INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Inventory Research

ISSN online: 1746-6970 - ISSN print: 1746-6962
https://www.inderscience.com/ijir

Learning under the inventory problem of economic order quantity: a behavioural study

Yan Wu, Kay-Yut Chen, Yan Lang

DOI: 10.1504/IJIR.2023.10054788

Article History:

Received:	18 September 2022
Last revised:	13 January 2023
Accepted:	11 February 2023
Published online:	06 July 2026

Learning under the inventory problem of economic order quantity: a behavioural study

Yan Wu*

San Jose State University,
San Jose, CA 95192, USA
Email: yan.wu@sjsu.edu
*Corresponding author

Kay-Yut Chen

University of Texas at Arlington,
Arlington, Texas 76019, USA
Email: kychen@uta.edu

Yan Lang

State University of New York at Oswego,
Oswego, NY 13126, USA
Email: yan.lang@oswego.edu

Abstract: Inventory management relies on the economic order quantity (EOQ) and newsvendor models. While the newsvendor model has received much attention in the field of behavioural operations management (BOM), the EOQ problem remains largely untouched. This study presents one of the first empirical examinations of inventory decisions under the deterministic EOQ model. The experiments analyse learning behaviours in response to stationary and non-stationary parametric environments. Results show that participants are less likely to repeat suboptimal decisions when parameters remain static. When confronted with cost parameter shocks, most players can improve decisions over time and benefit from past experiences. Two behavioural models, a modified experience-weighted-attraction (EWA) model and an exploratory behaviour-based error reduction model (ERM), are developed to analyse the decision-making process. These models reveal the impact of different behavioural traits on learning under EOQ. Additional experiments are conducted to improve empirical performance.

Keywords: economic order quantity; EOQ; inventory management; behavioural operations management; BOM; learning; reinforcement; probabilistic choice.

Reference to this paper should be made as follows: Wu, Y., Chen, K-Y. and Lang, Y. (2026) 'Learning under the inventory problem of economic order quantity: a behavioural study', *Int. J. Inventory Research*, Vol. 6, No. 5, pp.1–34.

Biographical notes: Yan (Diana) Wu received her PhD in Supply Chain Management and Operations Research from Penn State University. She is currently an Associate Professor at the School of Global Innovation and Leadership, Lucas College and Graduate School of Business. Her research interests are in the areas of supply chain management, behavioural operations management, experimental economics and artificial intelligence.

Kay-Yut Chen is a Professor of the ISOM Department of the University of Texas Arlington. Early in his career, he established behavioural economics research at HP Labs, a first in a corporation, after he received his PhD from Caltech in 1994. He won the Management Science 2014 Best Paper Award in Operations Management, and the 2012 INFORMS Revenue Management and Pricing Practice Award. His work has been featured in *Scientific American* (2006), *Newsweek* (2003), the *Wall Street Journal* (2000), *Financial Times* (2002) and others. He is the author of the book, *The Secrets of the MoneyLab*, published in Oct 2010.

Yan Lang is currently an Assistant Professor of Supply Chain Management at the State University of New York at Oswego. He obtained his PhD in Management Science and M.B.A. from the University of Texas at Arlington. His research interests revolve around several topics including behavioural operations management, inventory management, healthcare operations management, and cybersecurity.

1 Introduction and related literature

The economic order quantity (EOQ) and newsvendor problems are the two fundamental building blocks of inventory control theory. The former one, introduced by Harris (1913), is undoubtedly one of the oldest models in the inventory management literature. Its solution, the well-known 'square root formula', balances the tradeoff between the fixed ordering cost and the variable inventory holding cost. While the basic EOQ model assumes a known and constant demand, there exists a vast analytical literature on inventory and production models that generalises the inventory problem along numerous directions. Choi (2014) provides a comprehensive survey of these theoretical works, and indicates that many models are still widely accepted by industries today for their simplicity and effectiveness. For example, the popular enterprise resource planning (ERP) system developed by Oracle uses the EOQ model as built-in calculations for inventory planning and control.

Since Schweitzer and Cachon (2000), empirical inventory decision making has received much research attention. With the use of lab experiments, many studies seek to identify and explain behavioural departures, such as the 'pull-to-centre' effect, commonly observed under the newsvendor setting. In a typical newsvendor experiment, subjects are asked to repeat their decisions over multiple rounds, in which parameter settings (e.g., price, cost and demand distribution) are usually kept stationary. While the single-period newsvendor model predicts the optimal solution to be a unique one that balances the overage and the underage costs, behaviours of human subjects are often found to vary over time (Wanniarachchi et al., 2021; Truong et al., 2022). To capture the behavioural dynamics, the modelling framework of experience weighted attraction (EWA) is widely employed, where individuals are assumed to reinforce past actions, weighing decisions

chosen in the past more than counterfactual ones (Camerer and Ho, 1998, 1999). For a thorough discussion of the literature on newsvendor behaviours, we refer to the recent handbook of behavioural operations management (Donohue et al., 2018, Chapter 10).

It is worth noting that many of the explanations for newsvendor behaviours, such as mean anchoring, demand chasing or loss aversion, are driven by biases or preferences in making decisions under uncertainty. Due to the variable demand condition, outcomes of newsvendor decisions appear to be random, which makes it challenging for subjects to improve their performance through 'learning-by-doing' (Bolton and Katok, 2008; Bolton et al., 2012). In contrast to such an abundant literature, empirical decision making under the inventory problem of EOQ has not yet been explored much. We are one of the first few such experimental studies to examine learning behaviours under the infinite horizon deterministic EOQ model. We are interested in this particular inventory setting as it offers a nice contrast with the newsvendor literature. With no randomness in demand under the basic EOQ model, behaviours such as demand chasing or risk-aversion can be eliminated; and we can thus explore behavioural issues that are unique to decision making under certainty.

According to interviews conducted by Pan et al. (2022), despite the extensive adoption of ERP modules that recommend ordering decisions based upon EOQ analysis, inventory managers often override the system suggestions (by adjusting module inputs) during new product launches or in the face of fluctuating costs. As a result, inefficiencies are introduced by human biases and judgment errors and the manufacturers interviewed are negatively impacted. It is therefore important to understand the decision-making process under the inventory problem of EOQ, and experiment with relatively simple interventions that may help improve decision-making performance.

The main research focus of the study is how individuals learn and internalise the nonlinear trade-offs between the fixed ordering cost and the variable holding cost while making the inventory decisions of EOQ. We design and conduct a series of human-subject experiments, in which cost shocks are introduced in a way to reflect the constant changes in the real-world environment that an inventory manager has to face over time (even when outcomes from their decisions can be somewhat certain). In addition, we control the type of operational cost, i.e., the ordering cost, the holding cost, or both, that varies over time, generating three treatments in our first study. The manipulation of these cost parameters removes the possibility for subjects to arrive at optimal solutions simply by trial and error without understanding the trade-off to solve the inventory problem of EOQ. Moreover, it allows us to observe how individuals adapt from similar but not identical decision-making scenarios. To provide some benchmarks, we also include a 'static' phase in these experiments where the parameter setting stays unchanged.

We observe that, unlike the 'pull-to-centre' effect in the newsvendor experiments, aggregated orders in our EOQ experiments are skewed to the right – there exists a stronger tendency to over order and decision errors appear to be asymmetric. Moreover, when the parameter setting remains static, the majority of decision makers seldom repeat past decisions that are suboptimal and can converge quickly to the optimal solution. When cost parameters change dynamically, subjects react in a way that is qualitatively consistent with the nonlinear relationship required by the EOQ model; however, their performance improves at a much slower rate over time.

To further understand the underlying decision-making processes, we develop two competing behavioural models. The first is a modified version of the EWA model, which

incorporates behavioural adjustment based on 'similarities' between the past and current parameter settings into the reinforcement process. Note that under the non-stationary experimental setting of our study, reinforcement may not be an effective learning strategy as an individual may reinforce with the 'wrong' past experience. It is still an intuitive heuristic for individuals with cognitive limitations to follow. In parallel, we propose a new behavioural model, referred to as the *error reduction model* (ERM), to capture the exploratory behaviours observed under the static phase. Its core mechanism represents an exploratory process to reduce evaluation errors over time and past experience. The fundamental difference between the two models is that, under the ERM, previously attempted decisions that have been proven to be inferior by the deterministic feedback would have a lower chance (instead of a higher chance) to be chosen. We estimate these two (non-nested) models separately, and find that while the ERM explains the subset of observations from the static phase well, the modified EWA model provides a better overall fitting.

Model estimation results further confirm the positive impact of 'similarity adjustment' on learning, and reveal some interesting relationships between various behavioural traits. Based on these results, we design another set of experiments to investigate interventions to improve empirical performance. Results from our second study speak to the kinds of training and decision support tools that can be included in the development and adoption of ERP systems to further support inventory management decision makers.

To the best of our knowledge, there are three recent studies that are relevant to this research. Stangl and Thonemann (2017) analyse how inventory metrics affect actual decisions under several laboratory settings, and one of which they used is EOQ. They observe that individuals' assessment of performance differs under treatments that indicate inventory levels either by days of supply or by turnover ratios, despite the fact that the two metrics are equivalent. In their experiments, subjects were asked to determine the ordering costs of multiple products with different holding costs. In contrast, we confront decision makers with the direct EOQ decisions in our experiments. Pan et al. (2020) evaluates individuals' cognitive abilities using the cognitive reflection test (CRT), and they find that subjects with higher CRT scores on average perform better under a finite horizon EOQ model with stationary parameters. Pan et al. (2022) further examines the interventions of cognitive stress under a similar EOQ setting. They manipulate cognitive load by restricting when participants can order replenishment and engaging them in a pin memorisation competition. A Markov switching model that estimates the probability of switching to more profitable actions is used to describe observed behaviours in the experiments of Pan et al. (2020, 2022). In our study, each decision round simulates an inventory system over an infinite horizon independently and subjects are incentivised to minimise the long-run operational costs. In other words, we do not permit inventories to be carried over from one round to the next, as inventory decisions in our experiments are controlled to be independent over time. This simpler design enables us to eliminate reaction to leftover inventories, and focus on how individuals evaluate the fundamental nonlinear trading-off between the fixed and the variable costs under the EOQ problem, and how they adapt their behaviour in response to dynamic changes in the cost parameters.

The rest of the paper is organised as follows. In Section 2, we present details about how our first set of experiments are designed and implemented, followed by data analysis and result discussions. The two behavioural models are introduced, estimated and

evaluated in Section 3. In Section 4, we provide a second set of experiments that examine interventions to accelerate learning under EOQ. The study concludes in Section 5.

2 Study 1

2.1 Experimental design and implementation

We design the general experimental setting to represent the deterministic EOQ problem with no lead time. The decision task is framed as managing the inventory of a non-seasonal product over multiple 'weeks'. The product is consumed at a constant rate per week (D). If there is not enough stock to satisfy the next week's demand, a replenishment order (Q, determined by subjects) will be issued so that stockout would never occur in the game. There are two types of operational costs: a fixed ordering cost (K) for placing an order, and a variable holding cost (H) for carrying a unit of the product each week. The following formulas are introduced in the experiment to explain the calculations of the corresponding costs given an order quantity (Q):

$$\textit{Average ordering cost per week} = \frac{D}{Q} \times K$$

$$\textit{Average holding cost per week} = \frac{Q}{2} \times H.$$

The game consists of 50 *rounds*. In each round, we first endow subjects with the same amount of experimental money for managing the product and announce specific game parameters (regarding *D, H* and *K*); we next ask players to decide their orders, informing them that their decisions would be run for hypothetically infinite number of weeks given these parameters; after an order is placed, we provide the decision maker following results to review: the ordering and holding costs averaged over the simulated weeks according to the equations above, and profit for the round computed as the difference between the endowed revenue and the sum of two operational costs. The final cash payment to a participant is based upon her accumulated profits from the entire game.

Note that each of our game rounds is designed to mimic an independent inventory system over an infinite horizon – there is no inventory carryover between rounds. Since the profit in a round is the endowment minus the long-run average cost, generated over an infinite horizon given the ordering decision and parameters of that particular round, an individual's payoff in one round is completely unrelated to that in another round. This allows us to consistently incentivise subjects to minimise the long-run operational costs in the game.

