

Author Response to Replication of Ho and Zhang (2008)

Teck-Hua Ho and Juanjuan Zhang

June 11, 2022

We are honored to have been selected as one of ten *Management Science* studies to be replicated. We are grateful to the replication team for being transparent in their efforts and for sharing their data, code, and report (received on April 6, 2022).

Main Results of the Original Paper Were Replicated

Our original study tested alternative pricing-contract solutions to the inefficiency problem in the channel literature. Linear pricing contracts are inefficient because they lead to higher wholesale and retail prices than those that maximize total channel profit.

In standard economic theory, nonlinear pricing contracts offer a solution (Tirole 1988). For example, a two-part tariff (TPT) contract can restore channel efficiency. With a TPT contract, the manufacturer charges a fixed fee in addition to a wholesale price. This allows the manufacturer to lower the wholesale price, which induces the retailer to lower the retail price.

The main results of the original paper are as follows:

1. TPT does not solve the channel inefficiency problem because of loss aversion. Since the retailer is averse to paying a high fixed fee (an upfront loss), the manufacturer is forced to raise the wholesale price above what would otherwise have been offered without loss aversion, in order to maximize profit. The retailer responds by charging a higher retail price, to the detriment of channel efficiency.
2. Framing the TPT as a quantity discount (QD) improves channel efficiency even though the two contracts are economically equivalent. QD improves channel efficiency because it mitigates loss aversion by making it less salient. This allows the manufacturer to charge a higher fixed fee and a lower wholesale price, which induces the retailer to charge a lower retail price.

Both main results from the original paper were replicated (see the table on next page):

1. TPT was inefficient; in fact, its efficiency was even lower than in the original study.
2. Compared with TPT, QD produced a lower wholesale price, a higher fixed fee, and a lower retail price. All these effects were highly significant at the 0.001 level.

Variable	Original Study			Replication Study		
	TPT N = 264	QD N = 242	p-value	TPT N = 528	QD N = 517	p-value
Wholesale price	3.96 (1.17)	3.41 (1.25)	0.000	4.54 (1.45)	4.23 (1.66)	0.001
Fixed fee	5.24 (2.32)	6.95 (4.17)	0.000	4.34 (2.49)	6.02 (7.65)	0.000
Retail price (if accept)	6.86 (0.54)	6.71 (0.80)	0.029	7.06 (1.09)	6.80 (1.06)	0.001
Acceptance (%)	74.24 (43.81)	82.23 (38.30)	0.029	76.14 (42.67)	70.41 (45.69)	0.036
Efficiency (%)	69.51 (41.27)	76.37 (36.18)	0.047	65.18 (21.82)	62.62 (17.56)	0.328

Notes. Values in parentheses are standard deviations.

The only channel decision not replicated was the retailer's acceptance of the manufacturer's offer. The acceptance rate was lower in the QD condition of the replication study. Since the efficiency of a rejected contract was 0%, the overall channel efficiency became statistically indistinguishable between the TPT and QD conditions, contrary to the original study.

Note the following about the retailer acceptance rate:

1. The retailer acceptance rate is secondary to the main results of the original paper. As explained above, what is central to channel efficiency is charging lower wholesale and retail prices. To understand the logic, consider a manufacturer who charges a wholesale price equal to the marginal cost of production. This efficient contract may end up being rejected if the manufacturer leaves the retailer with too little surplus by charging a high fixed fee. As modeled in our original paper, this could happen if the manufacturer cannot perfectly predict the retailer's acceptance decision.
2. In the replication study, the acceptance rate's p -value of 0.036 was much higher than those of the other three channel decisions, which were 0.001 or less. In addition, there were observations in the replication data where the manufacturer's offer was such that the maximum profit the retailer could have earned was negative. These manufacturer-subjects probably did not understand the experimental task. If we were to remove these observations from the replication data, the difference in acceptance rate between the TPT and QD conditions would have been statistically insignificant ($p = 0.174$), whereas the statistical significance of the other three channel decisions would have remained to be 0.001 or less.
3. The replication study reported participant frustration and implemented unplanned protocol deviations that might have affected the acceptance rate and lowered channel efficiency in the QD condition. We address this in the next section.

Participant Frustration and Unplanned Protocol Deviations in the Replication Study

In the replication study, participants had unanticipated difficulty in completing the experimental task, especially in the QD condition. As acknowledged in the “Unplanned Protocol Deviations” section of the replication report, participants were frustrated; those in the TPT condition took two hours to complete the task while those in the QD condition needed substantially longer time.

In an e-mail to us, the replication team shared the following observation:

The reason QD took longer is because of extra calculations required for Price A. This calculation involves $X + Y/(10-P)$ which is often not an integer. To do this calculation correctly also requires understanding order of operations, which to my great surprise it turns out that many people don't understand. So some tried to calculate it as $(X+Y)/(10-P)$. Even after calculating Price A correctly, they often ended up with a fraction and then had to multiply a fraction to calculate profit.

These problems might have affected retailers' acceptance decisions and lowered channel efficiency in the replication study, especially for subjects in the QD condition:

1. The literature has shown that negative emotions affect subjects' decisions of whether to accept a contract offer (e.g., Pillutla and Murnighan 1996).
2. Proposition 1 of the original paper shows that channel efficiency decreases with the complexity of the decision task (page 694). If subjects found the experimental task more complex in the QD condition, this alone could have lowered the QD's condition's efficiency in the replication study.

