A Heuristic Approach to Explore: Value of Perfect Information

Shervin Shahrokhi Tehrani and Andrew T. Ching[1]

Rotman School of Management

University of Toronto

September 20, 2017

Please download the latest version at

VPI

[1]Rotman School of Management, University of Toronto, 105 St. George St, Ontario M5S 3E6, Canada. Phone: (416)-879-6242. E-mail: s.shahrokhi13@rotman.utoronto.ca.

Abstract

How do consumers choose in a dynamic stochastic environment when they face uncertainty about the return of their choice? The classical solution to this problem is to assume consumers use dynamic programming to obtain the optimal decision rule. However, this approach has two drawbacks. First, it is computationally very expensive to implement because it requires solving a dynamic programming problem with a continuous state space. Second, a decision maker is assumed to behave "as if" she optimally processes information regardless of its cognitive tractability. To address these two issues, we propose a new heuristic decision process called Value of Perfect Information (VPI), which extends the idea first proposed by Howard (1966) in the Engineering literature. This approach provides an intuitive and computationally tractable way to capture the value of exploring uncertain alternatives. Intuitively, in VPI, a decision maker investigates a subset of information which are expected by her to improve her myopic decision outcome. We argue that our VPI approach provides a "fast and frugal" way to balance the tradeoffs between exploration versus exploitation. More specifically, the VPI approach only involves ranking the alternatives and computing a one-dimensional integration to obtain the expected future value counterpart. In terms of computational costs, we show that the VPI approach is significantly simpler than the standard dynamic programming approach, making it a very practical learning model for consumers to employ. Moreover, the VPI approach provides switching patterns which differ from the independence of irrelevant alternatives (IIA) property, and generates potentially more realistic asymmetric switching reactions to price promotions. Using individual level scanner data, we find evidence that our VPI approach is able to capture consumers' choice well.

Contents

"*Information has no value at all unless it has the potential to change a decision.*"
Sam L.Savage, The Flow of Averages

1 Introduction

How do consumers choose in a dynamic stochastic environment when they face uncertainty about the return of their choice? This question has been studied for a long time in marketing and economics (Ching et al. (2013) and Ching et al. (2017) for reviews). Researchers have mostly embraced the assumption that consumers are rational and adopted the dynamic programming approach to model consumer's choice under such environment. The main advantage of this approach is that it is very parsimonious. By simply supplementing one parameter (the discount factor)[1] to the current payoffs functions, the dynamic programming approach can achieve the optimal balance between exploring/experimenting unfamiliar alternatives and exploiting the more familiar ones.

However, due to the curse of dimensionality problem, it is very difficult to implement this approach in practice. The state variables of a learning problem are continuous, and hence one can only discretize the state space and solve for an approximated version of the model. The finer the discretization, the larger the state space and the higher the computational burden. This is why some researchers simply abandon the exploration incentive by assuming consumers are myopic. For researchers who use the dynamic programming approach, they have resolved to using approximation methods such as Keane and Wolpin (1994), Ching et al. (2012), Rust (1997), Imai et al. (2009), and Chow and Tsitsiklis (1991). Given the computational costs of solving for the optimal solution of a dynamic learning problem, it is natural to conjecture that consumers use a more heuristic approach to balance exploration vs. exploitation. Psychologists have often argued that the limitation of human's cognitive ability would lead to decision rules that are "fast and frugal" (Bettman et al. (1998), Payne et al. (1988), Payne et al. (1993)). Therefore, consumers' bounded rationality motivates us to examine more realistic decision processes to model consumers' choice behavior (Simon (1955), Tversky and Kahneman (1975), Hutchinson and Meyer (1994), and Gigerenzer and Goldstein (1996)). From a managerial point of view, accounting for consumers bounded rationality may also lead to a better demand model, which is essential for managers to set efficient marketing strategies.

[1] Note that identifying and estimating discount factor is a challenging problem (e.g., Ching and Osborne (2017)). In the literature, researchers usually set the discount factor according to the prevailing interest rate. Here, our VPI approach avoid this issue because it does not involve the discount factor parameter.

The objective of this paper is to propose a heuristic approach (Value of Perfect Information) and provide evidence that it can capture the actual consumer choice under uncertainty. We argue that this heuristic approach (1) can *significantly* reduce the computational burden by eliminating the costly dynamic programming steps in estimation process, and (2) is a more realistic and pragmatic approach to model consumers' behavior towards a exploration-exploitation trade-off.

The rest of the paper is structured as follows. Section 2 reviews related literatures to learning dynamic models and heuristic solutions. In Section 3, we set up a benchmark brand choice model. Section 4 posits the VPI approach with respect to other heuristic descriptive models in learning dynamics. Section 5 provides theoretical results to show that how VPI generates switching patterns which do not follow the independence of irrelevant alternatives (IIA) property; we argue that its asymmetric switching reactions to price promotions could better capture the observed consumer choice behavior. Sections 6 and 7 provide evidence to support Myopic-VPI's performance by using simulations and estimations on field data. By comparing the mean absolute errors of predicted switching patterns and the average of correctly predicted choices, we provide further evidence to support better VPI's performance to explain consumers' choice. Finally, Section 9 concludes paper.