The optimal solution to the EOQ problem is in the form of a square root function:

$$Q^* = \sqrt{2DK / H}.$$

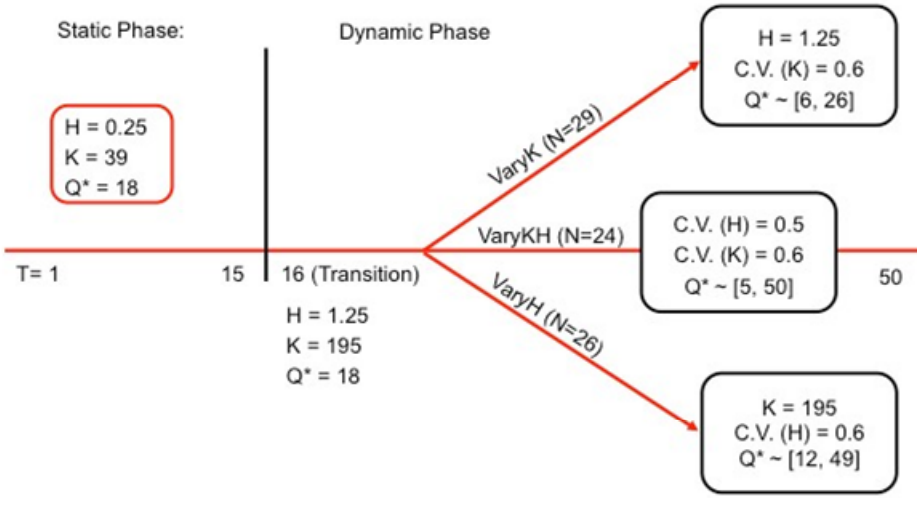
The fixed cost of K and the variable cost of H differ in how they interact with the order quantity to influence the total costs (i.e., nonlinear inverse relationship vs. positive linear relationship). To facilitate our research interests in the decision-making process under EOQ, we divide the 50 rounds into two phases with the demand (D) being normalised at

1 unit per week. It is made clear to all participants that the demand stays unchanged at this value throughout the entire game in the experiment instructions.

In the static phase, which runs from round 1 to 15 in Study 1, both the ordering cost and the holding cost are kept static. In other words, subjects confront the identical inventory problem repeatedly with a single prediction of optimal quantity (Q^*). From round 16 onwards, cost parameters are subject to dynamic change and players need to adapt their decisions accordingly yet still receive deterministic feedback. We require each subject to go through both phases of the game. This *within-subject* design allows us to contrast learning behaviour under different environments at the individual level. In the dynamic phase, we manipulate the type of operational costs to vary over rounds, with the purpose to test how decision makers deal with the trade-off between the fixed and variable costs. To prevent confounding, we control this variation to be *between-subject*, resulting in three treatments in Study 1: *VaryK* in which the ordering cost changes across rounds but the holding cost does not; *VaryH* where only the holding cost varies; and *VaryKH* with both costs being altered every round. Between the two phases, we create a 'transition' period of round 16, where both the ordering and holding costs are rescaled by the same factor but the optimal decision remains the same as that in the static phase.

In parameterising cost shocks in the dynamic phase of the game, since the fixed cost and the variable cost are at different scales, we attempt to control coefficients of variations (CVs) of the respective changing parameters (K and/or H) to be as close as possible for fair comparisons across treatments. In a particular round, the three conditions can be constructed 'in parallel' to generate either the same optimal order quantity with different total costs, or to produce the same lowest total costs given different optimal orders.[1] We opt for the second choice, as our candidate modelling frameworks for learning behaviour operate in the space of objectives rather than decisions. Lastly, to avoid any order effect, changing parameters under each condition are presented according to the same pre-generated sequence of random numbers. Figure 1 illustrates the overall design of Study 1, with corresponding game parameters and theoretical solutions.

Figure 1 Experimental design and parameter settings of Study 1 (see online version for colours)



We conduct all experiments at a large Midwest University. Subjects are undergraduate students recruited from several introductory business courses.[2] They receive course credits for full participation and cash payment for game performance. We make sure that these participants have not been exposed to the topic of EOQ at the time of the research. Subjects, upon arrival to the lab, are randomly seated at workstations separated by blinds. We provide written instructions with detailed examples for them to study. Subjects are encouraged to ask clarification questions but are prohibited from any other communications. The game is implemented using Z-tree and ordering decisions in each round are restricted to be non-negative integers.[3] Players are allowed to use calculators to assist decision making in the experiment. They can also access their own past performance (summarised in a history table) before and after each decision is made.[4] Please refer to Part I in the Appendix for sample instructions and screenshots of the game.

A total of 79 subjects take part in Study 1 (29 in *VaryK*, 26 in *VaryH* and 24 in *VaryKH*, respectively, due to random arrivals of participants). All subjects finished the entire game of 50 rounds, and thus experienced the within-subject factor of the parameter setting (i.e., static vs. dynamic). No subject participated in more than one treatment as required by the between-subject design. Lab sessions lasted for around 60 minutes and the same rate was used to convert game profits to cash payments at the end of an experiment. The average earning per subject is around $13.

2.2 Experimental results

We treat each subject as an independent sample and use non-parametric tests for statistical comparisons. Decision analysis is performed according to the game phases: static versus dynamic setting, with the transition period in between.

2.2.1 Behaviour under the static phase

In the first 15 rounds, subjects observe the same set of parameters repeatedly. We do not expect any significant behavioural difference across experimental sessions since the treatment manipulation does not happen until the dynamic phase. Corresponding statistical comparisons confirm with this assumption.[5] We thus pool observations from all 79 subjects in this phase. Applying the EOQ formula directly to parameters of static phase generates a non-integer solution of 17.66 units. In analysing the ordering decisions, we ignore rounding discrepancies of subjects and consider orders of 17 and 18 to be both optimal.

We first plot the order histograms for the first and the last round of static phase with some descriptive statistics in Figure 2. We can see that a wide range of orders with relatively equal frequency is attempted initially; yet by round 15, the order distribution converges to a sharp peak around Q^*. According to the two-sample Kolmogorov-Smirnov (KS) test, the observed distributional differences are highly significant (p-value < 0.0001). Figure 3 further illustrates how behaviours are developed over time, using order quantities and total costs aggregated by subjects in each round of the phase. A simple OLS regression shows that the total cost decreases significantly over time (p-value = 0.0002, R-square = 0.6750) – so learning clearly occurs under static parameter settings.

Figure 2 Order histograms in round 1 and round 15 (see online version for colours)

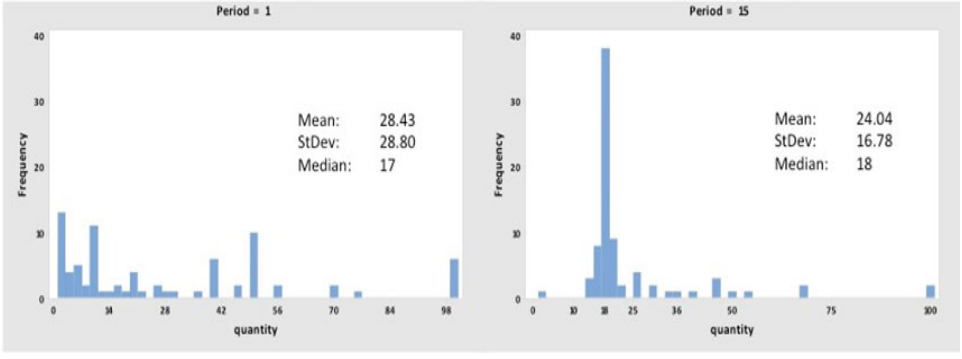
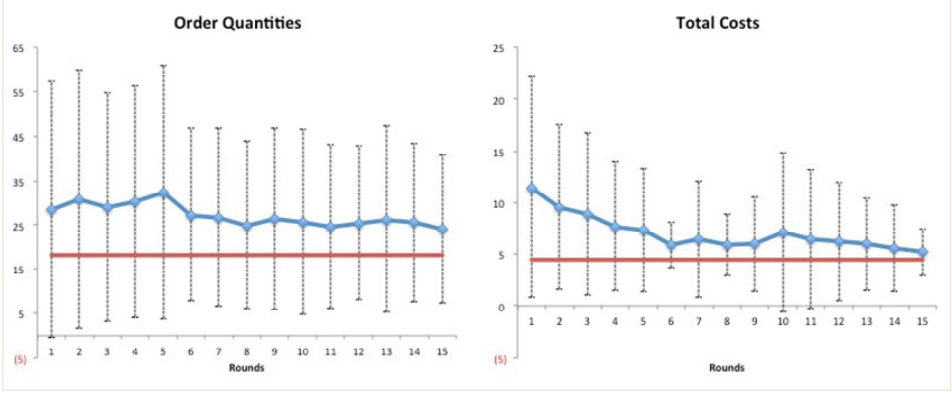
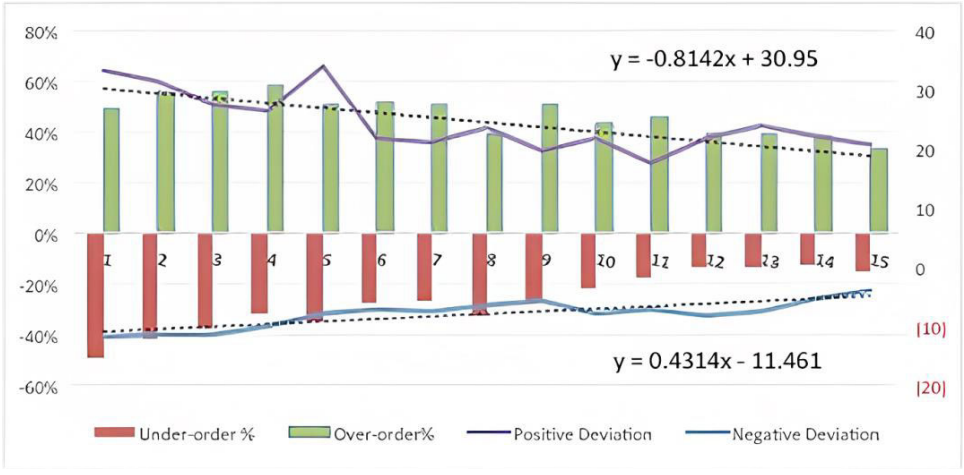


Figure 3 Aggregated ordering and cost behaviour in the static phase of study 1 (see online version for colours)



To further understand how performance is improved, we look into behavioural deviations from the optimal solution in both directions. In particular, we separate over-orders ($Q > Q^*$) and under-orders ($Q < Q^*$) and report their frequencies and magnitudes in Figure 4. It is worth noting that the tendency to err on both sides appears with similar frequencies in the very beginning. However, comparing all rounds in this phase, we find that subjects tend to over order more often (p-value = 0.0001 by the Sign test), and the positive deviations are significantly larger than the negative ones (p-value = 0.0007 by the Wilcoxon signed rank test). Regression analysis indicates that errors on both sides reduce significantly over time (p-value = 0.0019, R-square = 0.5369 for positive ones; p-value = 0.0001, R-square = 0.6902 for negative ones), at rates that are not statistically different from each other (p-value = 0.1001). As a result, the aggregated ordering behaviour appears asymmetric.

Figure 4 Decision deviations in the static phase of study 1 (see online version for colours)



At the individual level, more than 60% of the subjects (51 out of 79) have arrived at close-to optimal solutions ($Q^* \pm 2$) by round 15. Examining their order histories, we find that less than 40% of subjects (29 out of 79) duplicate previously attempted decisions that turn out to be 'inferior' more than once. This is consistent with using a heuristic of trial-and-error to explore the decision space for the optimal solution. In many newsvendor experiments, subjects are also asked to repeat ordering decisions under static parameter settings, but they receive feedback fluctuated by stochastic demands. Results from this literature indicate that decision makers tend to repeat or anchor on suboptimal decisions (Schweitzer and Cachon, 2000); and learning from experience is difficult due to random outcomes (Bolton and Katok, 2008). In contrast, our observations from static phase seem to suggest that the deterministic feedback under EOQ helps subjects 'move away' from suboptimal decisions effectively. However, it is not clear whether such behaviours are due to better assessment of the trade-off between the fixed and variable costs, or merely a result of trial and error in a reduced search space.

