Volume 119, issue 3, June 2013

ISSN 0165-1765

economics letters

ELSEVIER

ECONOMICS LETTERS

Vol. 119, issue 3 (2013) pp. 229–358

economics letters

www.elsevier.com/locate/ecolet

Abstracted/indexed in: Contents of Recent Economics Journals, Current Contents, Journal of Economic Literature, Mathematical Reviews, RePec, Social Sciences Citation Index, World List of Social Science Periodical and Zentralblatt MATH. Also covered in the abstract and citation database SciVerse SCOPUS®. Full text available on SciVerse ScienceDirect®

CONTENTS

G. Li, H. Peng and T. Tong, Simultaneous confidence band for nonparametric fixed effects panel data models 229

J. Hualde, A simple test for the equality of integration orders 233

A. Anagnostopoulos and Q. Li, Consumption taxes and precautionary savings 238

A.C. Dammert and J. Galdo, Program quality and treatment completion for youth training programs 243

J. Westerlund and J.-P. Urbain, On the estimation and inference in factor-augmented panel regressions with correlated loadings 247

M. Hanusch and P.M. Vaaler, Credit rating agencies and elections in emerging democracies: Guardians of fiscal discipline? 251

K. Boudt, J. Cornelissen and C. Croux, The impact of a sustainability constraint on the mean-tracking error efficient frontier 255

R. Midjord, Full implementation of rank-dependent prizes 261

D.A. Peel, Heterogeneous agents and the implications of the Markowitz model of utility for multi-prize lottery tickets 264

L. Bloch, Entry of firms and cost of disinflation in New Keynesian models 268

W. Cai and M. Pandey, Size-dependent labor regulations and structural transformation in India 272

S. Brown and K. Taylor, Reservation wages, expected wages and unemployment 276

P. Louis, E. Van Laere and B. Baesens, Understanding and predicting bank rating transitions using optimal survival analysis models 280

G. Wan and Y. Zhang, Chronic and transient poverty in rural China 284

T.H. Jørgensen, Structural estimation of continuous choice models: Evaluating the EGM and MPEC 287

A. Glass, K. Kenjegalieva and J. Paez-Farrell, Productivity growth decomposition using a spatial autoregressive frontier model 291

F. Walsh, The union wage effect and ability bias: Evidence from Ireland 296

T. Trimborn, Solution of continuous-time dynamic models with inequality constraints 299

H. Beladi, L. Liu and R. Oladi, On pollution permits and abatement 302

D. Vinogradov and E. Shadrina, Non-monetary incentives in online experiments 306

M. Kuecken and M.-A. Valfort, When do textbooks matter for achievement? Evidence from African primary schools 311

B. Bakó and A. Kálecz-Simon, Quota bonuses with heterogeneous agents 316

T. Gries and D. Meierrieks, Do banking crises cause terrorism? 321

S. Kallbekken and H. Sælen, 'Nudging' hotel guests to reduce food waste as a win–win environmental measure 325

G. Tian and Y. Zhang, When can we do better than autarky? 328

C.-Y. Wang, Y.-F. Chen and C.-W. Yu, Managerial optimism and post-financing stock performance in Taiwan: A comparison of debt and equity financing 332

(Contents continued on inside back cover)

0165-1765(201306)119:3;1-P

Economics Letters 119 (2013) 306–310

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Non-monetary incentives in online experiments☆

Dmitri Vinogradov[a,*], Elena Shadrina[b]

[a] University of Essex, United Kingdom
[b] Higher School of Economics, Russia

HIGHLIGHTS

- Online experiments with no monetary incentives often provide biased results.
- The bias can be remedied by controlling for subjects' inherent motivation.
- Subsample with high motivation generates results identical to labs.
- Difference between subsamples with high and low motivation cannot be explained by other factors.
- Time spent per question (attention) is a poor predictor of motivation although somewhat similar in effect.

ARTICLE INFO

Article history:
Received 19 April 2012
Received in revised form
18 February 2013
Accepted 8 March 2013
Available online 15 March 2013

JEL classification:
C71
C93
D01
D81

Keywords:
Incentives
Online experiments
Ellsberg experiment

ABSTRACT

Monetary incentives in online experiments are not always easy to implement. Yet online experiments are advantageous in terms of a natural decision-making environment, less stress on participants and a large number of the latter. Can we obtain plausible results from online experiments by using non-monetary incentives like altruism and curiosity? We investigate the role of non-monetary incentives in a simple Ellsberg-type experiment which can be easily compared to similar lab experiments.