2 Literature Review

We propose our Myopic-VPI approach by bridging three areas: Marketing, Consumer Behavior, and Artificial Intelligence. Before we introduce Myopic-VPI strategy, we review concepts from literatures on learning dynamics, the statistical and heuristic descriptive models, and related reinforcement learning techniques to clarify our approach.

2.1 Learning Dynamics

How do consumers choose in a dynamic stochastic environment when they face uncertainty about the return of their choice? There are substantive types of evidence in economics and marketing literatures that consumers involve a learning process (Ching et al. (2013) Ching et al. (2017)). Consumers know exploring new alternatives provide valuable information. The information helps them to improve their long-term payoffs. However, uncertainty about a new alternative's return incentivizes consumers to choose more familiar options by trusting only on previous experiences. This is called the exploration-exploitation trade-offs. In their seminal work, Erdem and Keane (1996) (EK) develop and estimate a learning model based on dynamic programming, and use it to study this trade-offs for frequently purchased experienced goods. But as we mentioned in the introduction, the high computational burden of dynamic programming has led some researchers abandon the exploration incentive and assume consumers are myopic (e.g., Mehta et al. (2004), Mehta et al. (2008), Narayanan and Manchanda (2009), Ching and Ishihara (2012)).

On the other hand, research based on experimental data finds evidence that consumers do consider the exploration-exploitation trade-offs. But both myopic and fully rational dynamic programming approaches fell short in fully capturing some dimensions of consumer choice patterns. This prompts us to propose a model which serves as a middle ground between these two extreme approaches.

In the behavioral economics and consumer behavior literatures, researchers suggested consumers' decision process should be "fast and frugal" (Bettman et al. (1998), Payne et al. (1988)). The growing of cognitive descriptive models creates a desire to introduce richer cognitively tractable approaches. For example Chintagunta et al. (2006) (p. 614) suggest that "*the future development of structural models in marketing will focus on the interface between economics and psychology.*" Our VPI model fits well in this line of reasoning.

2.2 Heuristic Models

In the statistic literature, several innovative decision processes were introduced in which the computational complexity have been resolved by decomposing the Bellman equation to suboptimal problems. The *multi-armed bandit* process is a benchmark dynamic stochastic process where the decision maker is uncertain about the reward distribution of each "arm" (i.e., alternative choice).[2] Gittins and Jones (1979) provide an Index Strategy solution to multi-armed bandit process by a *forward induction* argument. They assigned index (called *Gittins Index*) to each arm at each state. These indices are evaluated by solving a set of *one-dimensional* dynamic programming problem. Gittins (1979) proves that the Index Strategy is an optimal solution to a multi-armed bandit process. The Gittins index is equivalent to the *reservation price* concept in Pandora's optimal search rule in Weitzman (1979). The main assumption in a multi-armed bandit process is that all arms (i.e, chosen or the inferior alternatives) have stable reward distributions through time periods. The inferior alternatives' states are frozen at the next time period as they were.

[2]Note that the brand choice model, when there is no observed and unobserved shocks, could be interpreted as a multi-armed bandit process.

Later, Whittle (1988) relaxes the stability of arms assumption such that all arms evolve according to a Markov process regardless of chosen action. That is called the restless bandit process.[3] Under *indexability* condition, Whittle provides a generalization of Gittins index called *Whittle Index*.[4] He suggests a *heuristic decision process* in which the arm with highest index is chosen at each time step. His Index Strategy provides a *sub-optimal* solution in a restless bandit process. *Index Strategy* is computationally morel tractable. That has motivated several researchers in marketing and economics to use it as a behavioral decision process by agents (Miller (1984), Eckstein et al. (1988), Lin et al. (2015), and Dickstein (2014)).

Most existing research focuses on approximating the solution of DP problems. Geweke and Keane (2000) (GK) takes a different approach by relaxing the assumption that consumers solve a DP problem. They use a flexible functional form to approximate the expected future payoffs and estimate its parameters using data on choice and payoffs.[5] Ching et al. (2014) extend the GK approach to estimate a dynamic learning model. They argue that the dynamic learning problem has exclusion restrictions which allow them to estimate the expected future payoffs function even when only choice data is available. The estimated expected future payoffs function could be used to shed light on what heuristic rule consumers use to balance the exploitation vs. exploration incentives.

Why should researchers consider heuristic models? Thinking is a costly activity for decision makers (Camerer (2003)). The theory of bounded rationality (Simon (1955) and Simon (1956)) suggests that individuals' rationality is limited by incomplete information, and their time and cognitive ability constraints. Clearly, to make decision according to dynamic programming is cognitively very demanding. In the consumer behavior and behavioral economics literatures, researchers provide evidence that individuals decision-making process is "fast and frugal" (Bettman et al. (1998), Payne et al. (1988), Payne et al. (1993), Gabaix and Laibson (2000)). They argue that consumers prefer descriptive models which provide them higher ratio of $\frac{Payoffs}{Cost\ of\ Thniking}$.[6] Our heuristic model will fit this criterion.

[3]Note that the brand choice model in 3.1 could be regarded as a restless bandit process.