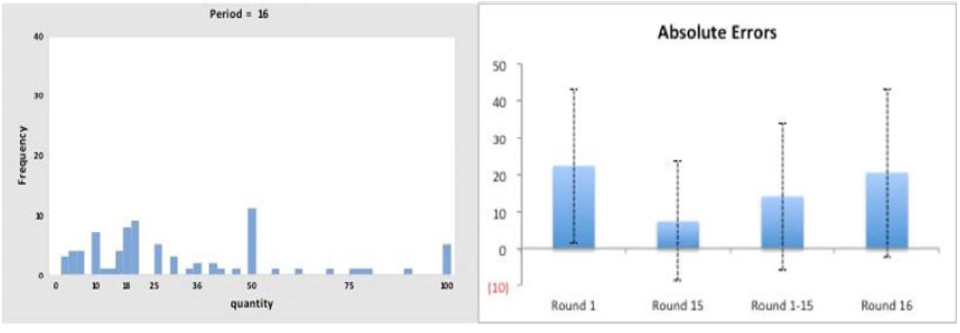
2.2.2 Behaviour in the transition period

We design the transition period to answer the above question. In round 16, both the ordering cost and holding cost are scaled by a factor of 5. As the ratio between the two types of costs remains the same as before, the optimal order should not change. In Figure 5, we plot the order histogram for this transition period, and compare its aggregated absolute errors ($|Q - Q^*|$) with Phase I. The order distribution in round 16 is found to be significantly different from that in round 15 (p-value = 0.0127 by the KS test), but weakly different from that in round 1 (p-value = 0.0512 by the KS test). Similarly, according to the paired Wilcoxon tests, the absolute errors by subjects in this round are significantly larger than those in round 15 (p-value = 0), yet are not statistically different from those in round 1 (p-value =0.1835), or those aggregated from round 1 to 15 (p-value = 0.1348).

It appears that rescaling of the cost parameters 'resets' the decision task for many subjects. Less than 30% of the subjects who have identified (close-to) optimal solutions in round 15 (15 out of 51) are able to duplicate the success in round 16. In response to the

scaled up holding and ordering costs, about 60% of subjects choose to increase their orders (versus 30% who choose to decrease) with greater magnitude. One possible explanation for this observation can be the mental accounting of the two types of operational costs under EOQ. The absolute increase in the fixed cost (from 39 to 195) is a lot larger than that in the variable cost (from 0.25 to 1.25). If decision makers were to overweigh the change in the fixed cost, they would then respond by over ordering. This argument is formally tested and partly supported using a behavioural model. Please see Section 3.4 for details.

Figure 5 Order histogram and absolute errors in the transition period of study 1 (see online version for colours)



2.2.3 Behaviour under the dynamic phase

From round 16, subjects need to respond to variations in cost parameters (of *K, H* or both depending on the treatment). We therefore use OLS regressions to analyse decisions in this phase. And in fitting these models, we take a log-log transformation of the data to help examine the 'square-root law'.[6] In particular, we are interested in

1 whether and how close decision makers follow the square-root law

2 how they react to the dynamic environment.

We present two regression models to address these questions correspondingly:

$$\log(Q_t) = Intercept + \alpha_1 \log(K_t) + \alpha_2 \log(H_t) \tag{1}$$

$$\begin{aligned} \left|\log(Q_t) - \log(Q_t^*)\right| = Intercept &+ \beta_1 \left|\log(K_t) - \log(K_{t-1})\right| \\ &+ \beta_2 \left|\log(H_t) - \log(H_{t-1})\right| \end{aligned} \tag{2}$$

In Model 1, we regress $\log(Q_t)$ of subjects in each round of the dynamic phase against the respective cost parameter(s): $\log(K_t)$ in *VaryK*, $\log(H_t)$ in *VaryH*, and both variables in *VaryKH*. If the decision-maker follows the square root law perfectly, $\alpha_1 = 0.5$ and $\alpha_2 = 0.5$. Thus, we can determine how closely the square root law is followed, by measuring how far the actual coefficients are different from these benchmark numbers. Note that we do not claim whether the subjects are using nonlinear responses since one can mimic, to some extent, though obviously not perfectly, the square root law with linear approximations.[7] In Model 2, the dependent variable measures the deviation from

the optimal prediction, and the independent variables describe the change in parameter settings compared to the last round, in terms of absolute ratios. Theoretical benchmarks and regression estimations for each treatment are shown in Table 1.

The overall models are found to be significant under all treatments (F-test p-values <0.05). Under Model 1, the coefficients of log(K) and log(H) measure the curvature of the EOQ function, which in theory are expected to be 0.5 and -0.5 in the log space, respectively. We observe corresponding estimates of the slopes to be statistically different from zero with the right signs. Subjects consider the operational costs in determining their decisions, at least qualitatively: increase orders when the fixed cost increases and when the variable cost decreases. We also directly compare these estimates with their theoretical benchmarks. The only insignificant result is found for the slope of log(H) under *VaryH* (0.466 vs. –0.5, p-value = 0.322). In the other two treatments, responses to cost variations are flatter than what they are supposed to be. Subjects do encounter difficulties in capturing the square-root function quantitatively. Especially under *VaryKH* with two varying costs, the coefficient of log(H) is no longer close to prediction.

Table 1 Responses to cost parameters in the dynamic phase of study 1

	Model 1			*Model 2*	
	Variables	*Predictions*	*Estimates*	*Variables*	*Estimates*
VaryK	Intercept	0.2350	1.3025***	Intercept	0.2241***
H = 1.25			(0.1265)		(0.0363)
D = 1	log(K_t)	0.5	0.3277***	$\lvert\log(K_t / K_{t-1})\rvert$	0.3058***
			(0.0255)		(0.0612)
	Overall	Adjusted R^2 =	0.1396***	Overall	Adjusted R^2= 0.0228***
VaryH	Intercept	2.9831	3.1466***	Intercept	0.4140***
K = 195			(0.0243)		(0.0423)
D = 1	log(H_t)	–0.5	–0.4660***	$\lvert\log(H_t / H_{t-1})\rvert$	0.1992**
			(0.0344)		(0.0716)
	Overall	Adjusted R^2 =	0.1673***	Overall	Adjusted R^2 = 0.0074**
VaryKH	Intercept	0.3466	2.5325***	Intercept	0.6897***
D = 1			(0.2441)		(0.0366)
	log(K_t)	0.5	0.1111*	$\lvert\log(K_t / K_{t-1})\rvert$	0.1090*
			(0.0482)		(0.0497)
	log(H_t)	-0.5	-0.2967***	$\lvert\log(H_t / H_{t-1})\rvert$	–0.0175
			(0.0483)		(0.0491)
	Overall	Adjusted R^2 =	0.0811***	Overall	Adjusted R^2 = 0.0051*

Notes: Significance codes: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.0001

In Model 2, we intend to examine the influence of dynamics in the parametric environment on subjects' behaviour. More specifically, we are interested in whether subjects can learn to further reduce their errors from past experience with a 'similar' parameter setting. Such similarity is measured by the absolute log ratio difference between cost in the current round and cost in the previous round. This speculation is

confirmed by the positive and significant slope estimates under *VaryK*, *VaryH* and partly under *VaryKH* (for changes in the fixed cost only): bigger the difference in parameter settings, larger the deviation from the optimal order.

2.2.4 Comparison between the two phases

We now look at behaviours developed over the entire game. In particular, we are interested in whether decisions are improved gradually, and if there is any behavioural difference between the two phases. Recall that the optimal order is identical under the static phase but varies across rounds and treatments in the dynamic phase. For a fair comparison, we use the absolute percentage error per round, *Ape* calculated as

$$\left|Q_{tt} - Q_t^*\right| / Q_t^*,$$

to characterise learning behaviour. We consider the following regression model:

$$APE_t = Intercept + \gamma_1 \times t + \gamma_2 \times Phase + \gamma_3 \times t \times Phase \tag{3}$$

There are three independent variables: round *t*, which goes from 1 to 50; *Phase,* a dummy variable that equals 1 if it is the dynamic phase; and an interaction term between the two. We fit Model 3 to each treatment separately, and corresponding outputs are summarised in Table 2.

Table 2 Learning behaviour in study 1

	VaryK	*VaryH*	*VaryKH*	*Pooled*
Intercept	1.2171***	1.2732***	1.3533***	1.2769***
	(0.1122)	(0.1067)	(0.1628)	(0.0745)
t	–0.0653***	–0.0532***	–0.0597***	–0.0596***
	(0.0123)	(0.0117)	(0.0179)	(0.0082)
Phase	–0.1168	0.1300	0.5189*	0.1575
	(0.1637)	(0.1557)	(0.2376)	(0.1087)
t x Phase	0.0508***	0.0324**	0.0399*	0.0414***
	(0.0128)	(0.0122)	(0.0186)	(0.0085)
Adjusted R^2	0.0293	0.0455	0.0303	0.0279

Notes: Significance codes: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001

Model 3 is highly significant under each treatment (F-test p-values < 0.0001). We find that *APEs* decrease significantly over time; yet they reduce at a slower rate in the dynamic phase as the positive interaction terms suggest. These observations are consistent across all conditions. It implies that while decision makers learn to improve their decisions under the EOQ problem in general, it is difficult for them to adapt to the dynamic environment even with deterministic feedback. Furthermore, under *VaryKH*, we notice *APEs* of the dynamic phase to be larger on average. It is probably due to the increased complexity of the decision task as two types of costs vary simultaneously under this treatment. To test for treatment effect, we pool the data from all conditions and run the same regression analysis with a treatment factor and its interactions between *t* and

phase. However, none of such terms turns out to be significant. We therefore leave them out from the final report in Table 2.

To summarise, decision makers – although they tend to over order – recognise the nonlinear trade-offs between the fixed and variable costs under the EOQ problem qualitatively. When the parameter settings are static, they are less likely to repeat suboptimal decisions from the past and can quickly converge to the optimal solution. This finding leads us to question if a different behavioural model than the popular reinforcement learning models would be more appropriate to describe the underlying decision-making process. When cost parameters vary dynamically, subjects respond to these changes in the right directions and are better at adjusting their behaviours to 'similar' settings. Yet they improve performance more slowly over time. While analysis on the lab observations helps identify these behavioural patterns, it cannot explain why and how they occur. In the next section, we resort to the behavioural modelling approach to search for the answers.

3 Behavioural models and estimations

In this section, we employ a behavioural modelling approach to probe deeper into the decision-making process under the inventory problem of EOQ. Results from our experiments indicate significant learning from past experiences despite the non-stationary environment. One established practice to model such behaviours is the experience-weighted attraction (EWA) framework (Camerer and Ho, 1999). It has been adapted by the behavioural operations management literature to explain, for example, the behavioural dynamics under the newsvendor experiments (Bostian et al., 2008; Wu and Chen, 2014). In these studies, subjects are found to reinforce past decisions due to stochastic realisations of demand. In our study, while the demand and decision outcomes are deterministic, subjects face shocks in operational costs under the dynamic phase. We hypothesise that reinforcement may well be a 'comfortable' strategy for many individuals with cognitive limitations, even though it can be a 'wrong' heuristic if the current parametric environment is very different from the past.

The central idea under EWA that past decisions are weighed more than counterfactual ones, on the other hand, seems to contradict our lab observations under the stationary parameter setting. The model predicts that previous, though suboptimal, decisions could have a higher chance to be picked again in the future than the unexplored choices. We observe that more than 60% of our subjects never repeat past decisions that have been proven to be inferior in the static phase. Hence, we developed a new behavioural model to reflect the logic of learning the correct valuations of possible alternatives, and referred to it as the ERM.

Moreover, we note from the experiments that decision errors are influenced by how similar the current parameter setting is to the past. We refer to such behaviour as 'similarity adjustment' and incorporate it into both the EWA and ERM models to attenuate the dynamic decision process. More specifically, we operationalise the idea of 'similarity' with a distance measure $\Delta_{tt'}$, interpreted as the difference between the parameter setting at time t and t'. To compute the distance factor of $\Delta_{tt'}$ under our experimental settings of EOQ, we normalise distances in the parametric space (i.e., the ordering cost K and inventory cost H) as follows:

$$\Delta_{tt'} = \sqrt{\frac{\left(K_t - K_{t'}\right)^2}{\sigma_K} + \frac{\left(H_t - H_{t'}\right)^2}{\sigma_H}}$$

where σ_K and σ_H are the standard deviations of K and H across rounds in the game. They act as normalisation factors so that the two types of costs are evaluated 'within the same range'.[8] The effect of 'similarity adjustment' has not been modelled or discussed in the behavioural operations management literature. Similar ideas can be found in math programming models for selecting workers for tasks of varying complexity (Nembhard and Osothsilp, 2005), and scheduling a set of tasks with some similarity (Nembhard and Bentefouet, 2012).

In the following subsections, we first show how similarity adjustment is specifically incorporated into the standard EWA formulation, followed by its estimation results. Next, we present the ERM formulations, a description of the estimation procedure and corresponding results. The empirical performance of these two models is then compared. Lastly, we discuss the behavioural issue of mental accounting and managerial implications of Study 1.