1. Introduction

Comparative advantages of online (Internet) experiments over lab experiments include (1) a large number and variety of participants, (2) natural setting, no pressure of an artificial lab environment, at (3) a relatively low cost. Yet there is an ongoing debate in the literature on whether and which forms of incentives should be used in Internet experiments. Typical incentives schemes range from a cash payment to all participants (frequently used in lab experiments, rare or impossible in Internet experiments due to anonymity concerns), lottery over a large stake (applicable in Internet experiments), performance based payments, or non-monetary incentives like a performance score relative to other participants. Duersch et al. (2009) analyze several of the above and come to a conclusion that without a cash complement the high score incentive alone leads to distorted results and therefore "significant and performance based financial incentives"[1] should be used in online experiments.

We challenge this view by considering non-monetary incentives based on such behavioral patterns as curiosity and altruism rather than on the sense of rivalry and desire to win. Human resource management emphasizes non-monetary factors among incentives, which is especially true for services offered by volunteers.

☆ We thank the Editor and the Reviewer for extremely useful comments and encouragement. Our thanks also go to Orhon Can Dağtekin and Qinyi Yue for their assistance with the data collection and Yuri Vorontsov for his help with the data analysis.

* Correspondence to: Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Tel.: +44 1206874893.
E-mail address: dvinog@essex.ac.uk (D. Vinogradov).

[1] Duersch et al. (2009, p. 122).

0165-1765/$ – see front matter
http://dx.doi.org/10.1016/j.econlet.2013.03.014

Our main idea is that curiosity and altruism (and possibly other behavioral factors not covered in this paper) alone can suffice to provide plausible results from an online experiment with no additional material incentives scheme.

We run online a standard experiment with well known lab treatment results reported in various studies. The only difference is that we do not incentivize participants the way it could be done in a lab but instead control for their inherent incentives to complete the task. Our findings do not reveal significant differences between lab experiments and online results for the group of subjects with high motivation. We obtain that non-monetary incentives are strong enough to generate a significant difference between subjects who take part in the complete experiment (high motivation) and those who quit after fulfilling a part of the task (low motivation).

2. The experiment

This paper draws on Ellsberg's (1961) two-colors experiment. There are two urns, A and B, containing 100 black and red balls each; subjects know that Urn B contains exactly 50 black to 50 red balls and there are *some* red and *some* black balls in Urn A. We ask subjects two questions: from which urn they would draw a ball if they were promised a prize for drawing a black ball, and if they were promised a prize for drawing a red ball. We refer to this experiment as a standard Ellsberg experiment (task). A "task" in our setting is a choice between urn A and urn B given varying information communicated to the participants. Subjects are also asked to perform several additional tasks.[2] They are free to choose whether they continue to answer the questions or quit the experiment at any stage. Responses are registered after they select to proceed to the next screen. The two Ellsberg questions are shown in the first screen (Part A); screen number two contains five additional tasks (Part B); screen number three contains two more tasks (Part C), and finally screen number four contains one question in which subjects are asked to self-assess their proficiency in statistics (Part D). The standard Ellsberg task is used only in Part A and employed in the current study for comparison with similar experiments. The role of additional tasks is explained below.

The study is based on three independent experiments conducted in 2011–12. The first experiment was run in July 2011 on www.surveymonkey.com. The total of 1000 participants was achieved in 20 days. Participants were invited via the Facebook account of a graduate student; in addition (thus suggesting that the majority of the subjects were young people though we did not control how many of them were university students), emails were sent to a group of academics in Europe and USA with an invitation to take part in the experiment and to spread the news. It was made clear that there was no prize for participation in this experiment. The data collected for each case include the start and end dates of the response, the IP address, and the answers to the questionnaire. On average, 76.4% of subjects completed the questionnaire in full (Parts A through D), whereas 21.8% only answered to Part A and did not proceed any further. We refer to this experiment (online, no monetary incentives) as Treatment 1.

