Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Efficiency and resistance to extinction of lottery-based incentives in human: Survey response behavior in 12-week real-world field experiment

Satoshi Nakano、Ryo Kato、Makito Takeuchi、Takahiro Hoshino

1 November, 2020
DP2020-020
https://ies.keio.ac.jp/en/publications/13455/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
1 November, 2020

Efficiency and resistance to extinction of lottery-based incentives in human: Survey response behavior in 12-week real-world field experiment

Satoshi Nakano、Ryo Kato、Makito Takeuchi、Takahiro Hoshino

Keio-IES DP2020-020

1 November, 2020

JEL Classification: D10; D91; M31

Keywords: Resistance to extinction; Financial incentive; Underreporting bias; Schedule of reinforcement; Survey response

Abstract

This study aims to improve survey response behavior by including schedules of reinforcement with financial incentives in a 12-week real-world field experiment. We found that lottery based-incentives produce more resistance to extinction and are also more cost-efficient than fixed honorariums, despite having the same expectations. We also found that immediate reinforcement and the high chance, low prize incentives reduce underreporting bias. In addition, we confirmed the resistance to extinction when switching incentives and sustained responses due to lottery-based incentives. Moreover, we evaluated the past reward winning experience and the habit of past responses, and found that experience influenced subsequent survey responses. Thus, participants' responses are dependent on incentive programs.

Satoshi Nakano
R&D Division, INTAGE Inc.
INTAGE Akihabara Building, 3 Kanda-Neribeicho, Chiyoda-ku, Tokyo
nakanosatoshi21@gmail.com

Ryo Kato
Research Institute for Economics and Business Administration, Kobe University
2-1, Rokkodai-cho, Nada-ku, Kobe, Hyogo
le.voiture.rouge@gmail.com

Makito Takeuchi
Faculty of Business Administration, Tohoku Gakuin University
1-3-1 Tsuchitoi, Aoba-ku, Sendai, Miyagi
makito.the.730@gmail.com

Takahiro Hoshino
Department of Economics, Keio University
2-15-45, Mita, Minato-ku, Tokyo
hoshino@econ.keio.ac.jp

Acknowledgement : This work was supported by KAKENHI 18H03209 2018-2022.

Efficiency and resistance to extinction of lottery-based incentives in human: Survey response behavior in 12-week real-world field experiment

Satoshi Nakano[1], Ryo Kato[2], Makito Takeuchi[3], Takahiro Hoshino[4]*

November 1, 2020

[1]R&D Division, INTAGE Inc., Tokyo, Japan.

[2]Research Institute for Economics and Business Administration, Kobe University, Kobe, Japan.

[3]Department of Business Administration, Tohoku Gakuin University, Sendai, Japan.

[4]Department of Economics, Keio University, Tokyo, Japan / RIKEN Center for Advanced Intelligence Project, Tokyo, Japan.

*Correspondence to: hoshino@econ.keio.ac.jp

Abstract: This study aims to improve survey response behavior by including schedules of reinforcement with financial incentives in a 12-week real-world field experiment. We found that lottery based-incentives produce more resistance to extinction and are also more cost-efficient than fixed honorariums, despite having the same expectations. We also found that immediate reinforcement and the high chance, low prize incentives reduce underreporting bias. In addition, we confirmed the resistance to extinction when switching incentives and sustained responses due to lottery-based incentives. Moreover, we evaluated the past reward winning experience and the habit of past responses, and found that experience influenced subsequent survey responses. Thus, participants' responses are dependent on incentive programs.

Keywords: Resistance to extinction; Financial incentive; Schedule of reinforcement; Survey response; Underreporting bias

JEL classification: D10, D91, M31

Introduction

Although many types of data, such as individual's location details and web log data, can now be collected automatically, data collected from surveys are still essential in certain academic studies, including studies in medical research *(1)*, political psychology *(2)*, sociology *(3)*, and economics *(4)*. However, survey response data often suffer from biases like overreporting and underreporting *(5)*. Existing studies have attempted to improve response rates by assessing insufficient effort responding (IER) *(6)* and instructional manipulation checks *(7)*. Contrarily, we focus on designing effective financial incentives for survey responses to examine whether probabilistic financial incentives drastically improve response rates.

Many experiments have shown that partial reinforcement—reinforcement that only occurs at certain ratios of time or intervals—is more likely to strength the aimed behavior and its resistance to extinction than continuous reinforcement—reinforcement that occurs every single time *(8, 9)*. In particular, many randomized controlled trials (RCT) using cash or gifts as reinforcements, called financial incentives, have been conducted to promote changes in behavior. Consequently, some research revealed that respondents offered financial incentives perform significantly better than those offered no financial incentives *(10)*, while other research indicated no significant effect *(10, 11)*. Financial incentives based on a variable ratio schedule, such as a lottery-based incentive, is also considered to change behaviors more effectively than no incentives *(12)*, though other research showed insignificant differences *(11)*. Research on survey response rates showed that financial incentives based on a fixed ratio of time tend to be effective *(13, 14)*, and financial incentives based on a variable ratio of time, namely lottery-based incentives, are more effective than no incentive *(15-17)*. Bosnjak and Tuten *(18)* compared fixed-ratio and variable ratio incentives in the context of web survey completion rates, and showed that, although the former has no advantages compared with offering no incentive, the latter doubled the completion rates.

We conduct a randomized experiment for online panel respondents to investigate: (1) whether partial reinforcement, or a lottery-based incentive, increases the survey response rate compared to the control group that has a weekly honorarium. Unlike the studies that used web survey response rates, this study is designed so that the expected incentive value is equal between the control and partial reinforcement groups; (2) which type of lottery increases the response rate: lotteries with small prizes but with higher chances, or large prizes but lower chances; (3) whether the effect of partial reinforcement lasts for a certain period; (4) whether immediate reinforcement is more effective than delayed reinforcement; (5) whether reinforced responses continue after lottery incentives cease (resistance to extinction); and (6) whether participants' response efforts are enhanced by past rewards or habit.

Data and Design

We collected scanner panel data in which respondents report every purchase of a beverage. The data were collected using INTAGE Inc., one of the largest Japanese marketing research companies. Every time participants purchase the product, they can respond to the survey by scanning its barcode. These data include the amount of purchases and the number of units purchased in stock-keeping units, regardless of where they were purchased. The product categories covered in this study are beverages including tea, coffee, milk, juice, carbonated drinks, and nutritional drinks. The reason for covering beverages is that many people, regardless of gender and age, frequently purchase beverages. Since the reward attached to one beverage is smaller than the price of the beverage, we assume that participants would not overreport. The field experiment was conducted over a 12-week period from March 8, 2020 to May 30, 2020.

We designed five reward programs (grouped A–E) and two time periods (pre-week: first eight weeks, post-week: following four weeks). The five groups consist of one control group (group A) with a weekly honorarium and four partial reinforcement groups (groups B–E). In pre-week, the control group receives a uniform weekly honorarium, while the other four groups receive lottery-based incentives. In post-week, we grant weekly honorariums to all groups A–E. This manipulation, in which the same person is rewarded in different ways between pre-week and post-week, is intended to capture the resistance to extinction when partial reinforcement is withdrawn. The partial reinforcement groups consist of a 2 × 2 experimental design. One of the factors relates to the probability and amount of lottery-based incentive, defined as: (1) high chance and low prize, and (2) low chance and large prize. Both expectation values are equal. This allows us to test whether rewarding participants with lower prizes but higher chances (or vice versa) will increase their survey responses. Another factor is the timing of the reinforcement, defined as: (1) immediate reinforcement: a lottery is drawn every time a product is scanned, and (2) delayed reinforcement: a lottery is awarded only once a day.

Table 1 shows the details of the incentive design for each group in pre-week. The expected values per week are 300 yen (about US$2.8) for all groups. The amount of 300 yen per week is reasonable compared to the reward amounts in the Syndicated Consumer Index (SCI), which is Japan's largest consumer individual-level scanner panel data managed by INTAGE Inc and has been in constant operation since 2012 with 50,000 panelists. In the case of lottery-based incentives, each group is awarded a lottery according to the odds of winning, the winning amount, and the number of lots per week, as shown in Table 1. On the days when the participants do not purchase anything, they can draw the lottery by reporting the fact that they did not purchase anything. Thus, we designated the number of lots per week at seven in groups B and C, and 10.1 in groups D and E[1)]. For groups D and E, the expected value per lottery is about 30 yen, which is much lower than the price of a beverage. Therefore, participants do not purchase additional goods to increase the likelihood of gaining a reward, as this is considered a loss.

In post-week, groups A–E receive a weekly honorarium, the same as the control group A in pre-week, of 300 yen per week, regardless of their scan record status. Participants in groups B–E were informed of this change to the reward program through an email on the morning of the first day of week 9 (May 3, 2020). We did not disclose to participants until week 8 that the rewards would change after week 9. Therefore, the data from participants up to week 8 are unaffected by the change in the incentive design.