[4]Here, the existence of indexes should be proved according to set up assumptions. For example, Lin et al. (2015) show that the brand choice model in 3.1 is indexable if (1) brands' reward distributions are *uncorrelated* and (2) observed and unobserved shocks are *independently* and *identically* idiosyncratic random variables. That is a non-trivial result constrained by structural assumptions on corresponded uncertain environment. One of advantages of VPI approach is its freedom to implementing it without the above structural constraints.

[5]Houser et al. (2004) apply the GK method to estimate a dynamic career choice problem in a laboratory setting. They find that one-third of subjects behave close to a DP solution, but the rest uses some kinds of heuristic decision rules.

[6] Note that evaluating the cost of thinking is a challenging task which needs an experimental setting by

Hutchinson and Meyer (1994) summarize evidence that the fully rational forward-looking model falls short in explaining some features of consumer choice behavior. Through their experimental studies, Meyer and Shi (1995) and Gans et al. (2007) provide evidence that consumers are neither myopic nor fully forward-looking. These experimental multi-arms bandit studies show that individuals value information which can be obtained through exploration; but their decision making process are far from optimal balance between exploitation vs exploration. Gabaix et al. (2006) provides evidence that individuals prefer to involve costly cognitive process of impending information only if it improves outcomes respect to myopic behavior. These studies suggest that consumers make their decisions *heuristically* to accomplish the following goals: (1) to achieve a better performance than what a myopic consumer would do, and (2) to save the cost of thinking. Our Myopic-VPI is designed in a way to achieve the above goals. Moreover, it reduces the computational burden significantly by eliminating DPs steps.

2.3 Reinforcement Learning

To design an efficient autonomous agent, which should perform in an unknown environment, computer scientists need to implement algorithms wherein the agent will be able to learn and act quickly with a negligible loss respect to the optimal solution. That philosophy provides us with interesting decision processes which are "fast and frugal". For example, in Computer Science (CS), the future component $\beta E[V(I_{it+1} \mid I_{it})]$, is regarded as the exploration bonus for agents. Interestingly, in reinforcement learning, an autonomous agent also prefers an efficient algorithm which provides the highest $\frac{Payoffs}{Time\ of\ Computation}$ (Meuleau and Bourgine (1999)).

Computer Scientists extensively study the exploration vs exploitation tradeoffs in learning models in the artificial intelligence (AI) and machine learning literature (Dearden et al. (1999), Dearden et al. (1998), Kaelbling (1993), Meuleau and Bourgine (1999), Watkins and Dayan (1992)). One heuristic rule stands out in terms of both computational tractability and performance. This approach is based on the Value of Perfect Information (VPI) concept first introduced by Howard (1966). It has been considered as a very efficient algorithm which outperforms rival approaches in AI (Teacy et al. (2008)). Our goal is to incorporate VPI in a structural learning model, and demonstrate that it could capture the actual behavior of consumers by using individual level consumer choice data.

measuring subjects' reaction time to make a decision based on a descriptive model. However, computational time and mathematical complexity of operations in a descriptive model have been shown as proxies and surrogates for cost of thinking (Bettman et al., 1998). As in literature, we assume computational time reflects required cost of thinking.

3 Standard Bayesian Learning Model

Here, we need a benchmark learning model to introduce the Myopic-VPI. We are following this literature, as in Ching et al. (2014), to model consumer learning behavior about brand quality of an experience good. Due to computational burden of the fully rational dynamic programming approach, researchers usually summarize consumers' uncertainty of product attributes in a one-dimensional weighted average of attributes called quality.[7] For simplicity, we will follow this assumption. But we note that one advantage of our VPI approach is that it can be applied to multi-dimensional learning problems.[8]

3.1 Benchmark Model

A consumer sequentially chooses brands from a set A containing J brands within a product category. Consumers do not know the true quality of brands. We assume that they have a normal prior beliefs about brand j's quality:

$$Q_j \sim N(Q_{j0}, \sigma_{j0}^2). \tag{1}$$

The values of Q_{j0} and σ_{j0}^2 may be influenced by advertising, pre-sale marketing campaign, word of mouth, reputation of manufacturer, etc. For simplicity, we assume that consumers have common prior beliefs on quality of brand j and is normally distributed. But note that our approach is not restricted by functional forms of distributions or utility.

If a consumer buys brand j, she will learn about the quality of brand j through consumption trial. This experience does not fully reveal the quality because of "inherent variability". According to Ching et al. (2014), inherent variability can be interpreted as existence of (a) quality variation of different units of a product, or (b) evaluation variation of quality of a product in different use occasions. Therefore, the purchase of brand j at time t only provides a noisy signal Q_{ijt}^E of the true quality Q_j as in:

$$Q_{ijt}^E = Q_j + \xi_{ijt} \quad \text{where} \quad \xi_{ijt} \sim N(0, \sigma_\xi^2) \tag{2}$$

The variance σ_ξ^2 captures the "inherent variability" which we refer to it as "experience vari-

[7] Crawford and Shum (2005) is an exception wherein patients learn about two attributes of each drug: (1) symptomatic and (2) curative attributes.

