end of the experiment, we asked subjects to fill in a questionnaire with some demographic data.⁶

Table 1: Summary of experimental treatments

treatment	setting	incentives	# of subjects
<i>net</i>	internet	high score	550
<i>lab-np</i>	laboratory	high score	55
<i>lab-f</i>	laboratory	high score + fixed payment	50
<i>lab</i>	laboratory	high score + performance pay	50

The game played in the experiment was a standard symmetric Cournot duopoly with linear inverse demand function $\max\{109 - Q, 0\}$ and constant marginal cost of 1. Each player's quantity q_i , $i = 1, 2$ was an element of the discrete set of actions $\{0, 1, \dots, 109, 110\}$. Player i 's profit function was given by

$$\pi(q_i, q_{-i}) := (\max\{109 - q_i - q_{-i}, 0\} - 1) q_i. \quad (1)$$

Given this payoff function, it is straightforward to compute the Nash equilibrium, which is at $q_1 = q_2 = 36$. Subjects played the Cournot duopoly repeatedly for 40 rounds with the same opponent. Each subject was matched against a computer that was programmed to one of a number of standard learning algorithms. Subjects were told that they would play against a computer but they received no further information about the algorithm. Computers were programmed to play according to noisy versions of one of the following decision rules: Best-response, fictitious play, imitate the best,

⁵In principle, subjects could have left the lab after receiving the 10 euros but no one did.

⁶Subjects were able to repeat the experiment (see Duersch et al., 2009). However, for the current paper we only use data from the first play.

reinforcement learning or trial & error.⁷

3 Results

Table 2 summarizes the main variables of interest, namely mean quantities of subjects and the average of their total profits. Note first that there are no significant differences for either quantities or profits between *net*, *lab*-np, and *lab*-f at any conventional level of significance of a two-sided Mann-Whitney U test. The average quantities in *net*, *lab*-np, and *lab*-f are almost the same at about 48 whereas in *lab* average quantity is only 43.14.⁸ All pairwise difference between *net*, *lab*-np, and *lab*-f on the one hand and *lab* on the other hand are significant at the 1% level. For *lab*, we also find higher total profits than in all three other treatments although this difference is not significant (see however the regression results below where we find a significant difference).

Table 2: Summary statistics

treatment	quantities (std. dev.)	profits (std. dev.)	decision time	aggressive
<i>net</i>	48.69 (10.76)	40,283 (16,894)	7	17%
<i>lab</i> -np	48.85 (12.09)	40,580 (16,955)	6	15%
<i>lab</i> -f	48.21 (10.56)	39,856 (16,806)	5	14%
<i>lab</i>	43.14 (7.42)	44,633 (13,480)	7	6%

Note: “Quantities” are averages over subjects and rounds. “Profits” are the mean of total profits across subjects. “Decision time” is median decision time per round in seconds. “Aggressive” is the % of subjects choosing quantities of at least 50 for at least 36 rounds.

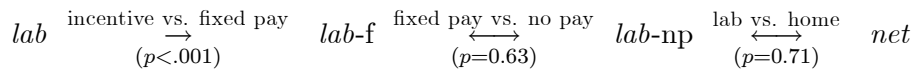
Confirming findings by Anderhub et al. (2001) and Shavit et al. (2001), variances on the *net* are higher than in *lab* (F-test, $p < 0.05$) for both, quantities and profits. However, this also seems to be driven by incentives as variances in *lab*-np and *lab*-f are not significantly different from those in

⁷See our companion paper for details about the learning processes (Duersch et al. 2009).

⁸In a different context, Shavit et al. (2001) find that bids in a lottery evaluation task are significantly higher on the internet than in a classroom experiment.

net.

The differences in mean quantities between *lab* and *net* can be summarized as follows.⁹



The first arrow indicates a significant increase in quantities between *lab* and *lab-f*. The next two arrows indicate that there are no significant differences at any conventional level between *lab-f*, *lab-np*, and *net*. We conclude that the difference between *lab* and *net* is primarily driven by the lack of monetary incentives in *net* and probably less so by the environment of the decision maker (internet vs. laboratory).¹⁰ Most importantly, given that even in the laboratory, there are substantial and highly significant differences in average quantities between *lab* and *lab-f*, we can reject the hypothesis that a high score provides adequate and non-distortionary incentives for all decision tasks.

Table 2 also presents the median decision times of subjects per round. Differences across treatments are small, which indicates that the differences between *lab* and the other treatments are not driven by differences in attention levels. We find, however, that subjects in *lab* seem to play less aggressively than those in the other treatments as they choose less frequently the kind of quantities a Stackelberg leader would choose.

For a more detailed look at the data, we ran OLS regressions to explain average quantities of subjects and average total profits (see Table 3). Confirming the MWU-tests above, Table 3 shows that average quantities in *lab* are significantly lower than in *net* (at the 1% level). But now also average profits are significantly higher in *net*. Again, there is no significant difference for average quantities or profits between *net*, *lab-f*, and *lab-np*. These effects are robust to inclusion of the learning algorithms the subjects were

⁹The *p*-values shown in parentheses refer to subjects' mean quantities using a two-sided Mann-Whitney U test, treating each subject as one observation.

¹⁰For the latter conclusion to hold, we make the (probably not too implausible) assumption that the marginal effect of providing monetary incentives is the same in the laboratory and on the internet.

