

Replication Report for “The Effect of Payment Schemes on Inventory Decisions: The Role of Mental Accounting”

By: Li Chen, Gürhan Kök, and Jordan Tong

Primary Team: Steve Leider and Samantha Keppler
University of Michigan

Secondary Team: Kyle Hyndman and Elena Katok
University of Texas at Dallas

Chen et al. (2013) find that in a newsvendor setting, orders are higher under an “own financing” payment scheme, where units are paid for up front ($-c$) and payment is earned per unit sold ($+p$), than under a “customer financing” payment scheme, where profits are earned per unit ordered ($+p - c$) and then payment is made per unit leftover ($-p$).

Hypothesis to replicate:

Subjects order higher quantities under the O-payment scheme ($-c$ per unit ordered, $+p$ per unit sold) than under the C-payment scheme ($+(p - c)$ per unit ordered, $-p$ per unit leftover), even though the two are mathematically equivalent.

Power Analysis

In the original study, the p -value is reported as $p < 0.0001$: “Orders under O were significantly greater than orders under C ($F(1, 95) = 35.84, p < 0.0001$)” (p. 442). This is based on an OLS regression model regressing average order quantity on treatment, and then conducting a contrast test. We contacted the study authors to secure the original study data, and the exact p -value is $p = 3.80 \times 10^{-8}$ for the O versus C condition (including the S condition, which we do not study in the replication). Dropping the S condition, and comparing only the O versus C conditions, the exact p -value is $p = 1.612 \times 10^{-6}$. We re-estimated the results using a regression with clustered standard errors by subjects. The p -value is still less than 0.05.

The original sample size is 50 participants (29 in treatment O and 21 in treatment C). To achieve 90% power based on the results of the standard OLS analysis, the required sample size is 15 participants. To achieve 90% power based on the results of the OLS analysis with clustered standard errors, the required sample size is 16 participants. The MS Replication Project team has adopted a policy of using the original sample size as a lower bound for replication. Further, the team agreed that no replication shall have fewer than 40 participants. In this case, the original sample size of $N = 50$ is binding.

Sample

The original study was conducted at Duke University. Participants were undergraduate and graduate students from a variety of

majors. The sample for the primary replication consists of subjects from the University of Michigan. The sample for the secondary replication consists of subjects from the University of Texas - Dallas. The target sample size for each replication is 50 subjects. Due to in-person laboratory interruptions from Covid-19, each replication was first conducted online. Subsequently, if the p-value associated with the primary hypothesis is greater than .05, that location would repeat the study in-person. In all cases, students are recruited from general laboratory populations.

For the online replication, the sample consists of 59 subjects from the University of Michigan (29 subjects in treatment O and 30 subjects in treatment C) and 57 subjects from the University of Texas - Dallas (27 subjects in treatment O and 30 subjects in treatment C). For the in-person replication at the University of Michigan, the sample is 55 subjects (28 subjects in treatment O and 27 subjects in treatment C). Some of the subjects failed the comprehension test at the end of the experiment, and their data are excluded as in the original study. Excluding these subjects, the target sample size of 50 is still achieved for each replication.

Materials

The experiment is an individual decision task. In the original study, data were collected manually. The materials involved were an instruction sheet, poker chips to represent order quantities, and three dice to represent random demand. For the online replication, the primary replication authors coded the experiment in Qualtrics using virtual representation of transactions and virtual dice to “roll.” The Qualtrics version of the original experiment follows as closely as possible the procedure the original study, and was kindly reviewed and approved by the study authors. At the University of Michigan, we also conduct the experiment as it was done originally.

Procedure

Both for the online and in-person replications, we follow the procedure of the experiment outlined in Study 1, Experimental Design (Section 4.1.1, p. 441) of the original study. First, instructions are given. Participants are informed they will be selling “widgets” and that customer demand for widgets in a given round is represented by the summed outcome of three standard dice. The widgets cost $c = \$1$ and sell for $p = \$2$. Participants start with \$100. At the beginning of each round, participants are asked to decide on an order quantity. Then, the participants roll three dice and demand is realized. The game is played for 25 rounds. Under the O condition, participants pay c for the widgets at the point they are ordered and then get p if they are sold. Under the C condition, participants get margin $p - c$ for the widgets at the point they are ordered and then pay back p for widgets not sold. The pre-registration report for the experiment is available at <https://aspredicted.org/5jq65.pdf>.