[8] As we see later, to apply Myopic-VPI, a researcher does not need to prove indexability of a dynamic problem. Essentially, Myopic-VPI can be applied to any learning dynamic model. For example, our approach can be simply generalized to model decision process in a multi-dimensional leaning environment as in Crawford and Shum (2005).

ability" as in Ching et al. (2014).

Since the prior belief and noisy signals are normally distributed, the posterior belief for perceived quality at $t = 1$ is

$$Q_{ij1} = \frac{\sigma_{j0}^2}{\sigma_{j0}^2 + \sigma_\xi^2} Q_{ij1}^E + \frac{\sigma_\xi^2}{\sigma_{j0}^2 + \sigma_\xi^2} Q_{j0} \tag{3}$$

$$\sigma_{ij1}^2 = \frac{1}{(1/\sigma_{j0}^2) + (1/\sigma_\xi^2)} \tag{4}$$

Equation (4) implies that $\sigma_{ij1}^2 < \sigma_{ij0}^2 = \sigma_{j0}^2$. That means the experience signal Q_{ij1}^E is valuable in the sense that it reduces consumer's uncertainty about the true quality Q_j. Note that

$$\sigma_{ij1}^2 - \sigma_{j0}^2 = \frac{-\sigma_{j0}^4}{\sigma_{j0}^2 + \sigma_\xi^2},$$

which is strictly decreasing in σ_{j0}^2. So exploring of a brand with higher uncertainty provides more valuable information for consumers.

Equations (3) and (4) generalize to any number of signals. If $N_{ij}(t)$ denotes the total number of experience signals received by consumer i through time t, then we have

$$Q_{ijt} = \frac{\sigma_{j0}^2}{N_{ij}(t)\sigma_{j0}^2 + \sigma_\xi^2} \sum_{s=1}^{t-1} Q_{ijs}^E d_{ijs} + \frac{\sigma_\xi^2}{N_{ij}(t)\sigma_{j0}^2 + \sigma_\xi^2} Q_{j0} \tag{5}$$

$$\sigma_{ijt}^2 = \frac{1}{(1/\sigma_{j0}^2) + N_{ij}(t)(1/\sigma_\xi^2)} \tag{6}$$

where d_{ijs} is an indicator for whether brand j is bought at time t by consumer i. Note that if consumer i receives infinite number of signals (i.e., $N_{ij}(t) \to \infty$), then $\sigma_{ijt}^2 \to 0$ and $Q_{ijt} \to Q_j$. That explains the infinite accumulation of information of experience signals of brand j is valuable such that it will reveal the true quality to consumer i. However, consumer i receives only finitely many random signals on quality of brand j in reality. Let $I_{it} = \{Q_{ijt}, \sigma_{ijt}^2\}_{j=1}^J$ denote the endogenous information state of consumer i at time t. Clearly, I_{it} is randomly evolved through time t according to the received noisy experience signals. Since the experience signals are random, the learning model *endogenously* generates heterogeneity across consumers by developing different information states among them through time t (Ching et al. (2013)).

Following Ching et al. (2014), we assume consumer i's (conditional indirect) utility of consuming brand j is:

$$u_{ijt} = f(Q^E_{ijt}) - w_p p_{ijt}, \tag{7}$$

where p_{ijt} is the observed price at time t for brand j. To capture risk-aversion behavior, we assume $f(Q^E_{ijt})$ takes the constant absolute risk aversion form (CARA). Since the experience signal Q^E_{ijt} is random, the consumer i considers the expected utility is given by:

$$E[u_{ijt} \mid I_{it}] = -exp\left(-r\left(Q_{ijt} - \frac{r}{2}\left(\sigma^2_{ijt} + \sigma^2_\xi\right)\right)\right) - w_p p_{ijt}, \quad j = 1, \cdots J, \tag{8}$$

where $I_{it} = \{Q_{ijt}, \sigma^2_{ijt}\}^J_{j=1}$ is the information state of consumer i at time t. Finally, we let $\theta = \{w_p, r, \{Q_{j0}, \sigma^2_{j0}, Q_j\}^J_{j=1}, \sigma^2_\xi\}$ be our set of parameters.[9]

Now, the main question is "How consumer i will choose her desired brand at each purchase time t?" The two extreme decision-making behaviors are (1) *Myopic* and (2) *Fully Forward-Looking* Behavior.[10] These two types of models are extensively studied in the marketing literature. Below, we will briefly explain the myopic and fully forward-looking models, and then argue that the heuristic decision-making process is a plausible middle-ground between these two extreme modelling approaches.[11]

3.2 Myopic Decision Maker

In myopic behavior descriptive model, a consumer only considers her current utility at time t. Let

$$V^M_j(\theta \mid I_{it}) := -exp\left(-r\left(Q_{ijt} - \frac{r}{2}\left(\sigma^2_{ijt} + \sigma^2_\xi\right)\right)\right) - w_p p_{ijt},$$

denote the expected current utility of consumer i from choosing brand j at time t. We call it the *myopic utility* denoted by superscript M. A myopic consumer i chooses brand j at time t if and only if

$$V^M_j(\theta \mid I_{it}) + e_{ijt} > V^M_k(\theta \mid I_{it}) + e_{ikt} \quad k \neq j, k = 1, \cdots J. \tag{9}$$

[9]Here we assume a risk averse utility functional form. Our arguments in this paper could be easily generalized to a risk neutral utility function.