3.1 The modified EWA model

3.1.1 The probabilistic choice formulation

The standard EWA formulation uses a probabilistic choice framework, under which individuals make random errors when they evaluate potential decisions, and each possible decision q_i receives an 'attraction' $w_t(q_i)$ that sets its probability of being selected. Similar to Bostian et al. (2008), and Wu and Chen (2014), we model the probability of choosing q_i at time t by

$$\frac{\exp\left(\gamma w_t\left(q_i\right)\right)}{\sum_{q \in Q} \exp\left(\gamma w_t(q)\right)},$$

where γ can be interpreted as the subject's tendency to make random errors. At $\gamma = 0$, the agent is completely random in choosing her decision among all possible choices, resulting in a uniform decision distribution (Chen et al., 2022; Lang et al., 2022).

When γ approaches ∞, the agent always selects the choice with the highest attraction according to the EWA procedures specified below. The attraction $w_t(q_i)$ for a particular choice q_i at round t is determined by the following recursive formulation from $t' = 0$ to t.

$$w_t'\left(q_i\right) = \frac{1-\rho}{1-\rho^{t'+1}} \pi^*\left(q_i, K_{t'}, H_{t'}\right)\left(1+\gamma_\Delta \Delta_{tt'}\right) + \rho \frac{1-\rho^{t'}}{1-\rho^{t'+1}} w_{t'-1}\left(q_i\right)$$

$0 \leq \rho \leq 1$ is the memory parameter. $\rho = 0$ means that a subject only considers what happened most recently (in the prior round) in estimating the value function, indicating a strong recency bias. As the memory parameter approaches 1, outcomes in all past rounds are considered equally.

Similarity adjustment

Typical EWA formulation does not usually call out the difference between t and t'. However, we propose to modify the first term of the standard EWA formulation with the factor $(1 + \gamma_\Delta \Delta_{tt'})$ to capture any behavioural adjustment due to similarities in the parametric environment. As shown earlier, $\Delta_{tt'}$ is a distance measure in the parameter space. We refer to γ_Δ as the *similarity adjustment* parameter and expect it to be negative, as individuals are more likely to underweight experiences from less similar environment, as opposed to a similar one. $\gamma_\Delta = 0$ represents the case where an individual does not respond to any difference between the current setting and past ones.

Reinforcement

In line with previous behavioural studies (e.g., Kalkanci et al., 2011, 2014), we consider reinforcement on the focal point of the most recent decision q_t versus all other 'counterfactual' alternatives. To operationalise this idea, we assign a different weight to the counterfactual alternatives and will test for possible underestimation by subjects. Thus, we formulate the utility function as:

$$\pi^*(q_i, K_{t'}, H_{t'}) \begin{cases} \pi(q_i, K_{t'}, H_{t'}) & \text{if} \quad q_j = q_t \\ r\pi(q_i, K_{t'}, H_{t'}) & \quad\quad\; q_j \neq q_t \end{cases}$$

The weight assigned to q_t is normalised to 1. When $r = 1$, an individual pays equal attention to the valuations of all choices. On the other hand, if r is 0, the individual only responds to the previous decisions and ignores all other alternatives, reflecting strong reinforcement bias. Intuitively, we expect $1 > \sigma \geq 0$: an agent weighs the counterfactual decisions less than the decision already made. $\pi(q_i, K_t', H_t')$ describes a value function for payoffs that would have been generated in case q_i were chosen given the ordering cost K_t and H_t in period t. $\pi(q_i, K_t', H_t')$ is calculated as the endowment R minus the average operational costs of the round. Hence,

$$\pi(q_i, K_{t'}, H_{t'}) = R - (K_{t'}/q_i + H_{t'} q_i/2).$$

Note that since (K_t, H_t) can be different from (K_t', H_t'), it is possible for individuals to reinforce with the 'wrong' input from past decisions under this model.

3.1.2 Estimations from the modified EWA model

The maximum likelihood method is used to estimate the modified EWA model for each individual subject based on observations from the entire game (50 rounds).[9] There are four behavioural parameters: random error (γ), reinforcement (r), memory (ρ) and similarity adjustment (γ_Δ). We do not expect all these behavioural effects to be relevant to everyone. Therefore, for each individual, we estimate the model and apply the likelihood ratio test to determine which behavioural factors nested within the model are statistically significant. An effect is included in the final model for a subject only if it is significant at the 5% level.

We first test the null hypotheses regarding similarity adjustment (H0: $\gamma_\Delta = 0$) and reinforcement (H0: $r = 1$, i.e., counterfactual decisions are underweighted). The following figure summarises by treatments, the percentage of subjects with these two traits being significant. Both types of behavioural tendencies are found to be prevalent for the majority of subjects across the board, and there is no treatment difference in the respective proportions at the 5% level.

Figure 6 Behavioural segmentation of subjects from the modified EWA model (see online version for colours)

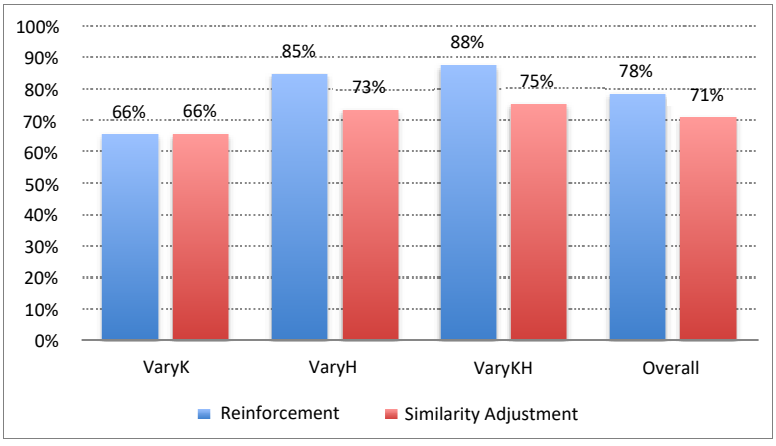


Table 3 Summary of behavioural estimates from the modified EWA model

		VaryK	*VaryH*	*VaryKH*	*Pooled*
Random error	Mean	1.27	0.68	0.34	0.79
(γ)	Median	0.83	0.34	0.25	0.34
	StDev	1.33	0.87	0.28	1.03
Reinforcement	Mean	0.78	0.61	0.56	0.64
(r)	Median	0.95	0.73	0.58	0.74
	StDev	0.28	0.38	0.27	0.32
Memory	Mean	0.71	0.81	0.89	0.80
(ρ)	Median	0.87	0.88	0.91	0.90
	StDev	0.36	0.22	0.07	0.26
Similarity Adjustment	Mean	–0.22	–0.28	–0.04	–0.18
(γ_Δ)	Median	–0.41	–0.41	–0.36	–0.40
	StDev	0.49	0.65	1.00	0.72

Table 3 summarises the statistics of all behavioural estimates conditioned on their significance. Note that estimates of similarity adjustment are negative for an overwhelming majority (more than 83%) of the individuals. This is consistent with the intuition that less attention is placed on past periods with parameter settings that are less similar (i.e., higher distance factor). Next, we compare estimates of these behavioural parameters across treatments using Wilcoxon tests. First, γ is significantly higher under *VaryK* than *VaryKH* (p-value < 0.01), which implies lower level of random noise in

observed decisions. Second, the reinforcement parameter (r) under *VaryK* is significantly higher (i.e., weaker tendency to reinforce) compared to the other two treatments (p-value = 0.0353 vs. *VaryH*, and p-value< 0.01 vs. *VaryKH*). As for similarity adjustment (γ_Δ), it is higher under *VaryH* than *VaryKH* (p-value = 0.0392). Lastly, there is no statistical difference in the memory parameter (ρ) across treatments.

We also check correlations among all four behavioural traits using data pooled from all experiments. The only significantly correlated pair of parameters we find is the random error and the reinforcement parameters (correlation coefficient = 0.59, p-value = 0.0000). This suggests that individuals who can evaluate their options with lower random noise (i.e., higher γ) do not reinforce as much (i.e., higher r). This result is, in spirit, in line with the correlation found between decision noises and anchoring in Wu and Chen (2014). It is also interesting to note that similarity adjustment (γ_Δ) is independent of all other behavioural factors including reinforcement.

The above estimation results help reveal more information regarding behaviours at the individual level as well as some treatment effects that experimental analysis cannot. It seems that *VaryKH* is the most challenging condition (probably due to the two simultaneously varying costs) for learning, as reflected by more decision noises (i.e., lower γ) and weaker responses to similarities in the parametric environment (i.e., lower γ_Δ). Subjects also appear to reinforce less when it is the fixed cost, instead of the variable cost, that is changing.

3.2 The ERM

The ERM, in contrast, is motivated by the exploratory behaviour observed under the static phase. We propose to capture the tendency to avoid repeating suboptimal decisions by a learning process that reduces evaluation errors over potential choices. More specifically, we consider that the learning rate may vary along two dimensions: past experience with the environment (i.e., how different is the current parameter setting compared to the previous ones), and time lapse (i.e., how far in the past when a previous decision is made).

3.2.1 The probit choice formulation

Similar to the EWA model, a probabilistic choice framework is used here. Most applications in behavioural operations, including the EWA model [another example is Su (2008)], assume that decisions follow multinomial logit distributions and random errors follow Gumbel distributions. However, the logit choice framework may not be suitable for the ERM due to its more restrictive assumption on the evaluation errors, which are i.i.d. across choices. We adopt the normal distribution to model decision errors as it generates less convergence issues and more consistent estimates, and hence turn to the probit choice framework (Hausman and Wise 1978). This framework has been successfully applied to multiple areas such as voting research (e.g., Dow and Endersby 2004), and marketing (e.g., Kim et al., 2017). The logit and probit formulations are similar in structure, while the latter requires more computational power as no closed-form solution is available. We estimate the model by the maximum simulated likelihood method, which is further described in the next subsection.

Let the evaluation of the decision $i \in \{1 \ldots N\}$, at time period $t \in \{1 \ldots T\}$, be $u_{it} = V_i + \varepsilon_{it}$, where V_i is the true evaluation of decision i. Note that V is only indexed by i

not t, since the true evaluation does not change with time. According to the probabilistic choice model, the decision maker chooses a decision i at time t with the highest u_{it}. That is, decision at time t:

$$q_t = \arg\max_i \{u_{it} : i = 1...N\}$$

We assume $\varepsilon_{it} \sim$ normal(0, σ_{it}) and that they are independent across i and t. Hence, all the dynamics of learning are captured in the evolution of σ_{it} over time.

Base learning factor

We assume that $\sigma_{it} = \sigma_0 e^{-z_{it}}$. This assumption is merely a change of variable for convenience and ease of interpretation. That is, instead of working in the σ_{it} space, we choose to frame learning in the space of $\log(\sigma_{it} / \sigma_0)$. In this framework, σ_0 can be interpreted as a *base error*, and z_{it} as a *learning factor*. That is, individuals start with a base level of error when they start the decision process. When learning occurs, the learning factor z_{it} increases. There is a diminishing return implicit to the assumption of the formulation. That is, the individual will only learn the true valuation of the decision i when $z_{it} \to \infty$.

Base learning

In this simple case, an individual improves her evaluations over all possible decisions equally in a steady rate:

$$z_{it} = \sum_{t'=1}^{t-1} \alpha$$

α can be interpreted as a 'base' level of learning with diminishing impacts to the actual errors. That is, the evaluation errors, of all decisions, reduce at the same relative rate over time. While we can simply write the expression above as $(t-1)\alpha$. We write it this way so that we can add features to the model in a convenient manner. In principle, this part of the model is similar to the time dependent logit-based quantal response equilibrium in Chen et al. (2012).

Similarity adjustment

As discussed before, the similarity adjustment behaviour is incorporated into the ERM as well. In particular, with the identical distance measure $\Delta_{tt'}$ used in the modified EWA model, it is formulated as:

$$z_{it} = \sum_{t'=1}^{t-1} (\alpha + \gamma_\Delta \Delta_{tt'})$$

γ_Δ is again referred to as the *similarity adjustment* parameter. We do not claim the above mathematical formulation to be the only one that captures the effect of dissimilarity in environment on learning. In fact, any decreasing function (for example, $\alpha e^{\gamma_\Delta \Delta_{tt'}}$ with $\gamma_\Delta < 0$) in $\Delta_{tt'}$ will also be applicable. We have attempted several formulations and found the above linear one to be appropriate for our data.