Table 1. The incentive design for the first eight weeks

Groups	Reward programs			Opportunity to draw lots	Odds of winning	Winning amount	Number of lots per week	Expectation value
A	The control group with weekly honorarium			—				300 yen/week
B	Variable ratio	Delayed	High chance, low prize	Up to once a day	33.0%	130 yen	7	300 yen/week
C	Variable ratio	Delayed	Low chance, high prize	Up to once a day	16.5%	260 yen	7	300 yen/week
D	Variable ratio	Immediate	High chance, low prize	At each scan record	33.0%	90 yen	10.1	300 yen/week
E	Variable ratio	Immediate	Low chance, high prize	At each scan record	16.5%	180 yen	10.1	300 yen/week

We recruited a total of 700 individuals, who were randomly divided into each group (140 individuals in each group). Following this, a total of 595 individuals (A: 118

[84.3%], B: 117 [83.6%], C: 119 [85.0%], D: 120 [85.7%], E: 121 [86.4%]) partook in the survey. A chi-square test was performed on a 5 × 2 contingency table (i.e., groups × permitted/denied), with $\chi^2(4)$ = 0.560 and p = 0.967. The results did not show a significant difference. Therefore, the rate of participation in the survey was not related to the incentive design.

In the analysis, we used three regression models (see Supplementary Materials). In Model I, we regress the number of weekly scan records y_{it} in participant i, week t on the group's dummy variables, weekly variables, and their interaction terms. This model is aimed to capture the variation in scan records with the passage of weeks. We then evaluate the past reward winning experience and the habit of responding on the number of scan records thereafter. Model 2 is being evaluated in pre-week. The dependent variable z_i is defined as the number of scan records from week 3 to week 8 in pre-week, and explanatory variables are the group's dummy variables, experience, and habit. Experience is defined as the actual amount of received rewards and the habit is defined as the number of past scan records. Both experience and habit are calculated in the first two weeks in pre-week. Model III has the same structure as Model II, but it evaluates the relationship between pre-week and post-week. We evaluate the impact of experience and habit in pre-week on the number of scan records in post-week. In addition, we also calculated the parameters of the two reinforcement factors (high chance, low prize and immediate reinforcement) by combining the parameters of the groups estimated in each model using delta method *(19)*. This is aimed to assess not only between groups, but also between reinforcement schedules.

Results

A holistic view of the experimental results

Figure 1 shows the expected number of scan records per week by group calculated using Model I. In pre-week, when the reinforcement schedule differed among groups, the number of scan records increase over time, especially in groups D and E, which have an immediate reinforcement. However, a decrease in scan records over time is observed in the control group A for all weeks. In addition, in post-week, during which weekly honorariums are paid uniformly to all groups, there is a decrease in the number of scan records in all groups. However, the resistance to extinction, which is measured as a persistent high number of scan records, is seen in groups D and E, whose number of scan records in the post-week is consistently higher than that of the other groups.

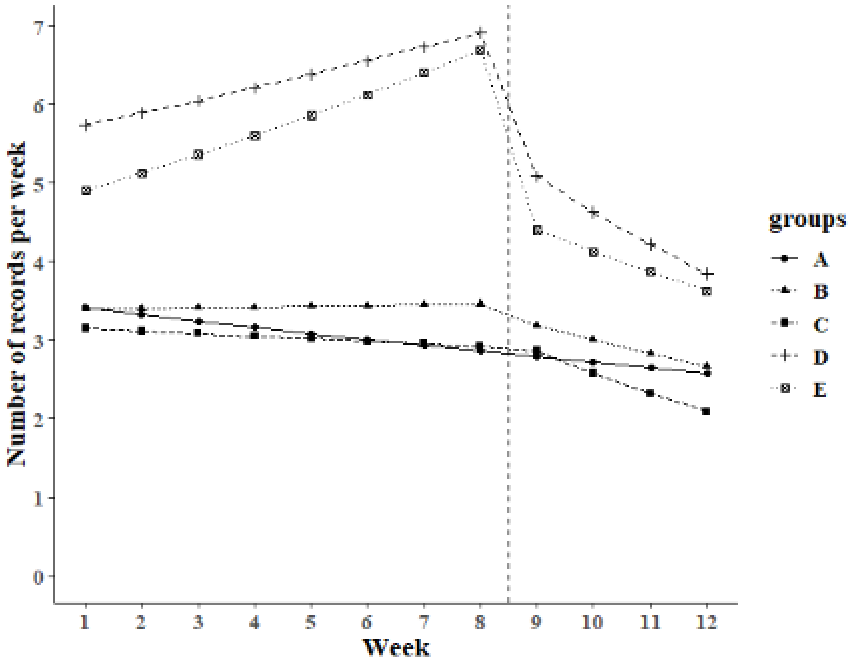

Fig. 1. The expected number of scan records per week by group. The lottery-based incentives were given to groups B–E until week 8, and weekly honorariums were given to all groups, A–E, from week 9.

Efficiency of monetary rewards and type of schedules

We find some interesting features regarding the number of scan records and efficiency of monetary rewards. In terms of rewards per record, group A earned 86 yen, group B earned 56 yen, group C earned 70 yen, group D earned 38 yen, and group E earned 43 yen (see Table S3). Groups D and E has more scan records, but the amounts of rewards per record are lower than for other groups. In contrast, group A, which was given a fixed weekly honorarium, has the highest amount of rewards per record. The reason for this lies in the cost of paying for uncooperative participants. The participants in group A were paid a fixed fee even if they did not complete a survey. As a result, uncooperative participants increased the total survey costs. Therefore, paying a lottery-based incentive to survey participants is more cost-efficient than paying a fixed honorarium.

Reporting behavior and the type of schedules

Figure 2(A) shows the differences in the number of scan records in pre-week for each group. As a test for the difference of means, we adopted a linear regression analysis with individual random effect considering multiple measurements from the same individual (see Supplementary Materials). The result indicates that groups D and E have significantly more scan records than the control group (group A) or the delayed reinforcement groups (groups B and C) have. On the contrary, no significant difference is observed in groups B and C compared to group A. Similarly, Figure 3(A) shows the impact of the groups on the number of scan records, excluding the effect of the passage of weeks estimated by Model I. This result also indicates that the participants in groups D and E have the highest survey responses. Figure 3(B) shows the effect of the passage of weeks in pre-week by group estimated by Model I. The coefficients for groups D and E are positively significant ($p < .001$), confirming an increase in the number of scan records over time.

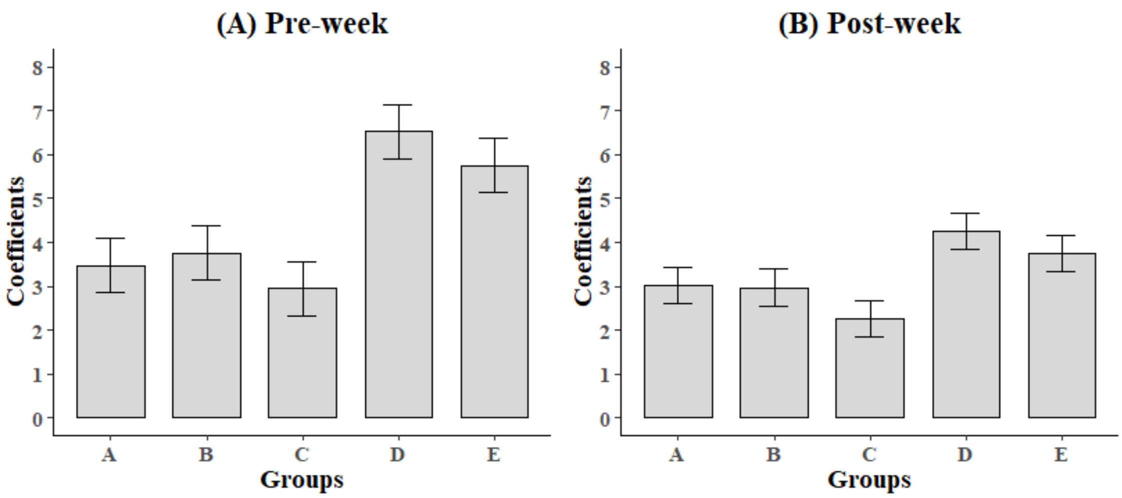

Fig. 2. Difference in scan records per week by groups. The bars indicate the mean of the number of scan records. The error bars reflect standard errors. In pre-week (panel A), the multiple comparisons using Shaffer's method shows that significant differences are found in groups C–D ($p < .001$), A–D, C–E, and B–D ($p < .01$), and A–E and B–E ($p < .10$). In post-week (panel B), significant differences are found in groups C–D ($p < .01$) and C–E ($p < .10$).