[10]Here, we assume that there is a homogeneous foresight among consumers toward the brand choice model. We are able to generalize our result for a heterogeneous foresight model in a similar way as in Lin et al. (2015).

[11]Ching et al. (2013) provides a detailed review of models which build on these two extreme settings.

The e_{ijt} is an $i.i.d$ unobserved preference shock by consumer i at time t. By assuming that e_{ijt} are independent and identically distributed extreme value Type I, the choice probabilities have a simple closed form:

$$P_i(j \mid I_{it}) = \frac{\exp(V_j^M(\theta \mid I_{it}))}{\sum_{k=1}^{J} \exp(V_k^M(\theta \mid I_{it}))}. \tag{10}$$

To construct the likelihood function, we let $j(t)$ denote the choice actually made at time t. Since quality signals are latent variables, simulated maximum likelihood method (SML) should be applied. $D_{it-1} = \{j(s)\}_{s=1}^{t-1}$ is the purchase history of consumer i through time t. Let $Q_{it-1}^E = \{Q_{ij(s)s}^E\}_{s=1}^{t-1}$ denote the actual values of all experience signals received up through t by consumer i. We define $P_i(j(t) \mid I_{i1}, D_{it-1}, Q_{it-1}^E)$ as the conditional probability of consumer i's choice at time t.

Because we do not observe Q_{it-1}^E, we integrate over the Q_{it-1}^E to obtain $P_i(j(t) \mid I_{i1}, D_{it-1})$, which will form the basis of our likelihood. If the panel choice date of individual i is available for T periods, then

$$\begin{aligned} &P_i(\{j(t)\}_{t=1}^T) \\ &\qquad = \prod_{t=1}^{T} P_i(j(t) \mid I_{it}) \prod_{t=1}^{T} P_i(j(t) \mid I_{i1}, D_{it-1}) \\ &\qquad = \int_{\{Q_{it-1}^E\}_{t=1}^{T=1}} P_i(j(t) \mid I_{i1}, D_{it-1}, Q_{it-1}^E) d_\mu \{Q_{it-1}^E\}_{t=1}^{T=1}. \end{aligned} \tag{11}$$

As explained in Ching et al. (2013), we are able to identify the relative means of prior beliefs, the relative uncertainties of prior beliefs, the true means of quality, and price sensitivity. These parameters could be estimated by SML.

There are several papers which implement myopic decision-making process (e.g., Roberts and Urban (1988), Mehta et al. (2004), Mehta et al. (2008) Camacho et al. (2011)). In a myopic learning model, the consumer only exploits her best expected current choice. There is no incentive to explore, i.e., a myopic consumer has no incentive to strategically consider the benefits of choosing any other brands to gain information to improve her performance in the future. The computational tractability is the main advantage of the myopic learning model. Psychologists consider the myopic model as the most effortless decision-making process. Here, a consumer does not need to think *deeply* about expected informational benefits of choosing inferior alternatives.

However, choosing a brand with a higher uncertainty (i.e., higher σ^2_{ijt}) could provide more valuable information since it reduces the uncertainty more significantly by experience trial. This is called exploration advantage. Clearly, the expected accumulated rewards of a myopic consumer is not optimal since exploration-exploitation tradeoff is totally neglected. For example, Erdem and Keane (1996) and Ching et al. (2014), provide evidence that the forward-looking models fit the data better than the myopic learning model. Those are consistent with experimental studies which show that people comprehend the value of exploration to obtain valuable information in a sequential decision process (Meyer and Shi (1995), Gans et al. (2007)). The tradeoff between exploitation vs exploration is an important factor in decision-making under uncertainty (Kaelbling et al. (1996)). In sum, although myopic learning model provides us with the best computational tractability from a practitioners' perspective, it may not be able to capture the actual consumer behavior. By following myopic model, the researcher ignores consumers' incentive to explore in the market. Thus, it provides inaccurate and biased estimation of parameters. *In sum, we should address exploration behavior in consumers' decision process according to field and experimental types of evidence.*

3.3 Fully Forward-Looking Decision Maker

In the fully forward-looking behavior, consumers optimally balance the exploration-exploitation tradeoff. Intuitively, choosing a more uncertain brand provides more valuable information for future choices. Meanwhile, consumer i needs to consider its current expected payoff which is based on exploitation of her prior belief. So she chooses a brand that gives her the highest expected myopic utility $V_j^M(\theta \mid I_{it})$ *plus* the expected discounted *future payoffs* $\beta EV[I_{it+1} \mid I_{it}, j]$. That is, consumer i chooses the highest alternative specific value function

$$V(j,t \mid I_{it}) = V_j^M(\theta \mid I_{it}) + \beta EV[I_{it+1} \mid I_{it}, j] + e_{ijt} \quad \text{for} \quad j = 1 \cdots J, \tag{12}$$