Temporal Decay

We expect the learning effect to diminish over time. That is, the impact of the last period on learning is higher than that of the period before, and so on. Hence, we further modify the formulation of the learning factor to be:

$$z_{it} = \sum_{t'=1}^{t-1} \left(\alpha + \gamma_{\Delta} \Delta_{tt'}\right) e^{-\gamma_t (t-t')}$$

where the parameter γ_t measures the decay of the learning effect over time.

3.2.2 Numerical estimation methodology

The numerical methods to estimate the ERM model outlined above merit some discussion. Since decision errors follow multivariate normal distributions, there exists no close-form solution to calculate the likelihood precisely. Hence, we turn to the maximum simulated likelihood estimation (Lee, 1992) with Monte Carlo integration to evaluate the likelihood function. Hajivassiliou et al. (1996) review Monte Carlo methods that are used to evaluate multivariate normal probabilities. While they focus on evaluating multivariate normal probabilities within rectangular constraints, we adapt their methods to help evaluate discrete choice probabilities. We start with the crude frequency simulator and then apply kernel smoothing to obtain a continuous simulated likelihood function. Next, importance sampling is incorporated to improve the efficiency of the simulator.

Consider a vector of utilities $\{u_i, 1 \ldots N\}$ for N possible decision choices. Let the errors $\{\varepsilon_i, 1 \ldots N\}$ be a set of independent normal random variables, with standard deviations $\{\alpha_i, 1 \ldots N\}$ as shown in 4.1. We drop the index t here because these errors are assumed to be independent over time, and hence we can evaluate the choice probabilities for each period, independent of other periods. The probability of choosing choice i is given by:

$$P(i) = P\left(u_i + \varepsilon_i \geq u_j + \varepsilon_j : \forall j\right)$$

We need to approximate $P(i)$, for all i, by drawing K times of ε_i. Given a set of draws $\{s_{ik}, i = 1 \ldots N, K = 1 \ldots K\}$ from the standard normal distribution, $u_i + s_{ik}\sigma_i$ will have the distribution of $u_i + \varepsilon_i$. Let $y_{ijk} = 1$ if $u_i + s_{ik}\sigma_i \geq u_j + \varepsilon_i$; and 0 otherwise, then the crude frequency simulator of $P(i)$ is expressed as:

$$P(i) = \frac{1}{K} \sum\nolimits_K \prod\nolimits_j y_{ijk}$$

Note that y_{ijk} equals to 1 when $i = j$ by definition, for all i and k. So intuitively, we can include i in the product. $\prod_j y_{ijk} = 1$ if and only if $u_i + s_{ik}\sigma_i \geq u_j + s_{jk}\sigma_j$ for all j. Hence, it can be interpreted as the indicator of the i^{th} option being the best. The rest is simply an average over K Monte Carlo simulations. Also note that this formulation is discontinuous in σ_i because y_{ijk} is not continuous in σ_i. As a result, the behavioural model (which is indeed a model of σ_i) is not continuous in $P(i)$. This poses a technical problem in the estimation of behavioural parameters. We solve this problem by using a *kernel smoothing* function, replacing y_{ijk} (Hajivassiliou et al., 1996). Specifically, we use the kernel function:

$$g(x \mid h) = \frac{1}{1+e^{-x/h}}$$

where h is a bandwidth parameter to control how fast $K(x \mid h)$ change from 0 to 1. y_{ijk} is then approximated by:

$$y_{ijk} = g\left(u_i + s_{ik}\sigma_i - \left[u_j + s_{jk}\sigma_j\right] \middle| h\right)$$

Another issue is the efficiency of the Monte Carlo method. We employ *importance sampling* (Hajivassiliou et al., 1996) to help reduce variance of the estimation results. Namely, a normal distribution with a shifted mean is used as the importance sampling distribution. And we find this method improves the efficiency of the Monte Carlo method substantially. The current sampling scheme is equivalent to sample choice i's utility with a normal distribution with mean u_i and standard deviation σ_i. A shift of x in the utility is equivalent to the use of an importance sampling normal distribution with mean $u_i + x$, and standard deviation σ_i. Therefore, the final probability calculated from this Monte Carlo simulation is:

$$P(i) = \frac{1}{K}\sum\nolimits_K \frac{\phi(u_i, \sigma_i)}{\phi(u_i + x, \sigma_i)} \prod\nolimits_j g\left(u_i + s_{ik}\sigma_i + x - \left[u_j + s_{jk}\sigma j\right] \middle| h\right)$$

where $\phi()$ is the normal density function.

In the final analysis, we use 5000 samples (K) in the Monte Carlo simulation and a kernel bandwidth (h) of 0.1. Testing shows that the standard deviation of the log likelihood estimates is generally less than half a percent of the mean, and that varying the bandwidth parameter would not change our main conclusions.

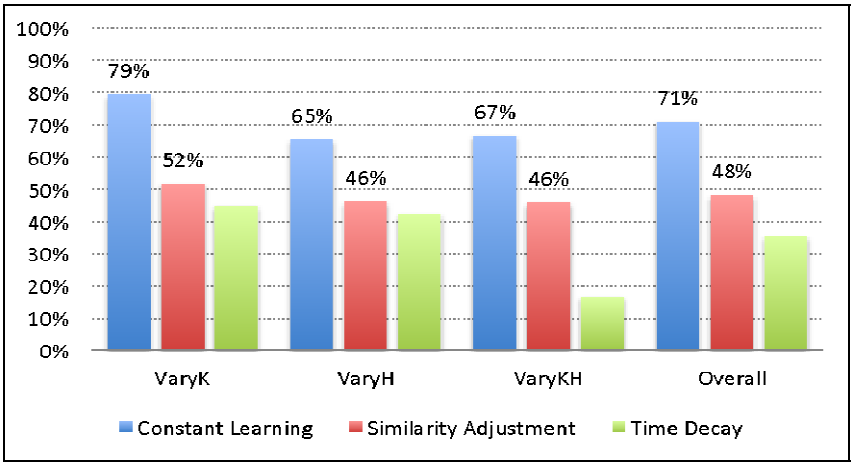
3.2.3 Estimations from the ERM

The maximum simulated likelihood method described above is used to estimate the behavioural model for each subject based on observations from the entire game. Under the ERM, there are four behavioural parameters: base error (σ_0), base learning rate (α), similarity adjustment (γ_Δ), and temporal decay (γ_t). Again, an effect is included in the final model for a subject only if it is significant at the 5% level. Figure 7 summarises the number of subjects with significant base learning rate ($\alpha > 0$), similarity adjustments ($\gamma_\Delta \neq 0$) and temporal decay ($\gamma_t \neq 0$). Note that, for a subject, similarity adjustment and temporal decay can be significant only if the base learning rate is significant.

Out of 79 subjects, we find that 47 of them have significant positive base learning rates (α). This is evidence that a majority of individuals can reduce errors in their evaluations of possible choices over time, even when the parameter setting is dynamic. Nine subjects are found to have significant negative learning rates, which imply that their decision errors increase as opposed to decrease over time, probably due to confusion. The remaining 23 subjects out of the population (about 29%) do not show any measurable learning behaviour, i.e., none of the behavioural parameters are significant. This observation can be explained using a probabilistic choice model with stationary errors over time. It is also possible that this group of subjects exhibits some small amount of learning (or confusion) behaviour, but we do not have enough statistics to identify the effects. It is not surprising that a small portion of the population cannot learn from past

experience or even get more confused when dealing with dynamics in the environment. Among subjects whose base learning rate is positive and significant, 81% of them (38 out of 47) also demonstrate significant similarity adjustment behaviour. This observation confirms our intuition that the learning rate is influenced by differences in the parameter settings: people can learn more effectively from a similar decision-making environment experienced in the past (i.e., a lower parameter distance measured in our study). In addition, we observe 28 out of these 47 subjects to have significant time decay parameters – they would place more weight and learn more from recent experience. Recall that we made historical information about past decisions available in all experiments, so memory limitations may have already been mitigated.

Figure 7 Behavioural segmentation of subjects from the ERM model (see online version for colours)



Across treatments, we find that the base error (σ_0) is significantly lower in the VaryK treatment compared to the other two treatments (p-values < 0.01 by the Wilcoxon test). This is in line with the random error parameter (γ) comparisons under the modified EWA model. We also detect higher base learning rates (α)in VaryK than in VaryKH (p-value = 0.0347). All other comparisons are not significant. For the relationships between behavioural parameters, we discover two significant correlations: the base learning rate is negatively correlated with the base error (p-value = 0.0067), which is not surprising in the sense that a poor start gives one more room to improve. More importantly, the learning rate is also negatively correlated with similarity adjustment (p-value < 0.0001). Hence, an individual who adapts faster in reducing evaluation errors is also more sensitive to changes in the parameter setting. Note that estimates of similarity adjustment parameters are almost all negative (except one subject), so a larger negative value indicates a bigger reduction in the learning rate for the same level of similarity. It is interesting to note that, under both types of behavioural models, there exist correlations between some behavioural factors. If we interpret lower base error, higher learning rate, as well as higher similarity adjustment as 'rational' behaviours, then the above results are in line with the intuition that a more 'rational' decision maker would exhibit rationalities in multiple traits. Table 4 summarises the statistics of these behavioural estimates, conditioned on their significance.

Table 4 Summary of behavioural estimates from the ERM

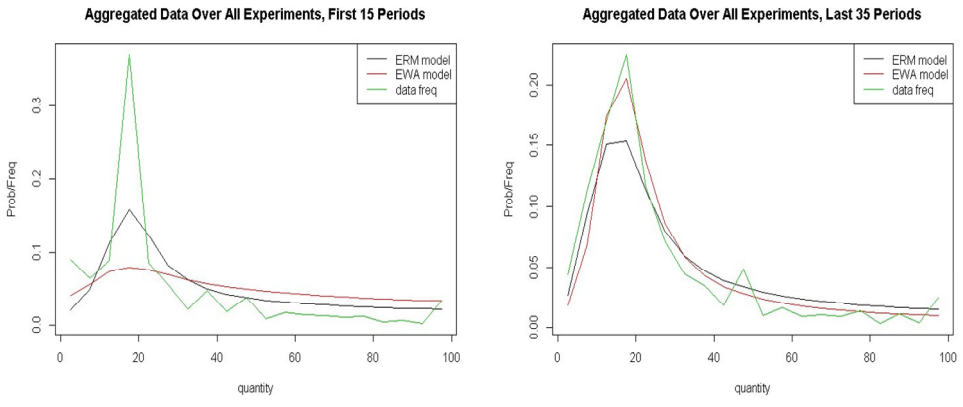
		VaryK	*VaryH*	*VaryKH*	*Pooled*
Base error	Mean	11.43	25.79	44.45	25.29
(σ_0)	Median	3.19	18.75	43.12	15.75
	StDev	14.48	24.37	32.83	27.26
Base learning	Mean	0.10	0.01	–0.02	0.04
(α)	Median	0.12	0.15	0.06	0.12
	StDev	0.33	0.76	0.41	0.51
Similarity adjustment	Mean	–0.25	–0.68	–0.12	–0.35
(γ_Δ)	Median	–0.10	–0.12	–0.07	–0.09
	StDev	0.47	1.14	0.19	0.73
Temporal Decay	Mean	0.23	0.38	0.19	0.28
(γ_t)	Median	0.11	0.16	0.12	0.11
	StDev	0.29	0.39	0.17	0.32

3.3 Model comparisons

Since the two behavioural models we propose represent completely different learning processes, they cannot be nested in one general model. We thus apply the Akaike Information Criterion (AIC) for model selection.[10] Recall that each model is estimated at the individual level using data from the entire game. The EWA model is found to be preferred for 78% of the subjects. Because individuals are independent, another way to compare is to aggregate the AICs over the population to arrive at a final score for each model. In this case, the EWA model (AIC = 29,009) also outperforms ERM (AIC = 31,508). However, if we only focus on *inferior* choices by excluding the optimal decision (18) and the near-optimal decisions (16, 17, 19 and 20), in the static phase (15 rounds) of the experiments, the ERM is better for 56% of the subjects. The ERM is not necessarily better than the modified EWA model in explaining the outcome of individuals' learning behaviour, but it seems to explain the process of how they explore the choices under a static parameter setting better. To illustrate this point, we provide the fittings of each model under the static and dynamic phases separately in Figure 8.