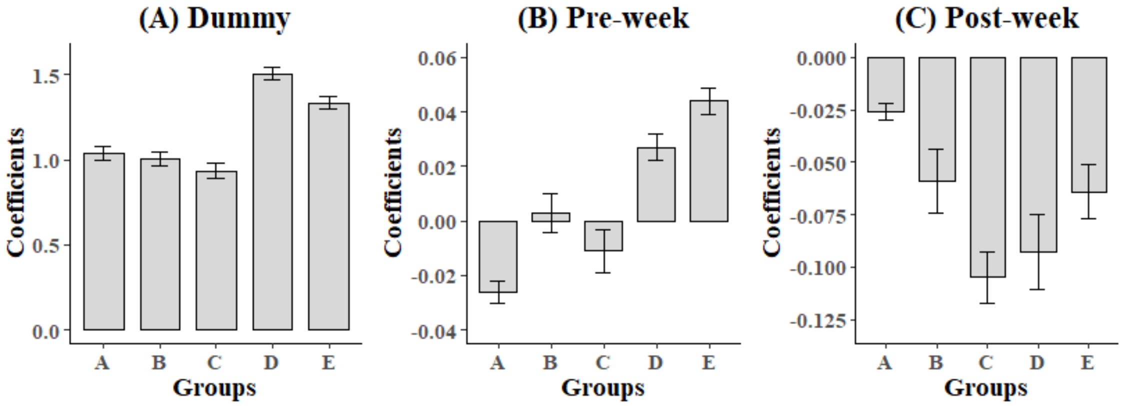

Fig. 3. Regression of the number of scan records per week on group dummies, pre-week, and post-week (Model I). The impact of the groups on the number of records excluding the effect of the passage of weeks (panel A). The effect of the passage of weeks in pre-week by groups (panel B). The effect of the passage of weeks in post-week by groups (panel C). The error bars reflect standard errors. In panel A, as a result of multiple comparisons using Shaffer's method, significant differences are found for groups C–D, B–D, A–D, C–E, B–E, and A–E ($p < .001$), and D–E ($p < .01$). In panel B, significant differences are found for groups A–E, A–D, C–E, B–E, and C–D ($p < .001$), A–B ($p < .01$), B–D ($p < .05$), and D–E ($p < .10$). In panel C, significant differences were found for groups A–C ($p < .001$), A–D ($p < .01$), and A–E ($p < .05$).

We summarized these results by group into two reinforcement schedules factors in Table 2 (Model I column). As shown in the coefficients of the dummy variables, the coefficients of both factors are positively significant, thus it shows that both reinforcement schedules are effective to improve survey responses. Compared to the coefficients between the two factors, immediate reinforcement is more effective than high chance, low prize reinforcement. In addition, looking at the coefficients of pre-week, both factors are positively significant, suggesting that both reinforcement schedules, especially immediate reinforcement, increase the number of scan records as the week progressed.

Table 2. Estimation results among factors of reinforcement schedules

Dependent variable		# of scan records per week in pre-week (Model I)				# of scan records after the third week (Model II)			# of scan records in post-week (Model III)		
		Coef	SE			Coef	SE		Coef	SE	
Control	Dummy	1.038	0.040	***							
	Week(A)	-0.026	0.004	***							
High chance, low prize	Dummy	2.506	0.067	***	Dummy	4.627	0.090	***	3.695	0.128	***
	Pre-week	0.029	0.008	***	Experience	0.955	0.072	***	1.466	0.158	***
	Post-week	-0.153	0.023	***	Habit	0.608	0.054	***	0.837	0.120	***
Immediate reinforcement	Dummy	2.837	0.064	***	Dummy	4.756	0.089	***	3.767	0.126	***
	Pre-week	0.071	0.007	***	Experience	0.828	0.063	***	1.623	0.149	***
	Post-week	-0.158	0.022	***	Habit	0.168	0.056	**	-0.808	0.138	***

Figure 4(A) shows the parameter estimates of the Model II, which is aimed to assess the impact of past winning experience and the habit of past responding behavior. The coefficients of experience are positively significant in all groups, suggesting that the experience of receiving the reward at the beginning improves subsequent survey responses. In contrast, the coefficients of habit are positively significant in groups B, C, and E. Particularly the large coefficients are obtained for groups C and B. Moreover, as shown in Table 2 (Model II column), which is summarized the results by two reinforcement schedule factors, for the high chance, low prize reinforcement, both experience and habit are related to subsequent survey responses, but habit is more effective. In contrast, for the immediate reinforcement, participants enhance their subsequent response efforts based on experience rather than habit.

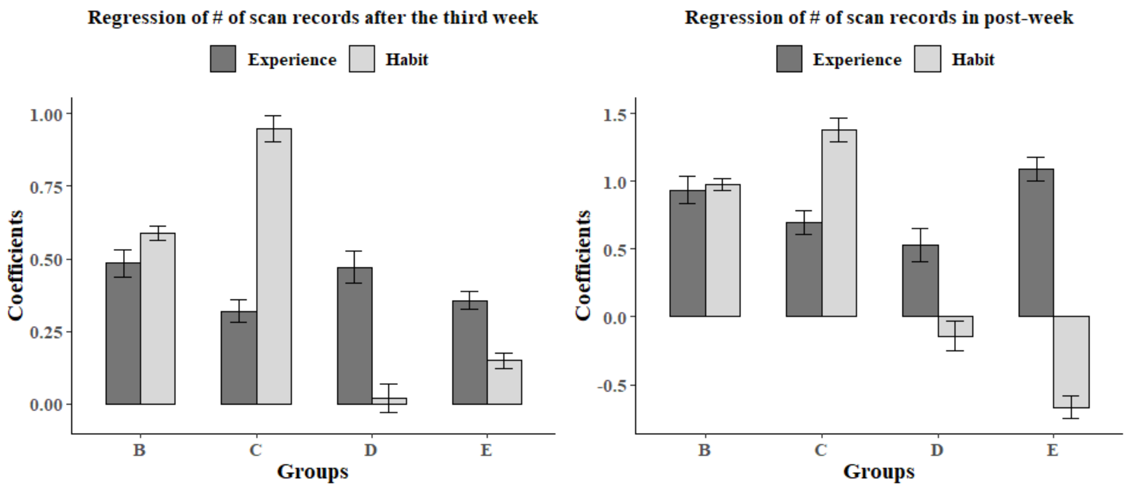

Fig. 4. Regression of the number of scan records on past winning experience and habit. The regression of number of scan records after the third week in pre-week (Model II) (panel A). The regression of number of scan records in post-week (Model III) (panel B). The error bars reflect standard errors. In Model II, for the coefficients of the experience, as a result of multiple comparisons using Shaffer's method, significant differences are found for groups B–C ($p < .05$), and C–D and C–E ($p < .10$). For the coefficients of habit, significant differences are found among all groups, C–D, C–E, B–D, B–E, and B–C ($p < .001$), and D–E ($p < .05$). In Model III, for the coefficients of the experience, significant differences are found for groups D–E ($p < .001$), C–E ($p < .01$), and B–D ($p < .10$). For the coefficients of habit, significant differences ($p < .001$) are found among all groups.

Resistance to extinction due to changes in the reinforcement schedule

Figure 2(B) shows the difference of the mean of the number of scan records in post-week by group. It is calculated similarly to the regression model with random effects in Figure 2(A). We find that the number of scan records in groups D and E are higher in both post-week and pre-week than in other groups. In addition, as shown for week 9 in Figure 1, the number of scan records for groups D and E of immediate reinforcement are high at the beginning of the post-week. Therefore, we confirm the resistance to extinction when switching incentives and sustained responses enhance by lottery-based incentives.

Furthermore, Figure 3(B) shows the effect of the passage of weeks. Among all groups, the coefficients are negative, indicating a decreasing trend. In addition, the speeds of decrease are greater in groups C, D, and E than in the control group A. Moreover, the coefficients on post-week in Table 2 (Model I column) are negatively significant ($p < .001$) for both high chance, low prize and immediate reinforcement, and the magnitude of the coefficients are comparable. This suggests that, regardless of how the reinforcement schedule was set up in pre-week, the decreasing rates of the number of scan records are comparable between the two factors.

Figure 4(B) shows the effect of winning experience and habit estimated by Model III. The coefficients of experience are significantly positive ($p < .001$) in all groups. Therefore, we confirm that the past winning experience of the lottery-based incentive affects subsequent response behavior, even if the incentive is switched to honorariums. However, the coefficients for habit are positively significant ($p < .001$) for groups B and C, and negatively significant ($p < .001$) for group E. In order to interpret these results more clearly, Table 2 (Model III column) shows the summaries from the perspective of the two factors. It shows that, in the case of immediate reinforcement, the experience, but not habit, positively affects the number of scan records, whereas, in the case of high chance, low prize, both experience and habit positively affects the number of scan records.

Discussion

We investigated the effects of financial incentives on survey response behavior. Specifically, we conducted a 12-week real-world field experiment in an RCT to examine the difference of reinforcement schedules. This field experiment was not intended to be an ad hoc survey, but rather a survey that requires long-term and ongoing responses, such as a panel survey, which has rarely been addressed in previous studies. As a result, this study suggests that the survey response effort varies substantially depending on the reinforcement schedule, even for less labor-intensive surveys, such as scan-type surveys. This means that participants' responses are dependent on the incentive program. Many experiments on animals and humans have suggested that partial reinforcement is stronger than continuous reinforcement, despite the incentives having the same expected value *(8, 9)*; however, this is the first study to show that similar results can be obtained in terms of survey response behavior. The results of this study are important to many researchers and practitioners in terms of the effect of incentives on the subject matter of surveys, a methodology that is widely used in the social sciences, medicine, and public health.