where e_{ijt} is $i.i.d$ unobserved shock. That means

$$j(t) = \arg\max_{j \in A} V(j,t \mid I_{it}). \tag{13}$$

Let $V_j^F(\theta \mid I_{it}) := V_j^M(\theta \mid I_{it}) + \beta EV[I_{it+1} \mid I_{it}, j]$ which is called the *fully forward-looking utility* of brand j denoted by superscript F. Note that (12) states exactly Bellman's principle of optimality. In contrast to fully forward-looking model, the myopic consumer neglects the future component $\beta EV[I_{it+1} \mid I_{it}, j]$ by focusing her judgement based on $\{V_j^M(\theta \mid I_{it})\}_{j=1}^J$. The future component captures the optimal value of exploration of brand j (called exploration bonus in the reinforcement learning). A fully forward-looking consumer optimally balances the tadeoffs between exploitation, i.e., $V_j^M(\theta \mid I_{it})$, vs exploration bonus, i.e., $\beta EV[I_{it+1} \mid I_{it}, j]$.

A fully forward-looking consumer i chooses brand j at time t if and only if

$$V_j^F(\theta \mid I_{it}) + e_{ijt} > V_k^F(\theta \mid I_{it}) + e_{ijt} \quad k \neq j, k = 1, \cdots J. \tag{14}$$

If e_{ijt} are independent and identically distributed extreme value, the choice probabilities have a simple closed forms:

$$P_i(j \mid I_{it}) = \frac{\exp(V_j^F(\theta \mid I_{it}))}{\sum_{k=1}^J \exp(V_k^F(\theta \mid I_{it}))}. \tag{15}$$

Note that (15) is identical to (10) except the future component is added. Theoretically, the future component could be solved by costly dynamic programming (DP) because of non-existence of *analytic solutions*. It should be evaluated numerically by iteration. The likelihood function could be constructed identically as in (11). Here, a researcher needs to approximate the future component at any information state. The fully forward-looking model could be estimated with a Bayesian or simulated maximum likelihood methods. Although a fully forward-looking model provides an optimal consumers' exploration behavior in contrast to myopic model, it raises two important issues: (1) its *computational tractability* in implementing learning models from practitioners point of views and (2) its *cognitive tractability* to capture human decision process from a more realistic point of view.[12]

The first issue is that the estimation of a forward looking model is computationally intractable. According to Lin et al. (2015), the problem is PSPACE-hard. That means that it is at least as hard as NP-hard problems, which are themselves suspected of being unsolvable in polynomial time. Moreover, the information states are continuous high dimensional spaces. Despite of using discretization methods (Chow and Tsitsiklis (1991), Rust (1997)), we still have the curse of dimensionality in solving the DP problem in fully forward-looking model. Let Ω denote the set of the finite discrete information states for each consumer. Therefore, we should evaluate the future component at $\mid \Omega \mid^J$-information states. Clearly, it has ex-

[12] Cognitive tractability is a concept in cognitive psychology. This literature shows that human minds/brains are finite systems with limited resources for computation. So Human cognitive capacities are constrained by computational tractability (Van Rooij (2008)). This is called bounded rationality in Economics (Simon (1956)) or "fast and frugal" decision process in Consumer Behavior (Gigerenzer and Goldstein (1996) and Bettman et al. (1998)).

ponential growth rate. In summary, the optimal approach to exploitation vs exploration tradeoff in the fully forward-looking model comes at the cost of computational intractability. Keane and Wolpin (1994), Ching et al. (2012), Rust (1997), and Chow and Tsitsiklis (1991) provide different approaches to reduce the computational burden of estimation of a dynamic choice model. Although they achieve partial solutions to overcome computational burden, their algorithms still suffer from the "curse of dimensionality."

The above issues suggest that consumers may not behave "as if" they are fully rational (in the sense that their choices are solution to the DP problem), given their limited cognitive ability (Van Rooij (2008), Ching et al. (2014), Meyer and Shi (1995), Gabaix et al. (2006), Gans et al. (2007)). The experimental evidence shows people's behavior are somewhere between myopic and fully forward-looking behavior (Meyer and Shi (1995) and Gans et al. (2007)). Instead, consumers maybe bounded rational and adopt a "fast and frugal" heuristic rule to balance the tradeoff between exploitation and exploration, even though it may seem suboptimal compared with the dynamic programming solution.[13]

In the next section, we are going to propose a new heuristic decision process which overcomes the computational and cognitive complexities of forward looking models by viewing exploration value differently in contrast to previous literature. The intuitive concept *the value of perfect information*, is extended to learning dynamic models. In our approach, a consumer makes her decision in a way to attain a higher outcome than myopic counterpart since she knows values of exploration; while she prefers to investigate only those types of information which potentially help her to improve outcomes in contrast to myopic behavior (since she is aware of her cognitive limitation). Interestingly, our approach can be considered as a generalization of "directed cognition" model (Gabaix and Laibson (2005)) to a dynamic uncertain environment (i.e., a multi-armed bandit problem). The objective is to show our new heuristic approach, Myopic-VPI, can be considered as a better "as if" model to (1) reduce computational and technical issues to estimate learning dynamic models in a more efficient way; and (2) explain more realistically consumers' choice behavior by regarding cognitive complexity of a decision process.

[13]However, if a consumer can reach the decision using a heuristic rule much faster than the dynamic programming approach, he/she can sample different alternatives much more often within a short period of time. If we consider a fixed window, a simple heuristic rule can lead to significantly higher payoffs.