Lastly, we compare both of our models with the static logit-based probabilistic choice model with only one behavioural parameter for the random errors.[11] We find that the ERM and the modified EWA model are both preferred, by the AIC criterion, to the probabilistic choice model for 84% and 62% of the subjects, respectively. The aggregate AIC of the EWA and ERM models are also lower (29,009 and 31,508, accordingly) than that of the probabilistic choice model (31,960). This result suggests that some type of systematic behaviour needs to be integrated in order to interpret the empirical EOQ decisions. Details of the estimations from the probabilistic choice model are available upon request.

Figure 8 Model fitting comparisons (see online version for colours)



3.4 Mental accounting

Recall that more subjects choose to increase their orders rather than to decrease when both the fixed and the variable costs are scaled up by the same factor. Suppose an individual 'mentally' weighs such an increase in the ordering cost more than that in the holding cost, she would then place a higher order quantity even though the optimal solution should remain unchanged.

To determine if the above explanation is true, we add one more parameter that allows overweighting of the fixed cost in the modified EWA model since it is preferred to the ERM model. Namely, we replace the actual ordering cost K with $(1 + w_K)K$, where w_K (restricted to be positive) is the additional weight assigned to the fixed cost in formulating the utility function. When $w_K = 0$, the correct K is used for evaluation. We estimate this parameter (along with the four original behavioural parameters) at the individual level using data from the entire game. Results from the likelihood ratio test for the null hypothesis of $w_K = 0$ show that only 5% of our sample (4 out of 79 subjects) overweigh the fixed cost significantly.

A similar analysis is performed using data from rounds 15 and 16 alone, by the probabilistic choice model. Since there are only two observations per subject, it is not appropriate to apply the modified EWA model (which is dynamic), and we have to pool data across subjects in this analysis. Under this case, w_K is found to be significant (p-value < 0.01) at the aggregate level. Thus, we conclude that although mental accounting may explain behaviour during the transition period, it does not appear to be very general across individuals, and is thus omitted from the final presentation of the behavioural models.

3.5 Discussions of Study 1

Results from Study 1 suggest that there are two different underlying decision-making processes: an exploratory learning one (captured in ERM) versus a reinforcement process (modelled by EWA). We would like to emphasise the distinction between them. Under the dynamic environment, reinforcing with inputs from inappropriate parameter settings in the past can impede rather than improve decision performance. Consequently,

reinforcement should not be viewed as a 'learning' strategy but rather a decision bias. This point can be further verified by a significant positive correlation, between the base learning rate (α in ERM) and the reinforcement parameter (r in the modified EWA), we discover (p-value = 0.0001). Since a higher r implies a lower tendency to reinforce, a subject who shows stronger ability to enhance her evaluation of decisions, for instance, through trial and error, would then reinforce less; and vice versa, a subject who tends to reinforce more would then be weaker in her exploratory learning ability.

From a managerial perspective, results from Study 1 suggest two directions to strengthen the positive effect of similarity adjustment on learning. First, exposing decision makers to a diverse set of decision scenarios through training, so that in the future, they have a higher chance to encounter something that is similar to the ones they have practiced before and benefit from such experience. Second, making past similar scenarios more salient to decision makers. One possible implementation is a decision support tool to pick out past or simulated scenarios that are close to the current one, and call managers' attention to the related decisions and outcomes. In the next section, we report another series of experiments to explore whether such approaches will result in any performance improvement empirically.

4 Study 2

Results from Study 1 show that subjects did pay attention to past experience in the game but did not necessarily benefit from it. In this section, we intend to experiment, under the EOQ problem, different strategies to accelerate learning. In particular, we propose two approaches to help 'train' the behaviour of similarity adjustment:

1 offer opportunities for subjects to experience a broader set of scenarios, in the hope that they can recognise the appropriate ones to adjust to

2 provide some decision supports that make similar scenarios from the past more salient.

We design new sets of experiments to examine each suggestion separately. To control for potential confounding effects, we focus on treatments where only one type of operational costs can vary in Study 2. In other words, *VaryK* and *VaryH* from Study 1 are treated as baselines.

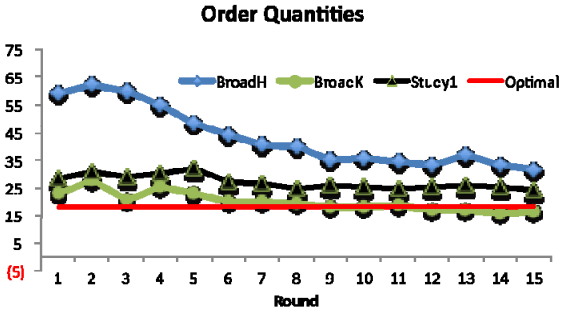
4.1 'Broad training' treatments

Recall that in Study 1, subjects begin with the static phase before experiencing any dynamic changes in the cost parameter. In this set of new experiments, we reverse the game sequence: subjects play in the dynamic phase (35 rounds) first and then move on to the static phase (15 rounds). Since the dynamic phase covers a wide range of cost parameters, it can be used to train the subjects. To test our conjecture that 'broad training' helps improve learning, we compare decisions in the static phase between the new experiments and Study 1, where no prior training is available.

We keep the experimental procedures and game parameters exactly the same as those in Study 1. The two new treatments conducted are coded as *BroadK* with a total of 22 subjects, and *BroadH* with a sample size of 20. We graph the average order quantities

over the 15 rounds of static phase for *BroadK*, *BroadH* and Study 1 in Figure 9 with some summary statistics.[12]

Figure 9 Aggregated ordering behaviour under 'broad training' treatments (see online version for colours)



Treatment	Statistics	First Round	Last Round
BroadK	Median	17.00	17.50
(n=22)	Average	22.91	16.50
	StDev	(21.55)	(3.64)
BroadH	Median	65.00	18.00
(n=20)	Average	59.25	32.50
	StDev	(33.49)	(28.08)
Study 1	Median	17.00	18.00
(n=79)	Average	28.43	24.04
	StDev	(28.80)	(16.78)

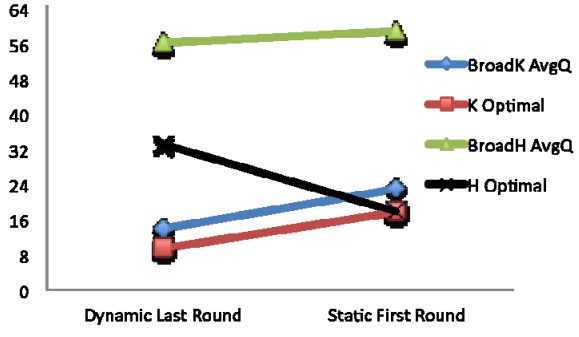
Regression analysis is performed to compare learning behaviours under these conditions over time. We apply Model 3 (with APE as dependent variable) to only the Static Phase and add a dummy variable for treatment. Results show a significant treatment effect and *BroadK* exhibits smaller errors, compared to study 1. This is evidence that broad training is successful in reducing decision errors in the *VaryK* condition. However, the learning rate is not statistically different between *BroadK* and Study 1. Comparing *BroadH* to Study 1 and *BroadK*, the former has a significantly higher learning rate, yet it is probably related to its apparently higher errors at the beginning. Details of the regression analysis can be found in the Appendix. Wilcoxon tests further reveal that, in the last round of the phase, the median order in *BroadK* is significantly lower than that of Study 1 (p-value = 0.039) and insignificant from the optimal prediction (p-value = 0.205 by the 1-sample test); and the gap between *BroadH* and Study 1 also becomes insignificant. Overall, there is mixed evidence about the intervention of 'broad training' on learning. It works in *BroadK* but seems to backfire in *BroadH*. The effect thus does not appear to be universal but likely to be context dependent.

To understand the large initial difference between *BroadK* and *BroadH*, we look at what happened previously - in the last round of the dynamic phase. Figure 10 plots the aggregated orders against the optimal solutions for *BroadK* and *BroadH*, accordingly. First, we observe that the tendency to over order persists in these new treatments. During the transition from the dynamic phase to the static one, subjects in *BroadK* experience a 29% decrease in the fixed cost and an 80% decrease in the variable cost. As a result, the optimal order should be almost doubled. We do find the empirical decisions to increase significantly as predicted (p-value = 0.0249 by the paired Wilcoxon test). As for *BroadH*, subjects encounter an 80% decrease in K and a 29% decrease in H with the optimal order to be roughly cut by half. Despite the parameter changes, there is no difference in the subjects' orders (p-value = 0.4687 by the paired Wilcoxon test).

We believe the reinforcement bias discovered in Study 1 may hold account for these observed differences. Recall that estimates from the modified EWA model indicate that the tendency to reinforce is significantly stronger when the holding cost varies (in *VaryH*). This implies that subjects in *BroadH*, when confronted with the Static Phase, are likely reinforcing with the wrong inputs – those from the dynamic phase. In contrast, subjects in *BroadK*, due to less reinforcement, may benefit from these past experiences.

In other words, there exist some pulling forces, as suggested by the significant correlation between similarity adjustment and reinforcement behaviours. The effectiveness of 'broad training' depends on such interactions.

Figure 10 Comparisons between BroadK and BroadH (see online version for colours)



	BroadK	BroadH
Dynamic Last round		
K	55.00	195.00
H	1.25	0.35
Q*	9.38	33.38
Average Q	14.09	56.25
Static First round		
K	39.00	
H	0.25	
Q*	17.66	
Average Q	22.91	59.25

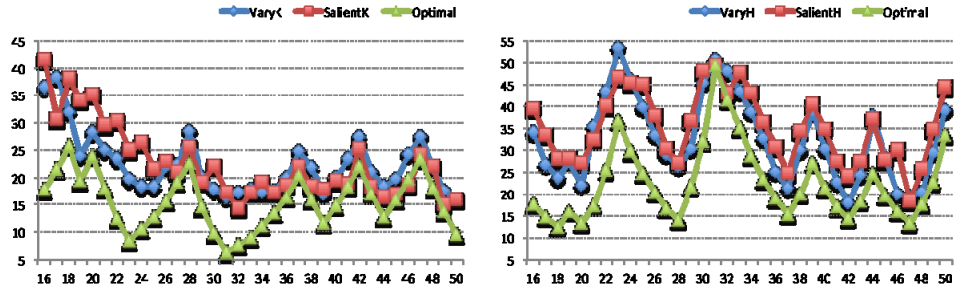
4.2 'Saliency' treatments

In Study 1, varying cost parameters of the dynamic phase are presented in a random sequence. This creates potential difficulty for subjects to identify similar scenarios from the past. Hence, we introduce the 'saliency' treatments: a table is added to the dynamic phase that lists and sorts (up to) five previous rounds by the closeness of the parameter settings to the current round. Again, all other game features are kept the same as those in Study 1. We conduct two more treatments, labelled as *SalientK* with 24 subjects, and *SalientH* with 25 subjects.

We do not expect any behavioural difference in the static phase between the 'saliency' treatments and Study 1, and have verified it statistically. Our main focus of data analysis here is on the dynamic phase, and more specifically, we compare *SalientK* with *VaryK*, and *SalientH* with *VaryH* to see if better performance can be reached with decision support. The aggregated order quantities in the respective treatments over the 35 rounds of the dynamic phase are shown in Figure 11. We can see that the tendency to over order still remains substantial under the 'saliency' treatments in the early rounds of the dynamic phase but reduces over time.

Data from *VaryK* and *SalientK*, and similarly, *VaryH* and *SalientH* are pooled to formally test these effects. In particular, we regress *Ape* against round t and a dummy variable for the treatment. The time trends are negative and highly significant in both data sets, suggesting learning occurs over time in all these treatments. Comparing *SalientK* with *VaryK*, we observe that subjects under *SalientK* start with higher deviations (i.e., positive and significant treatment effect) – they may be confused about the additional information at the beginning when the table is introduced. We also find that the interaction term between treatment and t is negative and significant, implying errors reduce faster under *SalientK*. Thus, it may take longer for decision makers to truly benefit from this intervention. Between *VaryH* and *SalientH*, however, neither the treatment nor its interaction with t is found statistically different from zero. Details of the regression outputs are attached in the Appendix.

Figure 11 Aggregated ordering behaviour under 'saliency' treatments (see online version for colours)



We again find mixed results, with the treatment effect to be more promising under the setting where the fixed cost (K) varies. This is also consistent with our previous speculation about the pulling force between similarity adjustment and reinforcement. When the influence of reinforcement is very strong (i.e., under *VaryH* and *SalientH*), providing decision support that picks out past similar decision-making scenarios for decision makers may not be sufficient to move them away from the reinforcement bias.