Our results showed that partial reinforcement, such as lottery-based incentive, enhanced the survey response behavior and reduced underreporting bias. In particular, immediate reinforcement and high chance, low prize incentives were most effective. In addition, switching from lottery-based incentives to weekly honorariums resulted in the resistance to extinction and sustained responses that were enhanced by lottery-based incentives. Although the extinction resistance is revealed to be enhanced more by the partial reinforcement than by the continuous reinforcement for animals *(8)*, we firstly showed that the similar phenomenon is observed for survey response behavior. Moreover, we evaluated the past reward winning experience and the habit of past responding behavior, and found that experience is related to subsequent survey responses. However, in the frequent reinforcement scenario, such as high chance, low prize incentives, both experience and habit were related to subsequent survey responses.

The implications of this study call for caution for many researchers and practitioners who design surveys. Particularly in a panel survey that requires long-term cooperation, a weekly honorarium may lead to a greater decrease in the willingness to respond to a survey over time than lottery-based incentives. To deal with the underreporting problem, researchers and practitioners may need to consider practical applications, such as giving lottery-based incentives, paying incentives according to survey response actions, or shortening the intervals between incentives. This study also suggests that a design that makes good use of the resistance to extinction is effective even if the incentives were based on weekly honorariums.

Furthermore, this study provides practical implications in terms of the management of monetary resources in the survey design. We found that when a fixed amount of money is given regardless of the status of cooperation in the survey, total expenditure was higher than when the incentives were given on a lottery-basis. This result was related to the fact that participants with lower levels of cooperation were also rewarded. Therefore, this study suggests that survey designers can more efficiently obtain the cooperation of participants if they pay a lottery-based incentive.

Finally, we discuss the limitations. The first is the issue of habituation in the case of a longer survey. In this study, the period of the lottery-based incentive was set at eight weeks, limiting the capture of the changes in the participants' behavior when partial reinforcement is applied over a longer period of time. The second limitation is checking the participants' survey responses to obtain more accurate data. In this study, the correctness of the participants' responses was assessed ex-post (see Robustness Checks in Supplementary Materials). However, it is not clear whether the participants answered correctly or not. Confirming the correctness of the responses to gain more accurate data by having the participants report visiting a store after obtaining data on their location, for example, would improve reliability and validity.

Endnote

1) The expected number of lotteries is calculated from the average number of weekly scan records for the beverage category in the SCI data. The SCI data are measured in a system very similar to the data collection in this study and are panel data that are actually operated as a business service. Therefore, we believe that the data can be used as a reference for designing incentives.

References and Notes:

1. M. M. Munn, H. Oura, M. Gallivan, K. Van Horne, A. W. Shouse, Engaging high school students in research on smoking behavior. *Science*. 341, 360-361 (2013).
2. S. P. Wojcik, A. Hovasapian, J. Graham, M. Motyl, P. H. Ditto, Conservatives report, but liberals display, greater happiness. *Science*. 347, 1243-1246 (2015).
3. C. S. Widom, S. J. Czaja, K. A. DuMont, Intergenerational transmission of child abuse and neglect: Real or detection bias? *Science*. 347, 1480-1485 (2015).
4. J. Das, A. Chowdhury, R. Hussam, A. V. Banerjee, The impact of training informal health care providers in India: A randomized controlled trial. *Science*. 354 (2016).
5. E. Lee, M. Y. Hu, R. S. Toh, Are consumer survey results distorted? Systematic impact of behavioral frequency and duration on survey response errors. *J. Mark. Res.* 37, 125-133 (2000).

6. J. L. Huang, P. G. Curran, J. Keeney, E. M. Poposki, R. P. DeShon, Detecting and deterring insufficient effort responding to surveys. *J. Bus. Psychol.* 27, 99-114 (2012).

7. D. M. Oppenheimer, T. Meyvis, N. Davidenko, Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* 45, 867-872 (2009).

8. B. F. Skinner, The extinction of chained reflexes. Proc. *Natl. Acad. Sci. U.S.A.* 20, 234 (1934).

9. A. C. Catania, *Learning* (Sloan Publishing, Cornwall on Hudson, NY, ed. 5, 2013).

10. S. M. Hall, R. G. Hall, G. DeBoer, P. O'Kulitch, Self and external management compared with psychotherapy in the control of obesity. *Behav. Res. Ther*. 15, 89-95 (1977).

11. R. W. Jeffery, R. R. Wing, C. Thorson, L. R. Burton, C. Raether, J. Harvey, M. Mullen, Strengthening behavioral interventions for weight loss: A randomized trial of food provision and monetary incentives. *J. Consult. Clin. Psychol*. 61, 1038-1045 (1993).

12. K. G. Volpp, L. K. John, A. B. Troxel, L. Norton, J. Fassbender, G. Loewenstein, Financial incentive–based approaches for weight loss: a randomized trial. *Jama*. 300, 2631-2637 (2008).

13. E. Duetskens, K. D. Ruyter, M. Wetzels, P. Oosterveld, Response rate and response quality of internet-based surveys: an experimental study. *Mark. Lett.* 15, 21-36 (2004).

14. S. D. Halpern, R. Kohn, A. Dornbrand-Lo, T. Metkus, D. A. Asch, K. G. Volpp, Lottery-based versus fixed incentives to increase clinicians' response to surveys. *Health Serv. Res*. 46, 1663-1674 (2011).

15. A. Frick, M. T. Bächtiger, U. D. Reips, "Financial incentives, personal information and drop-out rate in online studies" in *Dimensions of Internet Science* (Pabst Science Publishers, Lengerich, Germany, 1999), pp. 209-219.

16. S. R. Porter, M. E. Whitcomb, The impact of lottery incentives on student survey response rates. *Res. High. Educ.* 44, 389-407 (2003).

17. D. Heerwegh, An investigation of the effects of lotteries on web survey response rates. Field Methods. 18, 205-220 (2006).

18. M. Bosnjak, T. L. Tuten, Prepaid and promised incentives in web surveys: An experiment. *Soc. Sci. Comput. Rev.* 21, 208-217 (2003).

19. A. C. Cameron, P. K. Trivedi, *Microeconometrics: methods and applications*. (Cambridge University Press, New York, 2005), pp. 224-233.

20. A. W. Meade, S. B. Craig, Identifying careless responses in survey data. *Psychol. Methods*. 17, 437-455 (2012).

21. M. R. Maniaci, R. D. Rogge, Caring about carelessness: Participant inattention and its effects on research. *J. Res. Pers*. 48, 61-83 (2014).

22. J. L. Huang, N. A. Bowling, M. Liu, Y. Li, Detecting Insufficient Effort Responding with an infrequency scale: Evaluating validity and participant reactions. *J. Bus. Psychol*. 30, 299-311 (2015).

23. P. G. Curran, K. A. Hauser, I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *J. Res. Pers*. 82, 103849 (2019).

Supplementary Materials

A. Descriptive statistics by group

Table S1 shows the descriptive statistics by group. The average number of records per week in pre-week is higher in groups D, E, B, A, and C, in that order. However, in post-week, the records are higher in groups D, E, A, B, and C, in that order. Furthermore, corresponding to Figure 1, Table S2 shows the mean values of the number of scan records in the week by group. Group A, which received the weekly honorarium, has the highest average amount of incentives in pre-week. It should be noted that the lottery-based incentive groups B–E have a higher number of scan records, although the amounts of incentives are lower than for group A. Table S3 shows the efficiency of monetary rewards for calculating the rewards per record.

B. Models

B.1. Model I – the holistic view of the experimental results

We modeled the number of scan records y_{it} in participant i, week t, using the following zero-inflated Poisson (ZIP) regression (Model I).

$$Pr(y_{it} = k) = \begin{cases} p + (1-p)\exp(-\lambda_{it}) & ,if\ k = 0 \\ (1-p)\dfrac{{\lambda_{it}}^{y_{it}}\exp(-\lambda_{it})}{y_{it}!} & ,if\ k > 0 \end{cases} \quad \text{(B.1)}$$

where

$$\begin{aligned} \lambda_{it} = \beta_0 + \sum_{j=1}^{4} \beta_{1j} DUMMY_i(j) + \ \beta_2 WEEK_t \times DUMMY_i(A) + \\ \beta_3 PRE\_WEEK_t \times DUMMY_i(\bar{A}) + \\ \sum_{j=2}^{4} \beta_{4j} DUMMY_i(j) \times PRE\_WEEK_t \times DUMMY_i(\bar{A}) + \\ \beta_5 POST\_WEEK_t \times DUMMY_i(\bar{A}) + \end{aligned} \quad \text{(B.2)}$$

$$\sum_{j=2}^{4} \beta_{6j} DUMMY_i(j) \times POST\_WEEK_t \times DUMMY_i(\bar{A}) + \beta_7 GENDER_i + \beta_8 AGE_i$$

and

$$log\left(\frac{p}{1-p}\right) = \alpha_0 \tag{B.3}$$

The ZIP model assumes that y_{it} is 0 with probability p and Pois(λ_{it}) with probability $1-p$. The Poisson regression part of the ZIP includes dummy variables for the groups, weekly variables, and their interaction terms, gender, and age as explanatory variables. Specifically, $DUMMY_i(j)$ is a dummy variable of participant i's group (j = 1, …, 4, from group B to group E). The control group, group A, is the reference group that has the coefficient of the parameter is normalized to zero. $DUMMY_i(A)$ is a dummy variable which is equal to 1 if the participant belongs to group A, and $DUMMY_i(\bar{A})$ is a dummy variable which is equal to 1 if the participant does not belong to group A. $WEEK_t$ is a weekly variable that contains 1 to 12. In group A, which has a weekly honorarium in both pre- and post-week, we use the interaction term of $WEEK_t$ and the group A dummy. PRE_WEEK_t is a weekly variable that can be set to 1–8 in the pre-week and 0 in the post-week. Conversely, $POST\_WEEK_t$ is a weekly variable with 0 in the pre-week and 1–4 in the post-week. The interactions between PRE_WEEK_t or $POST\_WEEK_t$ and group dummy variables are set for the dummy variables of groups C-E (j=2, …, 4) with group B as the reference group. $GENDER_i$ is a dummy variable that is equal to 1 if participant i is male, and AGE_i is a continuous variable. β_0 is a constant term, and β_1-β_8 are coefficients of the explanatory variables. In addition, as in Equation (B.3), we do not set explanatory variables in the logit model part of the ZIP, but only a constant term.