Table 1: Prior Beliefs Towards Brand A and B

Brand	**High Quality Utility**	**Low Quality Utility**
A	100 with Probability 0.5	50 with Probability 0.5
B	260 with Probability 0.2	25 with Probability 0.8

4 The Value of Perfect Information

In this section, we present our proposed heuristic model called **Myopic-VPI** wherein consumers are exploring an alternative based on its *Value of Perfect Information* (VPI). In Howard (1966), the VPI was proposed as the expected improvement in future decision quality arising from the information acquired by exploration. The famous quote, by Sam L.Savage in the Flow of Averages, succinctly conveys the VPI's intuition:

> "*Information has no value at all unless it has the potential to change a decision.*"

The VPI concept is not only intuitive from a behavioral viewpoint; but also a superior technique in a dynamic leaning model. Teacy et al. (2008) show the VPI approach improves significantly the agents' performance in a complex uncertain environment compared to other heuristic approaches.[14] We will design VPI approach as a descriptive model wherein a decision maker (1) intuits exploration value based on potential of information to change her myopic judgement and (2) assists her to befit form strategic exploration without using a cognitively demanding dynamic programming approach.

4.1 A Simple Example

In this section, we explain the value of perfect information, through a very simple example, to intuitively introduce our VPI approach. We will generalize our idea in a formal definition in the next section.

Let's assume that a consumer wants to choose between brands A or B. She is going to purchase these products only for two periods (e.g., for two weeks). We assume the discount factor is $\beta = 0.9$. The consumer does not know the true quality state of brands. Thus, she has uncertainty about utility values by purchasing these products. However, she has a prior belief about each product's quality state as in Table 1. For simplicity, we assume each experience signal is a perfect signal which reveals true quality of brand j after purchasing it. So, one consumption trial is enough to know the quality state in this example.

[14] Note that the environment was designed in a way that the optimal solution was infeasible to implement or evaluate.

It is straightforward to show that the expected current utility of brand A and B are $V_A^M = 0.5(100) + 0.5(50) = 75$ and $V_B^M = 0.2(260) + 0.8(25) = 72$, respectively. So if she is a myopic consumer, she will buy brand A in the first week. If A ends up to be a high quality product, she will choose it in the second week too ($V_B^M < 100$); if A shows up as a low quality product, she will switch to B in the second week ($V_B^M > 50$). Here, a myopic consumer summarizes her prior knowledges to first moments of prior distributions (i.e., means). This generates a very fast decision process. *However, the negative side of her myopia is to ignore all other available information which have potential to improve her long-run payoffs.* For example, if A and B are low and high quality products respectively, she will lose 210 utils by following myopic behavior.

Her fully forward-looking counterpart will follow the Bellman optimality approach as follows:

$$\begin{aligned} V_A^F &= V_A^M + \beta E[I_2 \mid I_1, A] \\ &= 75 + \beta\Big[0.5 \max\{100, V_B^M\} + 0.5 \max\{50, V_B^M\}\Big] = 152.4. \end{aligned}$$

$$\begin{aligned} V_B^F &= V_B^M + \beta E[I_2 \mid I_1, B] \\ &= 72 + \beta\Big[0.2 \max\{260, V_A^M\} + 0.8 \max\{25, V_A^M\}\Big] = 172.8. \end{aligned}$$

A fully forward-looking consumer, at time $t = 1$, completely screen out all possible expected signals. This will help her to figure out the exploration value of each alternative. Since alternative B has higher uncertainty than A, exploring it is more promising than A. Clearly, a fully forward-looking consumer will buy brand B in the first week. If B ends up to be a high quality product, she will choose it in the second week too. If B turns out to be a low quality product, she will switch to A in the second week. *However, the fully forward-looking model ignores time constraints and cognitive costs due to processing all information nodes in her decision tree.*

Now we are in a position to informally introduce our VPI approach to overcome the caveats of the Myopic or Fully Forward-Looking descriptive models. A VPI consumer knows that, at time $t = 1$, A is the superior myopic choice than B. Meanwhile, she is aware that (a) myopic behavior can be harmful in a long-run perspective and (b) her cognitive limitations prevent her to follow the Bellman optimality approach. What should she do? Let's revisit the Sam L. Savage's quote.

"Information has no value at all unless it has the potential to change a decision."

If A and B are expected to end up as high quality and low quality respectively, her choice at time $t = 1$ will remain A as in the myopic behavior. Intuitively, these two possible outcomes do not have the potential to change her myopic choice A. She prefers to spend her cognitive resources more carefully to focus on the information which leads her to change her myopic choice A to a better option. Thus, the information which shows A is a high quality product, i.e., $A = 100$, has no value for her since it won't change the myopic decision. She assigns $Gain(A = 100) = 0$ without doing any costly cognitive operation. What should she do if A is a low quality product, i.e., $A = 50$? The answer is clear, she should switch to B since $V_B^M := 72 > 50$. Thus, by switching to B, she can expect to gain $V_B^M - 50 = 22$. She assigns $Gain(A = 50) = 22$ with doing the costly cognitive operation to evaluate the gain. However, these are ex-ante informational nodes before purchasing brand A. So, her expected gain by exploring brand A is

$$VPI_A = EGain(A) = 0.5 \times Gain(A = 100) + 0.5 \times Gain(A = 50) = 11, \tag{16}$$

which we call it the value of perfect information of A denoted by VPI_A. Therefore, the total value of choosing A, at time $t = 1$, is $V_A^{VPI} := V_A^M + VPI_A = 75 + 11 = 86$. That is equal to A's current expected payoff plus its exploration value.