We focus on non-monetary factors that made people respond to our invitation by making the following observations. First, each person has an option to ignore the invitation. Second, if curiosity or altruism lead them to open the weblink contained in the invitation they still have an option to quit the experiment after they see the first screen with the questions for part A without answering them. Participants who quit at this stage have the lowest motivation to complete the experiment and are not registered in our data. Participants with a higher motivation answer the questions of Part A, which is required to be able to proceed to the next screen, and register their answers by pressing the button "Next". We interpret this as a higher level of motivation than that of the above group. In a similar fashion, we interpret that motivation of participants who answer parts A and B is higher than that of participants who only answered part A but lower than that of participants who answered parts A, B and C. The highest level of motivation is demonstrated by subjects who complete the experiment in full. We focus on the fact that the majority of respondents have either answered Part A and quit or proceeded to further questions and completed the experiment in full. This leaves us essentially with two levels of motivation: low (only part A completed) and high (all parts completed). As soon as we do not offer any monetary prize, this motivation can only be explained by inherent motives such as curiosity or altruism.

Our main hypothesis is that non-monetary factors like curiosity and altruism provide adequate and non-distortionary incentives. To test this hypothesis we compare the behavior of subjects with low and high motivation in the online experiment with control groups chosen from earlier reported lab-based studies, both with real and hypothetical monetary incentives, and with our results from two additional experiments, described below.

Experiment 2 was conducted in a university classroom in Russia in June 2012, offering the total of 109 subjects an opportunity to win a prize of an equivalent of $100. After subjects have answered the questionnaire, we used real urns to draw black and red balls and identified how many times subjects gave "correct guesses". Each "correct guess" is an equivalent to a lottery ticket for the above prize. Thus subjects with a higher number of "correct guesses" have higher chances to win the prize. The main problem here is to model the ambiguity in Urn B, as in the end it is a real distribution of real balls in a real urn. For our experiment, it is important that this distribution is unknown to all the subjects, and that this is common knowledge (subjects know that nobody knows the distribution, etc.) We have achieved this by publicly selecting the distribution of balls in Urn B *after* all responses to the questionnaires were collected. The selection procedure was as follows: subjects were asked to give a number between 1 and 9, after which the fraction of answers above 5 determined the fraction of red balls in Urn B. In this treatment this fraction was 62%, thus we publicly placed 62 red and 38 black balls in the urn, after which balls were drawn as in the main task. Importantly, at the time of giving responses, subjects did not know the distribution. The completion rate was 100%, not surprising for a lab experiment. As a proxy for non-monetary motivation we used the fact that part of the subjects were our colleagues and friends (aged above 25, about 60% of the sample) and the rest of the subjects were postgraduate students (all aged under 25, about 40%) not related to and never taught by either of us. We expect that the cohort of colleagues and friends have a higher intrinsic motivation. This experiment is referred to as Treatment 2 (lab, with monetary prize).

In their informal feedback many subjects of Treatment 2 revealed that they would have participated in the experiment even without a monetary prize. Thus monetary incentives can affect subjects to different degrees. To control for this, we conducted experiment 3 in August–September 2012 online, mainly following the lines of Treatment 2 in what relates to the prize (£100 awarded by a type of a lottery as in Treatment 2). The distribution in Urn B (ambiguous) was taken 62% as in the previous treatment (not communicated to the subjects, thus preserving ambiguity). The invitations were sent out similarly to Treatment 1, resulting in 568 observations in total.[3] The main difference of this experiment

[2] The complete questionnaire and additional results are reported in the companion paper.

[3] 35% of the sample report themselves as employees, 16% as postgraduate and 30% as undergraduate students.

Table 1
Distribution of subjects who prefer the risky urn ("Risky" fraction) in experiments.