B.2. Models II and III – to evaluate the experience and the habit on the number of subsequent scan records.

We evaluate the past reward winning experience and the habit of past survey responses. We modeled the number of scan records z_i using the following Poisson regression. Assuming z_i~Pois(μ_i), μ_i is the mean and variance of the distribution. The Poisson regression model is a generalized linear model that can be expressed as $log(\mu_i) = \eta_i$ using the link function of the natural logarithm. The linear predictor η_i is set up with dummy variables for the groups, experience, group dummies×experience, habit, group dummies×habit, gender, and age, as follows:

$$\eta_i = \gamma_0 + \sum_{j=1}^{3} \gamma_{1j} DUMMY_i(j) + \gamma_2 EXPERIENCE_i + \sum_{j=1}^{3} \Upsilon_{3j} DUMMY_i(j) \times EXPERIENCE_i + \gamma_4 HABIT_i + \sum_{j=1}^{3} \Upsilon_{5j} DUMMY_i(j) \times HABIT_i + \gamma_6 GENDER_i + \gamma_7 AGE_i \tag{B.4}$$

where $EXPERIENCE_i$ is the past reward winning experience, which is defined as the actual amount of past incentives, and $HABIT_i$ is the habit of past survey responses, which is defined as the number of past scan records. Note that the variables of experience and habit are standardized so that the parameters can be compared when interpreting the results. γ_0 is a constant term, and γ_1 - γ_7 are coefficients of the variables. The participants of group A are excluded from the analysis because they are given a weekly honorarium. Therefore, the analysis was conducted for those in groups B–D (n=477). The reference group was set as group C.

We set up two models with different periods of analysis. In Model II, the total number of scan records after the third week in the pre-week is the dependent variable, and the experience and the habit, up to second week, are explanatory variables. In Model III, the dependent variable is the total records in post-week, and explanatory variables are up to eighth week (pre-week).

C. The results of Models I, II and III

C1. Estimation results of Model I

Table S4 shows the ZIP regression results of Model I. As shown in Supplementary Materials D.1, the results of ZIP with combinations of parameters by the delta method are used in the table. First, the main effects of the dummy variables for each group are positively significant ($p < .001$) for all of groups A–E. Moreover, the coefficients for groups D and E are larger than those of the others. The weekly variable for group A is negatively significant ($p < .001$), indicating a gradual decrease in the number of scan records as the weekly honorarium is continued. However, with regard to the weekly variables for the pre-week of groups B–E, groups D and E are positively significant ($p < .001$) and groups B and C are not. This result suggests that immediate reinforcement promotes an increase in the number of scan records. In addition, the weekly variables in post-week of groups B–E are negatively significant ($p<.001$) in all groups. Hence, it is suggested that switching to a weekly honorarium may reduce participants' willingness to respond to surveys, compared to a lottery-based incentive.

C2. Estimation results of Model II

Table S5 shows the results of the parameter estimation of the Poisson regression model (Model II). Note that the values in this table are after the combinations of parameters shown in Supplementary Materials D.2.

The experience is positively significant ($p < .001$) for all groups. It is suggested that participants who received more money at the beginning of the survey are more willing to respond in the subsequent period. The coefficients are especially large in the high chance, low prize reinforcement schedule groups (groups B and D). Multiple comparisons using Shaffer's method for parameter equality reveal significant differences between groups B–C ($p < .05$), and C–D and C–E ($p < .10$). The effects of past scan records are positively significant ($p < .001$) in groups B, C, and E. Multiple comparisons using Shaffer's method show that there are significant differences among all groups, C–D, C–E, B–D, B–E, and B–C ($p < .001$), and D–E ($p < .05$).

In additions, the parameters of habit are larger than those of experience for groups B and C, suggesting that both experience and habit have the effects. However, the parameters of habit are smaller in groups D and E than those of experience, suggesting that only past winning experience have an effect.

Note that the correlation coefficients between experience and habit are B = 0.494, C = 0.542, D = 0.888, and E = 0.906, which are not linearly dependent, but the coefficients are high. In this regard, we compared the information criterions for the model including only experience (LL = -5353.7, the Akaike Information Criterion (AIC) = 10727.3, BIC = 10769.0), only habit (LL = -5080.4, AIC = 10180.9, BIC = 10222.5), and both experience and habit (LL = -4897.3, AIC = 9822.6, BIC = 9880.9). Based on these information criterions, we adopted the model including both experience and habit in this study.

C3. Estimation results of Model III

Table S6 shows the results of the parameter estimation of the Poisson regression model (Model III). Because the coefficients of experience are positively significant ($p < .001$) for all groups, it is suggested that participants who received more money are more willing to respond when lottery-based incentives are switched to weekly honorariums. Multiple comparisons using Shaffer's method for parameter equality confirm significant differences in D–E ($p < .001$), C–E ($p < .01$) and B–D ($p < .10$).

The parameters of the habit are positively significant ($p < .001$) in groups B and C, and negatively significant ($p < .001$) in group E. Multiple comparisons using Shaffer's method confirm that there are significant differences ($p < .001$) in the magnitude of the parameter values between all groups. For groups E and D with immediate reinforcement, the effects of experience in pre-week are stronger than those of habit. However, for groups B and C with delayed reinforcement, both experience and habit are related to the number of scan records in post-week.

Note that the correlation coefficients between experience and habit are B = 0.558, C = 0.659, D = 0.987, E = 0.973. In this regard, we compared the information criterion for the model including only experience (LL = -3165.7, AIC = 6351.4, BIC = 6393.1), only habit (LL = -2974.2, AIC = 5968.4, BIC = 6010.0) and both experience and habit (LL = -2807.6, AIC =5643.2, BIC = 5615.2). Since the model including both is the best model, we adopted it in this study.

D. Combinations of parameters

D.1. Combinations of parameters for Model I

In this section, we show the combinations of parameters in Model I using the delta method *(19)*. Table S7 shows the results of the ZIP model (Model I) before the combinations of parameters shown in Supplementary Materials B.1. For this result, the design matrix $\boldsymbol{A}$ is used to integrate the parameters, as shown in Equation D.1. Here, we show the example of the combinations of the intercept and Dummy(j).

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_b \\ \delta_c \\ \delta_d \\ \delta_e \end{bmatrix}, \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_{1b} \\ \hat{\beta}_{1c} \\ \hat{\beta}_{1d} \\ \hat{\beta}_{1e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \widehat{\boldsymbol{\delta}} = \boldsymbol{A}\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}, \boldsymbol{A}\boldsymbol{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{A}^{\boldsymbol{t}}) \tag{D.1}$$

where δ_b is a combination of the intercept (the coefficient of group A) and group B. $\delta_c, \delta_d, \delta_e$ are similarly combined with the intercept and the coefficients, respectively. $\widehat{\boldsymbol{\beta}}$ is the parameter vector of the group's dummy variables. $\boldsymbol{V}(\widehat{\boldsymbol{\beta}})$ is the variance-covariance matrix of the parameters. For the pre-week and post-week variables, the coefficients of groups C, D, and E are combined with the reference group B.

For the results of each factor presented in Table 2 (Model I column), the parameters are combined using the design matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as shown in Equation D.2. We show the example of the combinations of the intercept and Dummy(j).

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix}, \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_{1b} \\ \hat{\beta}_{1c} \\ \hat{\beta}_{1d} \\ \hat{\beta}_{1e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \tag{D.2}$$
$$\widehat{\boldsymbol{\delta}} = \boldsymbol{B}\boldsymbol{A}\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}(\boldsymbol{B}\boldsymbol{A}\widehat{\boldsymbol{\beta}}, \boldsymbol{B}\boldsymbol{A}\boldsymbol{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{A}^{\boldsymbol{t}}\boldsymbol{B}^{\boldsymbol{t}})$$

where δ_0 is a parameter of weekly honorarium, δ_1 is a parameter of high chance low prize, and δ_2 is a parameter of immediate reinforcement. The parameters of the interaction terms between pre-week and group dummies, estimated with group B as the reference group, are combined as in Equation D.3. The interaction terms between post_week and group dummies are also operated in the same way.