Participants earn \$5 real dollars and a \$1 bonus for every 50 play dollars they have at the end of the game. For the online replication at the University of Michigan, participants were mailed a check. For the online replication at the University of Texas - Dallas, participants were mailed a gift card. For the in-person replication at the University of Michigan, participants were paid in cash at the end of the experiment.

Analysis

In the original study, the differences in order quantities between the O and C treatments are estimated by an OLS regression with the average order quantity as the dependent variable and a treatment dummy as independent variable. This is followed by a contrast test. While the original study looked at several treatments, the focus of the replication is only on the earnings difference between the O and the C treatment in Study 1. We conduct this same analysis.

For robustness, we will additionally estimate differences in order quantities using a linear regression with clustered standard errors by subject, with per round order quantity as the dependent variable and a treatment dummy as independent variable.

Differences from Original Study

There are some differences between the original study and the replication. The original experiments were conducted at Duke University. We conduct the experiments at the University of Michigan and University of Texas - Dallas. For the online replication, (i) we conduct the experiment on a computer rather than manually with poker chips and physical dice and (ii) we conduct the experiment remotely, not in-person. For the in-person replication, there are no notable differences from the original experiment.

Replication Results

We first test the results of the online replication. A total of 59 subjects at the University of Michigan and 57 subjects at the University of Texas - Dallas completed the online experiment. Following the original study’s inclusion criteria, also described in our pre-registration, we drop subjects who did not pass the comprehension test. Seven subjects failed each at the University of Michigan and University of Texas - Dallas, making the final sample for the online replication 52 and 50, respectively. In the data files, subjects that fail the comprehension test are indicated (ComprehensionFail=1).

Table 1 reports the original study results (A) versus the online replications at the University of Michigan (B) and University of Texas - Dallas (C), as well as the in-person replication at the University of Michigan (D). The p -value reported for the original study ($p = 1.612 \times 10^{-6}$) is for the analysis comparing only the O and C conditions, as this is the relevant comparison.

We find a significant difference in order quantities for the online replication at the University of Texas - Dallas ($p = 3.602 \times 10^{-3}$), but not at the University of Michigan ($p = 0.0754$). That is, the main result of Chen et al. (2013) replicates online using the University of Texas - Dallas subject pool. The direction of the finding is also supported, with significantly higher order quantities in the O condition (the “own financing” payment scheme) compared to the C condition (the “customer financing” payment scheme). The online replication results at each site are consistent when we conduct the analysis to predict order quantity including clustered standard errors ($p > 0.05$ at the University of Michigan and $p < 0.05$ at the University of Texas - Dallas).

The protocol set by the replication team is to conduct an in-person replication consistent with the original study whenever an online replication fails. Therefore, we additionally conduct an in-person replication at the University of Michigan. We find a significant difference in order quantities for the in-person replication at the University of Michigan ($p = 9.970 \times 10^{-3}$). The result is consistent when including clustered standard errors.

Unplanned Protocol Deviations

There were no unplanned protocol deviations.

Discussion

The result from Chen et al. (2013) that order quantities are higher under an “own financing” payment scheme than under a “customer financing” payment scheme is replicated. At the University of Texas - Dallas, the results replicated using an online, asynchronous version of the original in-person experiment. At the University of Michigan, the results did not replicate with the online version but did successfully replicate in person.

Table 1 Replication Results

A. Original Study			
<i>Payment Scheme</i>	<i>N</i>	<i>Mean Order Q (SD)</i>	$q^O - q^C$
O	29	11.728 (1.392)	1.979 ($p = 1.612 \times 10^{-6}$)
C	21	9.749 (1.058)	
B. University of Michigan (Online)			
<i>Payment Scheme</i>	<i>N</i>	<i>Mean Order Q (SD)</i>	$q^O - q^C$
O	27	10.864 (2.330)	0.899 ($p = 0.0754$)
C	25	9.965 (2.288)	
C. University of Texas - Dallas (Online)			
<i>Payment Scheme</i>	<i>N</i>	<i>Mean Order Q (SD)</i>	$q^O - q^C$
O	25	11.285 (2.577)	1.419 ($p = 3.602 \times 10^{-3}$)
C	25	9.866 (2.224)	
D. University of Michigan (In Person)			
<i>Payment Scheme</i>	<i>N</i>	<i>Mean Order Q (SD)</i>	$q^O - q^C$
O	25	10.962 (3.188)	1.128 ($p = 9.970 \times 10^{-3}$)
C	27	9.834 (1.660)	

References

Chen, Li, A Gürhan Kök, Jordan D Tong. 2013. The effect of payment schemes on inventory decisions: The role of mental accounting. *Management Science* **59**(2) 436–451.