	N	Risky	Real incentives	Comment
HE(1986)	274	0.77	No	0.77 of non-indifferent subjects (66% of the sample)
TVW(2008)	63	0.73	$15–16	Evaluation by others
KP(2003)	61	0.75	No	Standard Ellsberg
Current study				
Treatment 1	1000	0.71	No	Whole sample
	219	0.59	No	Low motivation
	765	0.74	No	High motivation
Treatment 2	109	0.55	$100 (lottery)	Lab, whole sample
	49	0.51	–	low motivation
	58	0.60	–	High motivation
Treatment 3	568	0.63	£100 (lottery)	Online, whole sample
	81	0.57	–	Low non-monetary motivation
	487	0.64	–	High non-monetary motivation
	166	0.63	–	Low monetary motivation
	403	0.64	–	High monetary motivation

HE(1986) = Hogarth and Einhorn (1986); TVW(2008) = Trautmann et al. (2008); KP(2003) = Kühberger and Perner (2003). For HE the fraction of subjects who chose the risky prospect is taken relative to the number of subject who are not indifferent between the two prospects.

(referred to as Treatment 3) is that subjects can leave their email address to be contacted if they win the prize (anonymity was promised, the prize was sent by a bank check), giving us a natural proxy for the extent to which subjects care of monetary incentives (dummy variable; if subjects did not leave email, their monetary incentive is treated as low, otherwise high). 166 subjects did not leave their email addresses, the remaining 402 did so.

3. Results

We use results for standard Ellsberg experiments conducted by Hogarth and Einhorn (1986) with 274 MBA students at the U. Chicago, Kühberger and Perner (2003) with 160 psychology students in Salzburg, and Trautmann et al. (2008) with 63 students from two universities in Holland (their result is slightly biased by a fear of negative evaluation) as a comparison benchmark. First, we only focus on the fraction of subjects who prefer the risky prospect over the ambiguous one ("risky" fraction). Table 1 summarizes the details of these experiments and compares them with our study.

In Treatment 1, the fraction of participants who prefer the risky urn when answering a single Ellsberg question, is slightly lower than in the benchmark studies if we consider the whole sample without taking motivation into account. However we observe a significant difference ($p < 0.005$)[4] between these fractions in the subsamples of completed (high motivation) and incomplete surveys (low motivation). The results from this online experiment with non-monetary incentives closely replicate those from the benchmark experiments if controlled for the level of motivation.

In Treatment 2, despite the monetary prize, the result on the whole sample is much lower, however the subsample with high motivation produces a higher fraction of those who prefer the risky urn (though statistically the difference in means is not significant, $p = 0.338$). Since both high and low motivation subsamples produce results below those in Treatment 1, a direct comparison with benchmark experiments is not possible.[5]

In Treatment 3 we compare subsamples with high/low motivation (as in treatment 1) and high/low monetary incentives (proxied by the willingness of the subjects to enter the prize draw). As reported in Table 1, there is a pronounced effect of non-monetary motivation ($p = 0.199$). Whether subjects care about the monetary prize or not, does not seem to have any effect: the "risky" fractions in the subsamples of low and high monetary incentives are about the average for the whole sample (insignificant difference, $p = 0.748$). However intrinsic motivation has a noticeable effect on the "risky" fraction in both subsamples, as shown in Table 2 (p-values for differences in means reported in the table). The effect of intrinsic motivation is stronger than that of monetary in terms of both frequencies of "risky" choices and p-values for differences between them.

Table 2
Distribution of subjects who prefer the risky urn ("Risky" fraction) in Treatment 3 in the subsamples with high and low intrinsic motivation (proxied by the completion of questionnaires) and with high and low monetary incentives (proxied by the willingness to enter the prize draw). *P*-values refer to differences in means in respective rows and columns.

	High monetary		Low monetary		*p*-value
	Risky	Total *N*	Risky	Total *N*	
High intrinsic	0.64	376	0.67	112	0.544
Low intrinsic	0.52	27	0.59	54	0.532
p-value	0.214		0.334		

Although our estimates of the "risky" fraction in Treatment 2 and Treatment 3 differ from the benchmark from other studies, we observe that subsamples with low motivations exhibit a larger deviation from the benchmark. Can distortions generated by a lack of motivation be explained by other factors associated with motivation? For example, familiarity with statistics can indicate that subjects are interested in the field and therefore would have a higher motivation than those with poor knowledge of statistics. We address this issue by analyzing the distributions of ambiguity attitudes in the sample. We identify an individual as *ambiguity averse* if (s)he prefers urn B (risky) to urn A (ambiguous) when asked to bet on red, and also prefers urn B to urn A when asked to bet on black. A subject is an *ambiguity lover* if (s)he prefers urn A both when asked to bet on red and when asked to bet on black. Finally the subject's choice is consistent with the expected utility paradigm if (s)he prefers different urns when asked to bet on different colors (*EU subject*). Table 3 summarizes the distribution of answers in various subsamples.