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_{3c} \\ \hat{\beta}_{3d} \\ \hat{\beta}_{3e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \tag{D.3}$$
$$\widehat{\boldsymbol{\delta}} = \boldsymbol{BA}\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}\left(\boldsymbol{BA}\widehat{\boldsymbol{\beta}}, \boldsymbol{BAV}\left(\widehat{\boldsymbol{\beta}}\right)\boldsymbol{A^t B^t}\right)$$

D.2. Combinations of parameters for Models II and III

Next, we show the combinations of parameters used in Tables S5 (Model II) and S6 (Model III). Table S8 shows the Poisson regression results of Model II before the combinations of parameters. This is the results of the model in Equation B.4 of Supplementary Materials B.2, where the reference group is group C. For this result, we combine the parameters using the design matrix. We show the example of the combinations of the intercept and Dummy(j) in Equation D.4. Similarly, Table S9 shows the Poisson regression results of Model III before the combinations of parameters.

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_b \\ \delta_d \\ \delta_e \end{bmatrix}, \widehat{\boldsymbol{\gamma}} = \begin{bmatrix} \hat{\gamma}_{1b} \\ \hat{\gamma}_0 \\ \hat{\gamma}_{1d} \\ \hat{\gamma}_{1e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \widehat{\boldsymbol{\delta}} = \boldsymbol{A}\widehat{\boldsymbol{\gamma}} \sim \boldsymbol{N}(\boldsymbol{A}\widehat{\boldsymbol{\gamma}}, \boldsymbol{AV}(\widehat{\boldsymbol{\gamma}})\boldsymbol{A^t}) \tag{D.4}$$

For the results of each factor presented in Table 2 (Model II and Model III columns), the parameters are combined using the design matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as shown in Equation D.5. We show the example of the combinations of the intercept and Dummy(j), where the reference group is group C. The parameters of experience and habit are also operated in the same way as in Equation D.5.

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, \widehat{\boldsymbol{\gamma}} = \begin{bmatrix} \hat{\gamma}_{1b} \\ \hat{\gamma}_0 \\ \hat{\gamma}_{1d} \\ \hat{\gamma}_{1e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \tag{D.5}$$
$$\widehat{\boldsymbol{\delta}} = \boldsymbol{BA}\widehat{\boldsymbol{\gamma}} \sim \boldsymbol{N}(\boldsymbol{BA}\widehat{\boldsymbol{\gamma}}, \boldsymbol{BAV}(\widehat{\boldsymbol{\gamma}})\boldsymbol{A^t B^t})$$

D.2. Difference test in the number of scan records between groups using a random effects model

We describe the random effects model and its combinations of parameters used to test of the differences in the number of scan records between groups. Here, we employ a linear regression model with the number of scan records y_{it} as the dependent variable and the group dummies as the explanatory variable, with individual's random effect. Since y_{it} has been measured multiple times from the same person, the measurements are correlated, and therefore we have adopted this method rather than simply used a test of the difference in means.

$$y_{it} = \beta_0 + \sum_{j=1}^{4} \beta_{1j} DUMMY_i(j) + \mu_i + \varepsilon_{it} \tag{D.6}$$

where μ_i is a random effect and assumed to be normally distributed with a mean of zero independent of the explanatory variables, and ε_{it} is an error term. $DUMMY_i(j)$ is a dummy variable for groups B to E, and, here, we analyze group A as the reference group. For the estimated parameters β_0 and β_{1j}, we performed the combinations of parameters shown in Equation D.7 and the results are shown in Figure 2.

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_a \\ \delta_b \\ \delta_c \\ \delta_d \\ \delta_e \end{bmatrix}, \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_{1b} \\ \hat{\beta}_{1c} \\ \hat{\beta}_{1d} \\ \hat{\beta}_{1e} \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \widehat{\boldsymbol{\delta}} = \boldsymbol{A}\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}, \boldsymbol{A}\boldsymbol{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{A}^{\boldsymbol{t}}) \tag{D.7}$$

E. An extension of Model I to include quadratic terms

Figure 1 shows that the number of scan records continued to increase in groups D and E in pre-week, while the number of scan records continue to decrease for all groups in post-week. This trend may be related to the weekly variables of Model I, estimated as a first-order term. Therefore, we extended the analysis by including quadratic terms for $WEEK_t$, PRE_WEEK_t and $POST\_WEEK_t$ in the Equation (B.2). Figure S1 shows the expected value of the number of scan records by group including quadratic terms. We find convex downward shapes in pre-week and convex upward shapes in post-week, suggesting that the number of scan records does not continue to fall in the post-weeks. However, after comparing the information criterions among the models with only first-order terms and the model including quadratic terms, the model with only first-order terms is adopted in several information criterions of Bayesian Information Criterion (BIC) and consistent AIC (CAIC). For the model with only first-order terms, AIC = 47027.8, BIC = 47144.6 and CAIC = 47161.6. For the model including quadratic terms, AIC = 47002.4, BIC = 47181.2, CAIC = 47207.2. Therefore, the model with only first-order terms is adopted in this study.

F. Robustness checks

F.1. Methods

In this section, we discuss the robustness checks of our results. Participants may respond falsely in a self-reported survey. In particular, when offered a form of lottery-based incentives, for which the participant is paid according to the response effort, dishonest scanning may lead to falsely implying that an effort was made. Hence, this study examines the robustness of the model's estimation results when participants are excluded for the following two cases.

First, the study excluded those who were suspected of falsely response with incorrect scan records. For example, we excluded those who scanned the same product multiple times in order to win the lottery, specifically, those who scanned the same stock-keeping units three or more times on the same day.

Second, this study excluded those who were prone to IER (insufficient effort responding). We used the following procedure to identify IER respondents. First, we conducted a questionnaire survey with participants before the experiment began and asked them about the IER scale. IER is defined as "a response set in which the respondent answers a survey measure with low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses" *(6, p. 100)*. In the case of this study, participants who are prone to IER may not cooperate earnestly in the survey and may be underreported. Although various indices have been developed to identify the participants who are prone to IER *(6, 20-22)*, we use the infrequency IER scale *(22)*, which is associated with short response time and inconsistent response to items. We use four items including “I can run 3 kilometers in 2 minutes,” “I am interested in pursuing a degree in managemental genetics,” “I have never used a computer,” “I work twenty-eight hours in a typical work day” [1)]. The response alternatives used a seven-point Likert scales (1 = strongly disagree, 7 = strongly agree). Although the infrequency IER scale originally consisted of eight items, in this study, four of the above-mentioned items were included in the questionnaire for ethical reasons presented by the marketing research firm that conducted the survey. These items are interspersed with the other items in the questionnaire to hide the fact that we were asking about IER from the participants.

The infrequency IER scale identifies IER based on the participants' agreement on items that do not exist or are not feasible. Therefore, we set IER respondents (IER1 = 1) if they gave an agreeable rating (from 5 = slightly agree to 7 = strongly agree) on any of the items and non-IER respondents (IER1 = 0) if they did not. In addition, in the item for the degree, Curran and Hauser *(23)* indicated that respondents could respond to the degree presented in the survey as interesting even if they did not know it, or they could not disagree because it might be a degree that existed. Therefore, we set only three items without using the item of the degree as IER2.

F.2. Results

We checked the robustness of the parameters by excluding: (1) those who were suspected of falsely responding with incorrect scan records, and (2) those who were prone to IER (IER1 and IER2). As the result of extracting those who were suspected of falsely responding, 11 participants remained (B: 2, D: 3, E: 6). In addition, for the IER, IER1 included 111 participants (A: 21, B: 21, C: 23, D: 28, E: 18), and IER2 included 84 participants (A: 14, B: 17, C: 17, D: 23, E: 13).

Tables S10 and S11 show the results for Model I. The results show no difference in significance or sign compared to the variables that become significant in Table S4, except for Pre_week(C) in the case of IER. In addition, in Pre_week(C), under the IER2 criterion, which is less likely to inadvertently remove non-IER respondents, the direction of the sign is the same and the result in Table S11 is marginally significant instead of non-significant in the original model, with a slight difference. Therefore, the implications of this study remain the same, suggesting that there is robustness in the results of this study.

Similarly, Tables S12 and S13 show the results for Model II. There is no change in the significance or sign of the parameters in the case excluding those who were suspected of falsely responding. However, in the case excluding those who were prone to IER, there is a tendency for the coefficient of experience to be positive and for the coefficient of habit to be negative in group D. However, the main findings of this study remain the same: in the original model, experience is more effective at improving the participant's survey response efforts than habit is. Therefore, we support the results of Model II. In addition, Tables S14 and S15 show the results for Model III. There is no change in the significance and sign of the results, suggesting that the results for Model III are robust.

Endnote

1) The scales of Huang et al. (2015) includes "I can run two miles in two minutes," and "I am interested in obtaining a degree in Parabanjology." However, since the survey was conducted among people in Japan, we use the familiar kilometer notation for the survey. In addition, since it is not possible to translate "Parabanjology" into Japanese, we have replaced the term with a non-existent degree.