5 Conclusions, implications and limitations

We present, to our best knowledge, one of the first behavioural studies that investigates empirical inventory decisions under the basic EOQ model. Experiments are designed to contrast learning behaviours under stationary versus non-stationary decision-making environment. We find that, when costs are static, subjects are less likely to repeat inferior decisions with deterministic feedback from the EOQ model. When confronted with fluctuating costs, a majority of the players are able to improve decisions over time by learning from similar past experience.

Previous behavioural studies have shown that in many operational contexts, empirical behaviours may deviate from normative predictions under a stationary environment, and are dynamic in nature when decision outcomes are under stochastic influence. This paper lends more support for such a viewpoint with a twist of non-stationary decision-making environment but deterministic outcomes. More notably, we identify two competing behavioural dynamics, a modified reinforcement process with similarity adjustment versus an exploratory process that reduces evaluation errors over time and experience. Our estimation results offer further evidence on when and which of the two behavioural models would be better used to analyse the empirical EOQ decisions. It is also noted that reinforcement and learning can be distinct behaviours given non-stationary parameter settings.

Some interesting correlations among several behavioural traits, i.e., the tendencies to reinforce and make random errors, the exploratory learning ability and adaptability to a similar environment are detected in Study 1. These results may have important implications on the assessments of employees during the screening and selection process for jobs related to inventory management, in addition to the standard psychometric tests commonly used by organisations. Moreover, insights from Study 1 suggest how

performance can be enhanced, and we conducted additional experiments to examine two interventions:

1 to increase the chance that an individual would encounter a 'similar' decision making scenario through experience-based practice

2 to provide decision support that can highlight past similar scenarios for decision makers to concentrate on.

Results from Study 2 imply that the effectiveness of these interventions depends on the relative strength of different behavioural traits (i.e., the pulling force between similarity adjustment and reinforcement bias) and the type of operational cost under the EOQ problem (i.e., the fixed or the variable cost) that fluctuates. Our research offers managerial guidance as to how additional functions or tools in ERP systems can be developed to enhance inventory planning and control in a non-stationary parametric environment.

We would like to discuss several limitations of this study and offer some suggestions for future research. First, we intentionally choose a simple experimental setting of the deterministic EOQ problem. Its simplicity serves to exclude known biases and preferences due to uncertainty, and thus helps highlight several insights of our behavioural models. It would be interesting to test whether the behavioural models we proposed apply to decision making under uncertainty, for example, an EOQ model with stochastic demand. Second, under our game setting, it is straightforward to measure how similar a past decision scenario is to the current one by normalised Euclidean distances in the parameter space. However, it remains an open question as to how to define 'similarities' under more complex, or less structured situations, and whether decision makers would still adjust their behaviours according to such similarities. Third, to test the conjecture of similarity adjustment, we control cost shocks in the experiments with relatively smaller variability (i.e., CV<1), follow-up research may consider to validate our results with greater dispersion in the cost shocks. In this study, we follow the standard behavioural economics approach to use undergraduate students as research subjects. A few studies employ inventory managers to play the newsvendor game (Bolton et al., 2012) and the beer distribution game (Tokar et al., 2012), and find results to be robust to education and professional experience. It will be interesting to further check the subject pool effect under the EOQ setting.

Lastly, our study suggests that management should consider the relative strength of different behavioural traits in constructing the interventions to improve inventory performance. This requires future research to develop methods to calibrate, and if possible, manipulate these behavioural factors directly. While we are able to identify correlations between certain behavioural traits through our behavioural models, we have only speculations as to how they are related. A more systematic investigation of the underlying cognitive or psychological mechanism is needed to understand why and how such relationships form.

Acknowledgements

Yan (Diana) Wu acknowledges the financial support provided by RSCA Seed Grant at San Jose State University that made this research possible.

Conflict of interest

Kay-Yut Chen has a potential research conflict of interest due to a financial interest with companies Hewlett-Packard Enterprise, Boostr and DecisionNext. A management plan has been created to preserve objectivity in research in accordance with UTA policy.

References

Bolton, G.E. and Katok, E. (2008) 'Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback', *Manufacturing and Service Operations Management*, Vol. 10, No. 3, pp.519–538.

Bolton, G.E., Ockenfels, A. and Thonemann, U.W. (2012) 'Managers and students as newsvendors', *Management Science*, Vol. 58, No. 12, pp.2225–2233.

Bostian, A., Holt, C. and Smith, A. (2008) 'Newsvendor 'pull-to-center' effect: adaptive learning in a laboratory experiment', *Manufacturing and Service Operations Management*, Vol. 10, No. 4, pp.590–608.

Camerer, C. and Ho, T. (1999) 'Experience-weighted attraction learning in normal form games', *Econometrica*, Vol. 67, No. 4, pp.827–874.

Camerer, C.F. and Ho, T.H. (1998) 'Experience-weighted attraction learning in coordination games: probability rules, heterogeneity and time-variation', *Journal of Mathematical Psychology*, Vol. 42, Nos. 2–3, pp.305–326.

Chen, K.Y., Wang, J. and Lang, Y. (2022) 'Coping with digital extortion: an experimental study of benefit appeals and normative appeals', *Management Science*, Vol. 68, No. 7, pp.5269–5286.

Chen, Y., Su, X. and Zhao, X. (2012) 'Modeling bounded rationality in capacity allocation games with the quantal response equilibrium', *Management Science*, Vol. 58, No. 10, pp.1952–1962.

Choi, T.M. (2014) *Handbook of EOQ Inventory Problems*, Springer, New York, NY.

Donohue, K., Katok, E. and Leider, S. (2018) *The Handbook of Behavioral Operations*, Wiley, NJ.

Dow, J.K. and Endersby, J.W. (2004) 'Multinomial probit and multinomial logit: a comparison of choice models for voting research', *Electoral Studies*, Vol. 23, No. 1, pp.107–122.

Hajivassiliou, V., McFadden, D. and Ruud, P. (1996) 'Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results', *Journal of Econometrics*, Vol. 72, Nos. 1–2, pp.85–134.

Harris, F.W. (1913) 'How much stock to keep on hand factory', *The Magazine of Management*, Vol. 10, pp.240–241.

Hausman, J.A. and Wise, D.A. (1978) 'A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences', *Econometrica: Journal of the Econometric Society*, Vol. 46, No. 2, pp.403–426.

Kalkanci, B., Chen, K.Y. and Erhun, F. (2011) 'Contract complexity and performance under asymmetric demand information: an experimental evaluation', *Management Science*, Vol. 57, No. 4, pp.689–704.

Kalkancı, B., Chen, K.Y. and Erhun, F. (2014) 'Complexity as a contract design factor: a human-to-human experimental study', *Production and Operations Management*, Vol. 23, No. 2, pp.269–284.

Kim, J.B., Albuquerque, P. and Bronnenberg, B.J. (2017) 'The probit choice model under sequential search with an application to online retailing', *Management Science*, Vol. 63, No. 11, pp.3911–3929.

Lang, Y., Su, J. and Chen, K.Y. (2022) *Strategic Disposal or Strategic Inventory? Theory and Experiments*, Working Paper, State University of New York at Oneonta, Oneonta, NY.

Lee, L.F. (1992) 'On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models', *Econometric Theory*, Vol. 8, No. 4, pp.518–552.

Nembhard, D.A. and Bentefouet, F. (2012) 'Parallel system scheduling with general worker learning and forgetting', *International Journal of Production Economics*, Vol. 139, No. 2, pp.533–542.

Nembhard, D.A. and Osothsilp, N. (2005) 'Learning and forgetting-based worker selection for tasks of varying complexity', *Journal of the Operational Research Society*, Vol. 56, No. 5, pp.576–587.

Pan, J., Shachat, J. and Wei, S. (2020) 'Cognitive reflection and economic order quantity inventory management: an experimental investigation', *Managerial and Decision Economics*, Vol. 41, No. 6, pp.998–1009.

Pan, J., Shachat, J. and Wei, S. (2022) 'Cognitive stress and learning economic order quantity (EOQ) inventory management: an experimental investigation', *Decision Analysis*, Vol. 19, No. 3, pp.189–254.

Schweitzer, M.E. and Cachon, G.P. (2000) 'Decision bias in the newsvendor problem with a known demand distribution: experimental evidence', *Management Science*, Vol. 46, No. 3, pp.404–420.

Stangl, T. and Thonemann, U.W. (2017) 'Equivalent inventory metrics: a behavioral perspective', *Manufacturing and Service Operations Management*, Vol. 19, No. 3, pp.472–488.

Su, X. (2008) 'Bounded rationality in newsvendor models', *Manufacturing and Service Operations Management*, Vol. 10, No. 4, pp.566–589.

Tokar, T., Aloysius, J.A. and Waller, M.A. (2012) 'Supply chain inventory replenishment: the debiasing effect of declarative knowledge', *Decision Sciences*, Vol. 43, No. 3, pp.525–546.

Truong, N.C.D., Wang, X., Wanniarachchi, H., Lang, Y., Nerur, S., Chen, K.Y. and Liu, H. (2022) 'Mapping and understanding of correlated electroencephalogram (EEG) responses to the newsvendor problem', *Scientific Reports*, Vol. 12, No. 1, pp.1–16.

Wanniarachchi, H., Lang, Y., Wang, X., Pruitt, T., Nerur, S., Chen, K.Y. and Liu, H. (2021) 'Alterations of cerebral hemodynamics and network properties induced by newsvendor problem in the human prefrontal cortex', *Frontiers in Human Neuroscience*, Vol. 14, p.598502.

Wu, D.Y. and Chen, K.Y. (2014) 'Supply chain contract design: impact of bounded rationality and individual heterogeneity', *Production and Operations Management*, Vol. 23, No. 2, pp.253–268.

Notes

1 Note that the overall cost curves still differ in the corresponding rounds across treatments, as identical total costs appear only at the optimal decisions.

2 These are core courses intended for sophomores and juniors majoring in business.

3 In theory, a subject can order anything larger than zero and thus incur unbounded costs. In Study 1, we calibrated a lump sum of $100 per round for the subject to manage the product inventory. To minimise the incidence of bankruptcy, we place an upper bound on the decision input implicitly. If an order placed is larger than 100 units, a message will pop up requiring the subject to reduce the quantity. We did not state this requirement directly in the instruction due to concerns of anchoring behaviour. The frequency of observing negative profits in a round is less than 1%. None of our subjects experiences a loss at the end of the game.

4 With specific cost functions discussed in the instruction and the help of the calculator, it is possible for a player to figure out the optimal solution to the EOQ problem analytically. In fact, we did observe a few such subjects (5 out of 79) who played the optimal decisions for more than 80% of the time. Excluding these subjects does not change any of our results, and thus they are kept in all of our analysis.

5 There is no statistical difference in subjects' order quantities across sessions in the Static Phase (all p-values > 0.4 by the Mann Whitney U test). Some significant variations occur in the cost comparisons during the first 5 rounds, which disappear in the later rounds of the phase.

6 Due to the tendency to over order, the empirical decision distribution does not appear to be symmetric. The above log transformation also helps address the issue of skewed data. For the regression analysis, we choose the ordinary least square models instead of mixed effect ones as they are less dependent on the normality assumptions. In addition, we bootstrapped p-values for coefficients with significance levels between 1 to 5%, and found our results to be robust.

7 We did use a standard regression with a square term to test if the responses to costs are nonlinear. We only find significant nonlinearities in the VaryH treatment, but not the others.

8 We have also tested a model that allows individuals to weigh the distance in the ordering cost and the holding cost differently, and find no evidence of such behaviours. Therefore, this feature is not included in the final model.

9 For the 5 subjects who have found the analytical solutions, since there are still enough variations in their decisions to obtain valid estimates from the two behavioural models, they are kept in the analysis.

10 The comparison is robust, and the results remain the same if the Bayesian Information Criterion (BIC) is used.

11 The standard formulation of

12 Note that the same parameters per round are used in Study 1 and Study 2. Since there is no statistical difference in the Static Phase among treatments of Study 1, we use pooled data from Study 1 for the comparisons.

Appendix

Part I: Sample experimental instructions

Game scenario

Imagine you are a manager at a store. You need to manage the inventory for a particular product that is carried at the store and purchased from a supplier. The product is non-perishable that can be carried over time. The product demand D is at a constant rate of *1 unit per week*. The current inventory management system of the store works as follows:

The store stocks a certain amount of the product. Its inventory level reduces at a steady rate as the demand for the product arrives over the weeks. Once the inventory drops to a level where there is not enough stock to satisfy the next week's demand, the system automatically informs the supplier to place a replenishment order. The size of the order quantity, which we call Q, needs to be determined by the manager (you). The supplier has guaranteed to deliver your entire order in the same day so that no stock-out would ever occur at the store. In other words, whenever the store is short of stock, the system will automatically bring the inventory level back to your selected Q with no delays.