The effect of motivation is visible in the lower fraction of ambiguity averse participants, higher fraction of ambiguity lovers and higher fraction of EU subjects in the low motivation subsample as compared to the whole sample and to the subsample with higher

[4] Here and below we report p-values from the T-test for difference in means on independent samples.

[5] Though it should be noted that in a sample with a noticeable fraction of ambiguity averse subjects, the "risky" fraction is expected to be significantly above 0.5, thus the low motivation subsample makes ambiguity aversion almost undetectable.

Table 3
Joint distribution of choices A and B when asked to bet on black or red.

	Treatment 1			Treatment 2			Treatment 3		
	N	A (red)	B (red)	*N*	A (red)	B (red)	*N*	A (red)	B (red)
All subjects	1000			109			568		
A (black)		*0.223*	0.056		*0.220*	0.083		*0.264*	0.032
B (black)		0.070	**0.651**		0.229	**0.468**		0.102	**0.601**
Low motivation	219			58			81		
A (black)		*0.297*	0.078		*0.224*	0.122		*0.321*	0.049
B (black)		0.110	**0.515**		0.265	**0.388**		0.111	**0.519**
High motivation	765			49			487		
A (black)		*0.200*	0.050		*0.211*	0.053		*0.255*	0.029
B (black)		0.059	**0.692**		0.175	**0.561**		0.101	**0.616**
Poor STATS	232			27			318		
A (black)		*0.185*	0.069		*0.296*	0.037		*0.245*	0.035
B (black)		0.082	**0.664**		0.111	**0.556**		0.110	**0.462**
Good STATS	322			16			167		
A (black)		*0.220*	0.040		*0.176*	0.353		*0.252*	0.018
B (black)		0.037	**0.702**		0.117	**0.353**		0.072	**0.659**

$A(c)$ and $B(c)$ = fraction of individuals who chose urn A or B respectively when asked to bet on color c = red or black. Ambiguity averse shown in bold; ambiguity-lovers in italic. "Poor/good STATS" = subjects with poor/good knowledge of statistics (self-assessment at 1–2 or 4–5 respectively on a five-point scale).

Table 4
OLS regressions.

	Treatment 1			Treatment 3					
	AA	AL	EU	AA	AL	EU	AA	AL	EU
MOT	0.159***	−0.084**	−0.074**	0.317***	−0.314***	−0.002	0.344***	−0.386***	0.414
	(0.042)	(0.037)	(0.029)	(0.081)	(0.075)	(0.054)	(0.090)	(0.084)	(0.061)
ATQ	0.005**	−0.002	−0.002*	0.005**	−0.0004	−0.004***	0.005**	−0.007	−0.005***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
MON							0.037	−0.096*	0.059
							(0.054)	(0.050)	(0.036)
Const.	0.431***	0.337***	0.232***	0.171***	0.596***	0.232***	0.131	0.698***	0.169**
	(0.060)	(0.053)	(0.042)	(0.099)	(0.093)	(0.017)	(0.115)	(0.106)	(0.078)
Adj. R^2	0.019	0.005	0.008	0.033	0.033	0.017	0.032	0.039	0.002
Obs.	815	815	815	462	462	462	462	462	462

Note: MOT is a dummy for high motivation. ATQ is the average time per question. MON is a dummy for monetary incentives as proxied by the willingness to enter the prize draw. Values in parentheses are standard errors.
*** Significant at 0.01 level.
** Significant at 0.05 level.
* Significant at 0.10 level.

motivation (statistical significance is in the regression analysis below). The lower rate of ambiguity averse subjects in the low motivation subgroup can be explained by a bias towards the uniform distribution if subjects with low motivation enter random answers.

Since it could also be argued that the low motivation subgroup differs from the rest of the sample in their familiarity with statistics, we also present the distribution of ambiguity attitudes for subjects with poor (self-assessment at 1 or 2) and good (self-assessment at 4 or 5) knowledge of statistics in the high motivation group (familiarity with statistics is unavailable for low motivation subsample by the choice of the motivation proxy in online treatments). Effects of familiarity with statistics and low motivation are qualitatively different: the fraction of ambiguity lovers in the subsample with poor stats in online treatments[6] is lower than in the whole sample and in the subsample with good stats, whereas the same fraction in the low motivation subsample is higher than in the whole sample. Therefore the results in the low motivation group cannot be explained by familiarity with statistics.