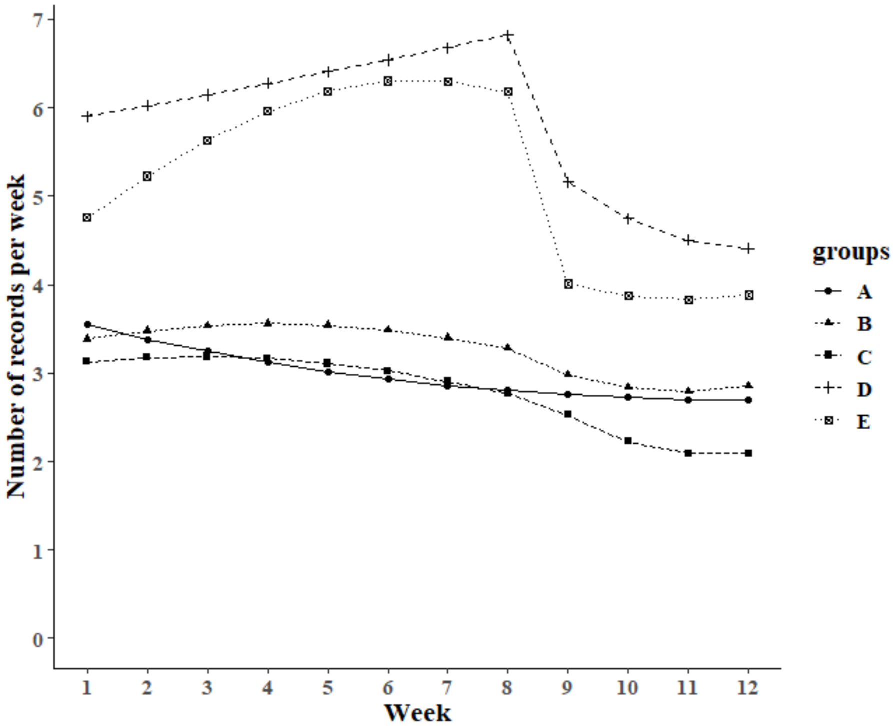

Fig. S1. The expected value of the number of scan records by group using a model including quadratic terms

Table S1. Descriptive statistics by group

		A	B	C	D	E	Total
n (participants)		118	117	119	120	121	595
n (participants-weeks)		1416	1404	1428	1440	1452	7140
Average number of records per week	mean	3.48	3.76	2.94	6.53	5.76	4.50
in the pre-week	sd	3.47	4.01	3.27	11.10	10.30	7.49
Average number of records per week	mean	3.02	2.97	2.26	4.25	3.76	3.26
in the post-week	sd	3.20	3.57	2.79	6.34	6.64	4.86
Acutual amount of rewards	mean	2.40	1.68	1.66	1.96	2.00	1.83
in the pre-week (thousand yen)	sd	0.00	0.89	1.08	2.36	2.32	1.81
Acutual amount of rewards	mean	0.60	0.41	0.39	0.45	0.46	0.43
in first two weeks of the pre-week (thousand yen)	sd	0.00	0.24	0.29	0.52	0.51	0.41
Age	mean	41.2	41.4	42.1	42.0	41.0	41.5
	sd	11.1	10.7	10.9	10.6	11.2	10.9
Male (%)		51.7	51.3	51.3	49.2	51.2	50.9

Table S2. Mean values of the number of scan records in the week by group

		groups A	B	C	D	E
weeks	1	3.95	3.64	3.34	5.94	5.27
	2	3.85	3.97	3.17	6.47	6.02
	3	3.76	3.88	3.05	6.84	5.79
	4	3.81	3.81	3.06	6.32	6.07
	5	3.24	3.95	2.73	6.53	5.37
	6	3.05	3.47	2.82	6.23	5.64
	7	3.08	3.68	2.77	6.93	5.39
	8	3.06	3.67	2.61	6.98	6.50
	9	2.87	3.33	2.45	4.51	3.90
	10	3.07	2.84	2.18	4.22	3.85
	11	3.20	2.96	2.35	4.08	3.54
	12	2.92	2.75	2.07	4.18	3.74

Table S3. Efficiency of monetary rewards

	A	B	C	D	E
Records per week	3.48 (3.47)	3.76 (4.01)	2.94 (3.27)	6.53 (11.10)	5.76 (10.30)
Rewards per week (yen)	300 (0)	211 (111)	207 (135)	246 (295)	250 (290)
Rewards per record (yen)	86	56	70	38	43

Note. For records per week and rewards per week, we show the mean (sd). For rewards per record, we divided the mean of rewards per week by the mean of records per week.

Table S4. The ZIP regression results of Model I

Poisson regression part				
	β	SE	z-value	
Dummy(A)	1.038	0.040	25.746	***
Dummy(B)	1.002	0.042	23.674	***
Dummy(C)	0.934	0.045	20.662	***
Dummy(D)	1.504	0.036	41.233	***
Dummy(E)	1.333	0.037	35.753	***
Week(A)	-0.026	0.004	-5.932	***
Pre_week(B)	0.003	0.007	0.421	
Pre_week(C)	-0.011	0.008	-1.390	
Pre_week(D)	0.027	0.005	5.316	***
Pre_week(E)	0.044	0.005	8.512	***
Post_week(B)	-0.059	0.015	-3.842	***
Post_week(C)	-0.105	0.012	-8.799	***
Post_week(D)	-0.093	0.018	-5.288	***
Post_week(E)	-0.064	0.013	-5.104	***
Gender	0.345	0.012	28.528	***
Age	0.010	0.001	16.090	***
Logistic regression part predicting records to be 0				
	β	SE	z-value	
Intercept	-0.858	0.026	-32.410	***
Log-likelihood	-23496.9			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S5. The results of Poisson regression of Model II

	γ	SE	z-value	
Dummy(B)	2.149	0.048	45.186	***
Dummy(C)	2.010	0.048	41.748	***
Dummy(D)	2.478	0.047	53.023	***
Dummy(E)	2.278	0.046	49.544	***
Experience(B)	0.485	0.047	10.237	***
Experience(C)	0.320	0.039	8.312	***
Experience(D)	0.471	0.054	8.637	***
Experience(E)	0.357	0.032	11.115	***
Habit(B)	0.588	0.025	23.814	***
Habit(C)	0.946	0.045	21.174	***
Habit(D)	0.020	0.048	0.409	
Habit(E)	0.149	0.028	5.262	***
Gender	0.069	0.018	3.715	***
Age	0.019	0.001	20.507	***
Log-likelihood	-4897.3			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S6. The results of Poisson regression of Model III

	γ	SE	z-value	
Dummy(B)	1.704	0.068	24.984	***
Dummy(C)	1.615	0.071	22.655	***
Dummy(D)	1.991	0.067	29.651	***
Dummy(E)	1.776	0.066	27.085	***
Experience(B)	0.935	0.103	9.054	***
Experience(C)	0.696	0.091	7.679	***
Experience(D)	0.531	0.120	4.416	***
Experience(E)	1.092	0.087	12.484	***
Habit(B)	0.977	0.044	22.271	***
Habit(C)	1.379	0.089	15.496	***
Habit(D)	-0.141	0.111	-1.263	
Habit(E)	-0.667	0.081	-8.219	***
Gender	-0.146	0.027	-5.497	***
Age	0.016	0.001	12.200	***
Log-likelihood	-2807.6			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S7. The ZIP regression results of Model I before the combinations of parameters

Poisson regression part				
	β	SE	z-value	
Intercept	1.038	0.040	25.746	***
Dummy(B)	-0.036	0.044	-0.817	
Dummy(C)	-0.104	0.047	-2.204	*
Dummy(D)	0.466	0.039	11.854	***
Dummy(E)	0.295	0.040	7.361	***
Week(A)	-0.026	0.004	-5.932	***
Pre_week	0.003	0.007	0.421	
Pre_week(C)	-0.013	0.010	-1.325	
Pre_week(D)	0.024	0.008	2.855	**
Pre_week(E)	0.042	0.008	4.916	***
Post_week	-0.059	0.015	-3.842	***
Post_week(C)	-0.046	0.020	-2.358	*
Post_week(D)	-0.034	0.023	-1.453	
Post_week(E)	-0.005	0.020	-0.258	
Gender	0.345	0.012	28.528	***
Age	0.010	0.001	16.090	***
Logistic regression part predicting records to be 0				
	β	SE	z-value	
Intercept	-0.858	0.026	-32.410	***
Log-likelihood	-23496.9			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S8. The Poisson regression results of Model II before the combinations of parameters

	γ	SE	z-value	
Intercept	2.010	0.048	41.748	***
Dummy(B)	0.139	0.032	4.304	***
Dummy(D)	0.468	0.030	15.445	***
Dummy(E)	0.268	0.031	8.716	***
Experience	0.320	0.039	8.312	***
Experience×Dummy(B)	0.165	0.061	2.703	**
Experience×Dummy(D)	0.150	0.067	2.256	*
Experience×Dummy(E)	0.037	0.050	0.748	
Habit	0.946	0.045	21.174	***
Habit×Dummy(B)	-0.357	0.051	-6.981	***
Habit×Dummy(D)	-0.926	0.066	-14.078	***
Habit×Dummy(E)	-0.797	0.053	-15.144	***
Gender	0.069	0.018	3.715	***
Age	0.019	0.001	20.507	***
Log-likelihood	-4897.3			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S9. The Poisson regression results of Model III before the combinations of parameters

	γ	SE	z-value	
Intercept	1.615	0.071	22.655	***
Dummy(B)	0.089	0.049	1.812	†
Dummy(D)	0.376	0.047	7.922	***
Dummy(E)	0.161	0.047	3.421	***
Experience	0.696	0.091	7.679	***
Experience×Dummy(B)	0.239	0.137	1.738	†
Experience×Dummy(D)	-0.166	0.150	-1.101	
Experience×Dummy(E)	0.396	0.125	3.155	**
Habit	1.379	0.089	15.496	***
Habit×Dummy(B)	-0.401	0.099	-4.062	***
Habit×Dummy(D)	-1.519	0.142	-10.668	***
Habit×Dummy(E)	-2.046	0.121	-16.962	***
Gender	-0.146	0.027	-5.497	***
Age	0.016	0.001	12.200	***
Log-likelihood	-2807.6			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$.