Every time the system places an order, the supplier charges the store an ordering/shipping cost of K, independent of the shipment size, for processing and delivering the order within the same day. In addition, the store incurs an inventory holding cost H, for carrying one unit of the product for one week. The average total operational cost per week under this inventory system is calculated as follows:

$$\textit{Average ordering cost per week} = \frac{D}{Q} \times K$$

$$\textit{Average holding cost per week} = \frac{D}{2} \times H$$

$$\textit{Average total cost per week} = \textit{Average ordering cost} + \textit{Average holding cost}$$

For example, suppose in a round of the game, we have D = 1 per week, K = \$100 per order placed, H = \$1 per unit per week. If you choose an order quantity of Q = 1, your average ordering cost per week = (D = 1) ÷ (Q = 1) × (K = \$100) = \$100. Since demand is 1 per week and you choose Q = 1, it implies that you will be ordering one time per week, and for the time you order, you are charged with \$100. Average holding cost per week = (Q = 1)/2 × (H = \$1) = \$0.5. Given your Q = 1, the beginning inventory at each week is 1 unit, the ending inventory at each week is 0. The average inventory level held for each week is (0 + 1) / 2 = 0.5 unit, multiplying the \$1 per unit per week holding cost leads to the average holding cost per week of \$0.5. The total cost per week will be 100 + 0.5 = \$100.5. If you choose an order quantity of Q = 100, your average ordering cost per week = (D = 1) ÷ (Q = 100) × (K = \$100) = \$1. Since demand is 1 per week and you choose Q = 100, it implies that you will be ordering one time every 100 weeks, less frequently than before. Average holding cost per week = (Q = 100)/2 × (H = \$1) = \$50. Given your Q = 100, the average inventory level held for each week is (100)/2 = 50 units, higher than before. The average total cost per week will be 1 + 50 = \$51. Note that smaller the Q, more frequent ordering but less inventory carried on average; larger the Q, less frequent ordering but more inventory carried on average.

Your task

The game consists of *50* rounds. In each round, the parameters of the holding cost H and the ordering cost K will be provided for you to decide your order quantity Q (a non-negative integer). Your decisions will be run for (hypothetically) multiple weeks with the demand D always fixed at 1 unit per week. The computer will then return the average cost performance (i.e. ordering, holding and total costs per week calculated according to the equations discussed before) given your decision for the round.

In this experiment, from round 1 to 15 (*or from round 36 to 50 in 'broad' training treatments*), both the holding cost H and the ordering cost K stay unchanged; from round 16 to 50 (*or from round 1 to 35 in 'broad training' treatments*), the ordering cost K (*or the holding cost H*) stays unchanged but the holding cost H (*or the ordering cost K*) is changing every round.

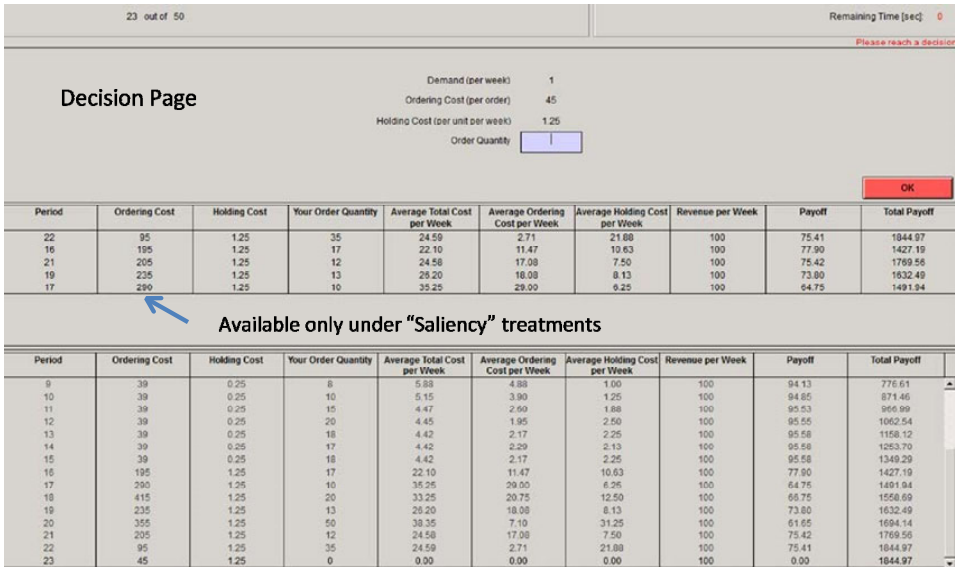
In each round, a fixed amount of *\$100* is given to you for managing the product. The average weekly costs generated from your decision will be subtracted from this lump sum to calculate your payoff for the round. At the end of the game, the investigator will announce the conversion rate, and your payoff accumulated in the entire game will be

converted to US dollars and paid to you in cash. Note that the more profit you earn in each round, higher will be your final cash earnings.

Feedback

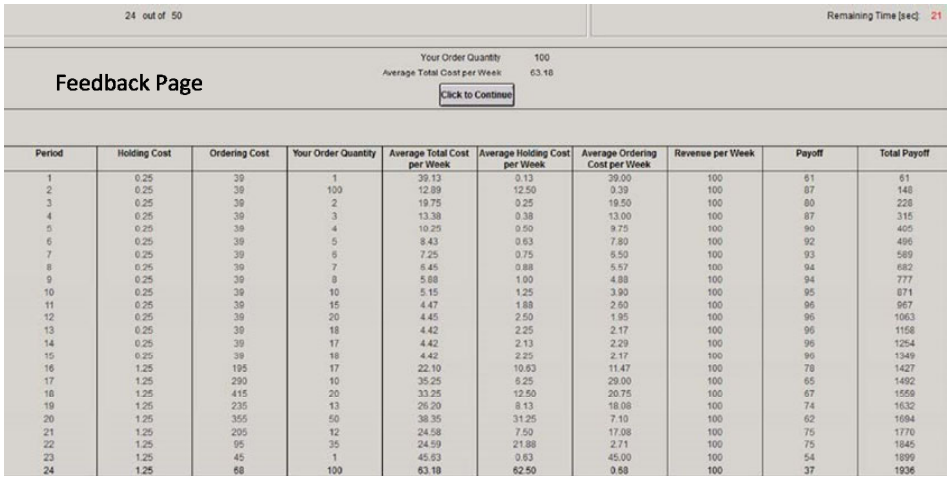
A history table that summarises your past decisions, costs and payoff results will be shown both before and after you make the decision in each round. In addition (*for 'saliency' treatments only*), in Period 16 to 50, before you make the decision, another table will be shown to list (up to) five past decision rounds with holding costs H (*or ordering costs K*) that are closest to the holding cost (*or ordering cost*) of the current round.

Part II: screenshots of the game interface (example from the 'saliency' treatment) (see online version for colours)



Period	Ordering Cost	Holding Cost	Your Order Quantity	Average Total Cost per Week	Average Ordering Cost per Week	Average Holding Cost per Week	Revenue per Week	Payoff	Total Payoff
22	95	1.25	35	24.59	2.71	21.88	100	75.41	1844.97
16	195	1.25	17	22.10	11.47	10.63	100	77.90	1427.19
21	205	1.25	12	24.58	17.08	7.50	100	75.42	1769.56
19	235	1.25	13	26.20	18.08	8.13	100	73.80	1632.49
17	290	1.25	10	35.25	29.00	6.25	100	64.75	1491.94

Period	Ordering Cost	Holding Cost	Your Order Quantity	Average Total Cost per Week	Average Ordering Cost per Week	Average Holding Cost per Week	Revenue per Week	Payoff	Total Payoff
9	39	0.25	8	5.88	4.88	1.00	100	94.13	776.61
10	39	0.25	10	5.15	3.90	1.25	100	94.85	871.46
11	39	0.25	15	4.47	2.60	1.88	100	95.53	966.99
12	39	0.25	20	4.45	1.95	2.50	100	95.55	1062.54
13	39	0.25	18	4.42	2.17	2.25	100	95.58	1158.12
14	39	0.25	17	4.42	2.29	2.13	100	95.58	1253.70
15	39	0.25	18	4.42	2.17	2.25	100	95.58	1349.29
16	195	1.25	17	22.10	11.47	10.63	100	77.90	1427.19
17	290	1.25	10	35.25	29.00	6.25	100	64.75	1491.94
18	415	1.25	20	33.25	20.75	12.50	100	66.75	1558.69
19	235	1.25	13	26.20	18.08	8.13	100	73.80	1632.49
20	355	1.25	50	38.35	7.10	31.25	100	61.65	1694.14
21	205	1.25	12	24.58	17.08	7.50	100	75.42	1769.56
22	95	1.25	35	24.59	2.71	21.88	100	75.41	1844.97
23	45	1.25	0	0.00	0.00	0.00	100	0.00	1844.97



Period	Holding Cost	Ordering Cost	Your Order Quantity	Average Total Cost per Week	Average Holding Cost per Week	Average Ordering Cost per Week	Revenue per Week	Payoff	Total Payoff
1	0.25	39	1	39.13	0.13	39.00	100	61	61
2	0.25	39	100	12.89	12.50	0.39	100	87	148
3	0.25	39	2	19.75	0.25	19.50	100	80	228
4	0.25	39	3	13.38	0.38	13.00	100	87	315
5	0.25	39	4	10.25	0.50	9.75	100	90	405
6	0.25	39	5	8.43	0.63	7.80	100	92	496
7	0.25	39	6	7.25	0.75	6.50	100	93	589
8	0.25	39	7	6.45	0.88	5.57	100	94	682
9	0.25	39	8	5.88	1.00	4.88	100	94	777
10	0.25	39	10	5.15	1.25	3.90	100	95	871
11	0.25	39	15	4.47	1.88	2.60	100	96	967
12	0.25	39	20	4.45	2.50	1.95	100	96	1063
13	0.25	39	18	4.42	2.25	2.17	100	96	1158
14	0.25	39	17	4.42	2.13	2.29	100	96	1254
15	0.25	39	18	4.42	2.25	2.17	100	96	1349
16	1.25	195	17	22.10	10.63	11.47	100	78	1427
17	1.25	290	10	35.25	6.25	29.00	100	65	1492
18	1.25	415	20	33.25	12.50	20.75	100	67	1559
19	1.25	235	13	26.20	8.13	18.08	100	74	1632
20	1.25	355	50	38.35	31.25	7.10	100	62	1694
21	1.25	205	12	24.58	7.50	17.08	100	75	1770
22	1.25	95	35	24.59	21.88	2.71	100	75	1845
23	1.25	45	1	45.63	0.63	45.00	100	54	1899
24	1.25	68	100	63.18	62.50	0.68	100	37	1936

Part III: Regression analysis in Study 2

Table A1 Learning behaviour comparisons under the static phase

DV = APE	*Study 1 vs. BroadK*	*Study 1 vs. BroadH*	*BroadK vs. BroadH*
Intercept	0.8524***	2.6427***	2.6427***
	(0.1190)	(0.1489)	(0.1529)
t	–0.0582***	–0.1288***	–0.1288***
	(0.0131)	(0.0163)	(0.0168)
Treatment	–0.4245**	1.3657***	–1.7902***
	(0.1346)	(0.1668)	(0.2115)
t x Treat	0.0014	–0.0692***	0.0706**
	(0.0148)	(0.0183)	(0.0233)
Adjusted R^2	0.0806***	0.1268***	0.2562***

Notes: Significance codes: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001
treatment = 0 for study 1 and 1 for BroadK or BroadH

Table A2 Learning behaviour comparisons under the dynamic phase

DV = APE	*SalientK vs. VaryK*	*SalientH vs. VaryH*
Intercept	1.1101***	1.4114***
	(0.1190)	(0.1148)
t	–0.0149***	–0.0210***
	(0.0034)	(0.0033)
Treatment	0.5767**	–0.0859
	(0.1769)	(0.1369)
t x Treat	–0.0144**	0.0026
	(0.0051)	(0.0052)
Adjusted R^2	0.0412***	0.0534***

Notes: Significance codes: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001
treatment = 0 for VaryK/VaryH and 1 for SalientK/SalientH.