Table 3: OLS regressions

	dependent variable:			
	avg. quantity		avg. total profit	
<i>lab</i>	-4.68***	(1.69)	4594**	(1941)
<i>lab-f</i>	-1.00	(2.87)	893	(3299)
<i>lab-np</i>	-1.62	(2.82)	-1249	(3251)
economics	-1.49	(1.00)	3375***	(1147)
game theory	-2.10*	(1.18)	-461	(1358)
researcher	-1.63	(1.41)	3332**	(1615)
female	-2.57***	(0.92)	-484	(1056)
best reply	3.30***	(1.22)	10261***	(1405)
fictitious play	-2.45**	(1.21)	8960***	(1397)
imitation	-1.98	(1.25)	-22280***	(1433)
trial & error	1.00	(1.25)	3558**	(1432)
start 35	-0.17	(1.03)	1816	(1181)
start 45	-0.99	(1.14)	-842	(1309)
constant	70.45***	(11.33)	26258**	(13017)
Observations	705		705	
adj. R^2	0.07		0.49	

Note: *** significant at 1%-level; ** significant at 5%-level; * significant at 10%-level; Standard errors in parentheses. Explanatory variables are dummies for the experimental setting (*lab*, *lab-f*, *lab-np*) with *net* the default, dummies for whether subjects had any training in economics or game theory, dummy for subjects who are researchers, dummies for female subjects, and dummies for the learning algorithm (best reply, fictitious play, imitation, trial & error, with reinforcement as default), and dummies for the starting value of the computer (35 or 45 with 40 as default).

matched against and the starting values of the computer. Interestingly, female subjects tend to choose lower quantities. Finally, it is reassuring that subjects with some training in economics and researchers obtained significantly higher profits.

4 Conclusion

Our experiment provides some methodological lessons with respect to internet experiments. The option of providing incentives through a high score table seems attractive since it is inexpensive and reduces the organizational burden on the experimenter. However, we show in this experiment that results obtained with a high score differ significantly from those obtained with the usual cash incentives. This holds independently of whether the experiment is run on the internet or in the laboratory. Additional problems that could occur if a highscore is used, are distortions due to other regarding preferences (since the comparison to the payoffs of others is so obvious in a highscore) and distortions due to the fact that subjects might try to make to top position in the highscore at all cost rather than obtaining the best expected placement, which may encourage excessive risk taking. For future (internet) experiments, we would thus suggest the use of significant and performance based financial incentives.¹¹

References

- [1] Anderhub, V., Müller, R., and Schmidt, C. (2001), “Design and evaluation of an economic experiment via the Internet”, *Journal of Economic Behavior & Organization*, 46, 227-247.
- [2] Bosch-Domenech, A., Montalvo, J., Nagel, R., and Satorra, A. (2002), “One, two, (three), infinity, . . . : newspaper and lab beauty contest experiments”, *American Economic Review* 92 (5), 1687–1701.

¹¹A recent internet experiment by Oechssler et al. (2008) shows that there are also significant differences between incentives through cash payments for each subject and through a lottery scheme with large cash prizes even if both schemes are equivalent with respect to expected value.

- [3] Bossaerts, P. and Plott, C. (2004), “Basic principles of asset pricing theory: evidence from large-scale experimental financial markets”, *Review of Finance* 8, 135–169.
- [4] Charness, G., Haruvy, E., and Sonsino, D. (2007), “Social distance and reciprocity: the internet vs. the laboratory”, *Journal of Economic Behavior & Organization*, 63, 88-103.
- [5] Drehmann, M., Oechssler, J. and Roeder, A. (2005), “Herding and contrarian behavior in financial markets”, *American Economic Review*, 95(5), 1403-1426.
- [6] Drehmann, M., Oechssler, J. and Roeder, A. (2007), “Herding with and without payoff externalities — an internet experiment”, *International Journal of Industrial Organization* 25, 391–415.
- [7] Duersch, P., Kolb, A., Oechssler, J., and Schipper, B. (2009), “Rage Against the Machines: How Subjects Play Against Learning Algorithms”, forthcoming *Economic Theory*.
- [8] Forsythe, R., Nelson, F., Neumann, G.R., and Wright, J., (1992), “Anatomy of an experimental political stock market”, *American Economic Review* 82, 1142–1161.
- [9] Forsythe, R., Rietz, T.A., and Ross, T.W. (1999), “Wishes, expectations and actions: a survey on price formation in election stock markets”, *Journal of Economic Behavior & Organization* 39, 83–110.
- [10] Greiner, B., Jacobsen, H.-A., Schmidt, C. (2002), “The virtual laboratory infrastructure for online economic experiments”, in: Kremer, K., Macho, V. (eds.), *Forschung und Wissenschaftliches Rechnen. Gesellschaft für Wissenschaftliche Datenverarbeitung*, Göttingen, 59–73.
- [11] Güth, W., Schmidt, C., and Sutter, M. (2003), “Fairness in the mail and opportunism on the internet—a newspaper experiment on ultimatum bargaining”, *German Economic Review* 4 (2), 243–265.

- [12] Lucking-Reiley, D. (1999), “Using field experiments to test equivalence between auction formats: magic on the internet”, *American Economic Review* 89 (5), 1063–1080.
- [13] Oechssler, J., Roider, A., and Schmitz, P. (2008), “Cooling-Off in Negotiations - Does It Work?”, mimeo, University of Heidelberg.
- [14] Shavit, T., Sonsion, D., and Benzion, U. (2001), “A comparative study of lottery-evaluation in class and on the Web”, *Journal of Economic Psychology*, 22, 483-491.