Table S10. The results of Model I excluding those who were suspected of falsely responding

Poisson regression part				
	β	SE	z-value	
Dummy(A)	1.143	0.041	28.047	***
Dummy(B)	1.099	0.043	25.487	***
Dummy(C)	1.038	0.046	22.794	***
Dummy(D)	1.527	0.038	40.141	***
Dummy(E)	1.413	0.039	36.328	***
Week(A)	-0.025	0.004	-5.898	***
Pre_week(B)	0.003	0.007	0.503	
Pre_week(C)	-0.010	0.008	-1.315	
Pre_week(D)	0.025	0.005	4.705	***
Pre_week(E)	0.025	0.006	4.311	***
Post_week(B)	-0.059	0.016	-3.785	***
Post_week(C)	-0.103	0.013	-8.140	***
Post_week(D)	-0.091	0.018	-5.155	***
Post_week(E)	-0.052	0.013	-3.932	***
Gender	0.277	0.012	22.176	***
Age	0.008	0.001	13.148	***
Logistic regression part predicting records to be 0				
	β	SE	z-value	
Intercept	-0.839	0.027	-31.570	***
Log-likelihood	-21686.8			

Note. *** p < .001, ** p < .01, * p < .05.

Table S11. The results of Model I excluding those who were prone to IER

IER1					IER2			
Poisson regression part								
	β	SE	z-value		β	SE	z-value	
Dummy(A)	0.946	0.046	20.530	***	0.972	0.044	21.913	***
Dummy(B)	0.938	0.049	19.310	***	0.914	0.047	19.249	***
Dummy(C)	0.840	0.051	16.326	***	0.854	0.050	17.166	***
Dummy(D)	1.305	0.044	29.865	***	1.287	0.043	30.097	***
Dummy(E)	1.206	0.043	28.297	***	1.190	0.042	28.610	***
Week(A)	-0.025	0.005	-5.223	***	-0.026	0.005	-5.697	***
Pre_week(B)	-0.012	0.008	-1.543		-0.004	0.007	-0.540	
Pre_week(C)	-0.022	0.009	-2.468	*	-0.016	0.008	-1.881	†
Pre_week(D)	0.019	0.006	2.925	**	0.017	0.006	2.777	**
Pre_week(E)	0.046	0.006	7.776	***	0.050	0.006	8.652	***
Post_week(B)	-0.088	0.018	-4.979	***	-0.085	0.017	-4.857	***
Post_week(C)	-0.089	0.015	-6.070	***	-0.093	0.014	-6.404	***
Post_week(D)	-0.086	0.020	-4.403	***	-0.089	0.019	-4.682	***
Post_week(E)	-0.080	0.014	-5.557	***	-0.072	0.014	-5.115	***
Gender	0.352	0.014	25.176	***	0.355	0.014	25.970	***
Age	0.011	0.001	15.999	***	0.011	0.001	16.312	***
Logistic regression part predicting records to be 0								
	β	SE	z-value		β	SE	z-value	
Intercept	-0.851	0.029	-28.910	***	-0.842	0.029	-29.460	***
Log-likelihood	-18007.6				-18895.1			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$.

Table S12. The results of Model II excluding those who were suspected of falsely responding

	γ	SE	z-value	
Dummy(B)	2.250	0.049	46.067	***
Dummy(C)	2.122	0.049	43.227	***
Dummy(D)	2.474	0.048	51.232	***
Dummy(E)	2.289	0.048	48.079	***
Experience(B)	0.514	0.048	10.639	***
Experience(C)	0.331	0.038	8.607	***
Experience(D)	0.496	0.054	9.192	***
Experience(E)	0.345	0.036	9.593	***
Habit(B)	0.592	0.025	23.723	***
Habit(C)	0.939	0.044	21.152	***
Habit(D)	0.071	0.048	1.483	
Habit(E)	0.166	0.031	5.395	***
Gender	0.049	0.019	2.571	*
Age	0.017	0.001	17.440	***
Log-likelihood	-4011.4			

Note. *** p < .001, ** p < .01, * p < .05.

Table S13. The results of Model II excluding those who were prone to IER

	IER1				IER2			
	γ	SE	z-value		γ	SE	z-value	
Dummy(B)	1.989	0.055	36.455	***	1.957	0.053	36.706	***
Dummy(C)	1.790	0.057	31.615	***	1.784	0.055	32.720	***
Dummy(D)	2.268	0.053	42.671	***	2.236	0.052	43.165	***
Dummy(E)	2.126	0.053	40.374	***	2.109	0.051	41.090	***
Experience(B)	0.374	0.051	7.363	***	0.395	0.050	7.886	***
Experience(C)	0.538	0.047	11.409	***	0.503	0.045	11.221	***
Experience(D)	0.918	0.068	13.546	***	0.857	0.066	12.908	***
Experience(E)	0.308	0.036	8.455	***	0.304	0.036	8.494	***
Habit(B)	0.627	0.032	19.796	***	0.630	0.031	20.324	***
Habit(C)	0.967	0.050	19.215	***	1.071	0.048	22.369	***
Habit(D)	-0.386	0.060	-6.407	***	-0.334	0.059	-5.650	***
Habit(E)	0.188	0.032	5.930	***	0.186	0.031	5.950	***
Gender	0.131	0.022	5.861	***	0.153	0.022	7.069	***
Age	0.022	0.001	19.903	***	0.023	0.001	21.149	***
Log-likelihood	-3892.1				-4147.68			

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S14. The results of Model III excluding those who were suspected of falsely responding

	γ	SE	z-value	
Dummy(B)	1.645	0.070	23.339	***
Dummy(C)	1.578	0.073	21.616	***
Dummy(D)	1.941	0.069	28.158	***
Dummy(E)	1.740	0.068	25.698	***
Experience(B)	0.998	0.107	9.316	***
Experience(C)	0.697	0.091	7.697	***
Experience(D)	0.535	0.122	4.379	***
Experience(E)	1.016	0.087	11.705	***
Habit(B)	0.957	0.044	21.576	***
Habit(C)	1.370	0.089	15.419	***
Habit(D)	-0.144	0.113	-1.280	
Habit(E)	-0.523	0.081	-6.494	***
Gender	-0.123	0.027	-4.503	***
Age	0.017	0.001	12.275	***
Log-likelihood	-2572.7			

Note. *** p < .001, ** p < .01, * p < .05.

Table S15. The results of Model II excluding those who were prone to IER

IER1								
	γ	SE	z-value		γ	SE	z-value	
Dummy(B)	1.885	0.077	24.345	***	1.922	0.076	25.270	***
Dummy(C)	1.845	0.084	22.087	***	1.819	0.081	22.524	***
Dummy(D)	1.998	0.075	26.666	***	2.048	0.074	27.772	***
Dummy(E)	1.877	0.074	25.306	***	1.938	0.073	26.604	***
Experience(B)	0.846	0.111	7.655	***	0.841	0.110	7.662	***
Experience(C)	0.672	0.109	6.143	***	0.834	0.104	8.050	***
Experience(D)	0.820	0.151	5.426	***	0.660	0.146	4.513	***
Experience(E)	0.785	0.108	7.260	***	0.863	0.106	8.180	***
Habit(B)	1.268	0.068	18.772	***	1.205	0.066	18.284	***
Habit(C)	1.896	0.126	15.063	***	1.532	0.101	15.138	***
Habit(D)	-0.285	0.141	-2.019	*	-0.134	0.137	-0.978	
Habit(E)	-0.380	0.100	-3.796	***	-0.450	0.098	-4.600	***
Gender	-0.167	0.032	-5.224	***	-0.214	0.031	-6.805	***
Age	0.013	0.002	8.532	***	0.013	0.002	8.211	***
Log-likelihood	-1954.2				-2077.2			

Note. *** p < .001, ** p < .01, * p < .05.