It is worth mentioning that participants in the original study neither had problems with the calculations nor reported frustration. They understood and completed the task within 1.5 hours in both the TPT and QD conditions.

To mitigate frustration, the replication team implemented an unusual and unplanned protocol deviation in 75% of the experimental sessions. Specifically, the team used built-in software to correct calculation mistakes made by participants. This protocol deviation might have changed the nature of the experimental task.

Finally, the replication study was performed at one location although the plan was to include two separate locations. At this one replication location, the actual number of subjects did not meet the pre-registered goal, reducing the power of the replication study from the pre-registered target of 90% to around 80%.

Conclusion

The main results of the original study were replicated. The replication study confirmed (1) that TPT is inefficient and (2) that QD helps the manufacturer and retailer design more efficient contracts compared with TPT.

Contrary to the original paper, the replication study found a lower acceptance rate in the QD condition, which made the overall efficiency statistically indistinguishable between the TPT and QD conditions. However, acceptance rate was not the primary focus in our original study. Its result in the replication study was less stable than the other channel decisions. Moreover, participants' unanticipated frustration in the QD condition alone might have affected their acceptance rate and lowered channel efficiency. The unplanned protocol deviation to correct participants' calculations could also have altered the nature of the experimental task.

Future research should seek to understand why participants in the replication found it difficult to complete the task in the QD condition even though QD is economically equivalent to TPT. Could the reason be the composition of the subject pool? Was QD computationally more challenging than TPT and, if so, how would the computational challenge differ across subject pools? It would also be interesting to examine whether a better protocol deviation, such as several rounds of practice, would improve participants' understanding of the experimental task and alleviate their frustration.

References

- Ho, T.-H., Zhang, J. 2008. Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science*, 54(4), 686-700.
- Pillutla, M.M., Murnighan, J.K. 1996. Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208-224.
- Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.

Author Response to Replication of Ho and Zhang (2008) Addendum

Teck-Hua Ho and Juanjuan Zhang

February 23, 2023

We thank the replication team for informing us of their second wave of replication, conducted after our original response (dated June 11, 2022, above). We received their revised replication report on February 14, 2023.

In this second wave, the replication team collected additional data at the original replication site (at the University of Texas at Dallas) and added a secondary replication site (at the University of Michigan). We appreciate not only the effort but also the improved quality of data collection in this second wave of replication – there were no reports of subject frustration or confusion about the task, unlike in the first wave of replication.

Unfortunately, at the Michigan site, the target number of observations ($N = 1,348$) was not met, reducing the statistical power of the study from the pre-registered target of 90% to 60%. With this reduced sample at Michigan, efficiency was higher in the quantity discount (QD) condition than in the two-part tariff (TPT) condition ($p = 0.111$), which was in the same direction as our original study.

Would the efficiency result be statistically significant had the Michigan site met the target number of observations? We tested this hypothesis by pooling the data collected in the second wave of replication (396 from UT Dallas and 629 from Michigan). Indeed, efficiency is higher in the QD condition than in the TPT condition at $p = 0.039$, replicating our original finding (see the table below).

Variable	Original Study			Replication Study (Second Wave)		
	TPT $N = 264$	QD $N = 242$	p -value	TPT $N = 508$	QD $N = 517$	p -value
Wholesale price	3.96 (1.17)	3.41 (1.25)	0.000	4.56 (1.42)	4.32 (1.48)	0.009
Fixed fee	5.24 (2.32)	6.95 (4.17)	0.000	4.17 (2.55)	5.18 (2.75)	0.000
Retail price (if accept)	6.86 (0.54)	6.71 (0.80)	0.029	7.16 (1.01)	6.98 (0.92)	0.007
Acceptance (%)	74.24 (43.81)	82.23 (38.30)	0.029	78.15 (41.36)	80.27 (39.83)	0.403
Efficiency (%)	69.51 (41.27)	76.37 (36.18)	0.047	66.33 (39.59)	71.25 (36.77)	0.039

Notes: Values in parentheses are standard deviations.

The replication team chose to pool data by site, combining the data obtained in both waves at UT Dallas. This is problematic since the data collected at UT Dallas in the first and second waves were not comparable (due to subject frustration and confusion in the first wave).

In the second wave of replication, once again the contract terms (i.e., wholesale prices, fixed fees, retail prices) were significantly more efficient in the QD condition than in the TPT condition. It appears that these results are very robust regardless of subject heterogeneity, frustration, and confusion during the study. These were the main results of the original study, as highlighted in our original response (dated June 11, 2022, above).

We remain concerned about the continued use of the unplanned protocol deviation to correct subjects' calculation mistakes. This is problematic from a replication perspective. Hence, the conclusions drawn should be treated with caution.

Moreover, it may be interesting for the replication team to analyze the degree of calculation mistakes. This can be a good proxy for subject understanding of the experimental task. The level of understanding of the task remains a limitation to the generalizability of the replication results.

In summary, with higher-quality data collection, the second wave of replication study replicated our original findings on both the efficiency outcome and its underlying mechanism. These findings suggest that subject frustration and confusion can affect replication results.