Finally, we ask whether motivation can be detected by other means, like average time spent on a question (ATQ). ATQ can be associated with attention paid to the task. In treatment 1 the mean ATQ is 401.49 s, whereas the median ATQ is only 22.2 s. This is due to observations (subjects) with extremely high ATQ (more than an hour). As this could be explained by a broken Internet connection, any other technical factors, as well as distraction, lack of attention, motivation, etc., which we cannot explore any further, we drop observations with ATQ above 40 s. This leaves us with 815 observations.[7] For this subsample we check if fractions of ambiguity averse (AA), ambiguity loving (AL) and EU — consistent subjects (EU) resulting from our experiment can be explained by motivation and ATQ. OLS regression results are in Table 4.[8] Confirming the results from Table 3, motivation is found to have a significant effect on the obtained fractions of subjects with given ambiguity attitudes (significance improves if ATQ is dropped from the regressions). Although ATQ is found significant for AA and EU fractions, its impact is small (for subjects who spend 40 s on a question the probability of being AA is only 10% higher than for the median subject) yet has the same sign as motivation. In Treatment 3 we have the advantage of comparing the effects of monetary and non-monetary incentives, and therefore run similar regressions with and without a dummy for monetary incentives. Crucially, monetary incentives have no influence on the results (insignificant and relatively small

[6] In the lab treatment the number of observations in subsamples with poor and good stats is insufficient for conclusions.

[7] For comparison, 90, 95 and 99 percentiles yield mean ATQ of 60, 100 and 340 s respectively. Dropping observations using adequate thresholds for these percentiles yields neither significant not meaningful results.

[8] Logit and probit yield similar results in terms of significance but are less tractable in terms of the interpretation of the coefficients.

coefficients) if intrinsic motivation is controlled for (though dropping intrinsic motivation makes monetary incentives significant).

4. Conclusion

Our results have implications for the design of online (Internet) experiments. We find that motivation of subjects is responsible for possible biases in the results from Internet experiments as compared with those held in labs (both with monetary incentives and without). Motivation in our setting is explained by curiosity or altruism as reflected in subjects' willingness to participate and fully complete the experiment despite being aware that there is no remuneration or prize awarded upon the completion of the experiment. Subjects with high motivation can be detected by the number of questions answered (number of rounds played). The subsample with high motivation yields results closer to those obtained in labs with monetary incentives. Running an experiment with a control for both intrinsic and monetary incentives we obtain that the latter are insignificant for the results (although regain significance once non-monetary motivation is dropped from the regressions)

An Internet experiment should therefore be designed in such a way that it allows for the control of motivation. As in our example, this can be done through designing several rounds of questions and checking results for subsamples with different completion levels. Time per question (or time per round) is a poor indicator of motivation although we obtain that its impact on the observed variables has the same sign as motivation. Using monetary incentives can still be a better motivation than inherent factors like curiosity and altruism and would possibly dominate them thus making the control of motivation less important in experiments with sufficiently high monetary remuneration. Our results rather indicate that if monetary incentivization is not a feasible option then controlling for motivation is a useful tool to avoid possible biases in results.

References

Duersch, P., Oechssler, J., Schipper, B., 2009. Incentives for subjects in internet experiments. Economics Letters 105, 120–122.
Ellsberg, D., 1961. Risk, ambiguity and the savage axioms. The Quarterly Journal of Economics 75 (4), 643–669.
Hogarth, R.M., Einhorn, H.J., 1986. Decision making under ambiguity. The Journal of Business 59 (4–2), 225–250.
Kühberger, A., Perner, J., 2003. The role of competition and knowledge in the Ellsberg task. Journal of Behavioral Decision Making 16, 181–191.
Trautmann, S.T., Vieider, F.M., Wakker, P.P., 2008. Causes of ambiguity aversion: known versus unknown preferences. Journal of Risk and Uncertainty 36 (3), 225–243.