Now, she should evaluate brand B too. What should she do if B is a high quality product, i.e., $B = 260$? The answer is clear, she should switch to B since $V_A^M := 75 < 260$. Thus, by switching to B, she can expect to gain $260 - V_A^M = 185$. She assigns $Gain(B = 260) = 185$ with doing the costly cognitive operation to evaluate the gain. On the other hand, the information which shows B is a low quality product, i.e., $B = 25$, has no value for her since it won't change the myopic decision. She assigns $Gain(B = 25) = 0$ without doing any costly cognitive operation. As the above, these are ex-ante informational nodes before purchasing brand B. So, her expected gain by exploring brand B is

$$VPI_B = EGain(B) = 0.2 \times Gain(B = 260) + 0.8 \times Gain(B = 25) = 37. \tag{17}$$

Therefore, the total value of B, at time $t = 1$, is $V_B^{VPI} := V_B^M + VPI_B = 72 + 37 = 109$. Since $V_B^{VPI} > V_A^{VPI}$, a VPI consumer will also buy brand B in the first week. If B ends up to be a high quality product, she will choose it in the second week too; while if B shows up as a low quality product, she will switch to A in the second week. This is identical choice behavior as in the optimal model.

It is clear that a VPI consumer shows a "partially" forward looking behavior to (a) enjoy strategic exploration value; (b) recognize her cognitive limitations in fully analyzing her decision tree.[15] VPI approach should be considered as a suboptimal solution which decision maker chooses it as a "fast and frugal" decision process. For example, if the consumer's prior belief of B, in Table 1, changes such that B returns only 180 utility units with probability 0.2, then the VPI consumer behavior will be the same as the myopic counterpart. This shows that the VPI should be regarded as a bounded rational descriptive model relative to cognitively costly fully forward-looking behavior. We will show how this intuitive concept can be generalized in the context of multi-armed bandit problems. More importantly, we will build a structural model to investigate if the VPI approach is supported by consumer choice data.

4.2 Value of Perfect Information for Exploration

Let's consider the benchmark model in 3.1 in time t at information state I_{it}, where $I_{ijt} = \{Q_{ijt}, \sigma^2_{ijt}\}$. Without loss of generality, we assume that

$$V_1^M(\theta \mid I_{it}) > V_2^M(\theta \mid I_{it}) > \cdots > V_J^M(\theta \mid I_{it}). \tag{18}$$

It means consumer i knows that the 1st product is the superior current choice regardless of any exploration. At the probabilistic choice level, it implies

$$P_i(1 \mid I_{it}) > P_i(2 \mid I_{it}) > \cdots > P_i(J \mid I_{it}). \tag{19}$$

Eq (19) says that if consumer i behaves myopically, the first brand will have the highest chance to be chosen by consumer i. We assume myopic behavior is the least costly cognitive approach to make a decision – consumer i only needs to consider her prior belief without incurring the cost of thinking of strategic exploration.

Intuitively, consumer i has an assessment of quality distributions at time t. Consumer i wants to know her expected benefit to explore brand j by purchasing it at t if its true quality is Q_j^*. What can be gained by learning the true value Q_j^*? How would this knowledge change consumer's i's expected future rewards? The key idea of VPI relies on the following intuition: If this knowledge does not change the best myopic choice of consumer i at time t (which is the 1st brand by (18)), then the expected future rewards would not change. **Thus,**

[15]Note that our above argument can also be implemented in the context of the optimal search behavior proposed by (Weitzman, 1979) called Pandora's Box Problem. Also, it is generalizable to any multi-armed bandit process without any barrier such as indexability.

that information has no value. So, the only interesting scenarios are those where the new knowledge Q_j^* does change her myopic choice at t. The VPI approach assumes that consumer i bears cognitive cost of thinking on consequences of Q_j^* if and only if she expects an improvement on her performance. This can happen in only two cases: (a) when the new knowledge determines that a brand previously considered sub-optimal is revealed as the best choice (given consumer i's belief), and (b) when the new knowledge indicates that a brand that was previously considered the best is actually inferior to other brands. Each scenario can be detailed as follows:

(a) Let $j \neq 1$. If consumer i hypothetically thinks of quality of brand j to be Q_j^*, then j will provide utility $u_{ijt}^* = -\exp(-r.Q_j^*) - w_p p_{ijt}$ for her. In case (a), the new knowledge indicates that the purchase of brand j (where $j \neq 1$) is a better choice for her if $u_{ijt}^* > V_1^M(\theta \mid I_{it})$. Thus, consumer i expects to gain $u_{ijt}^* - V_1^M(\theta \mid I_{it})$ by virtue of purchasing j instead of the best myopic choice. That is, the exploration bonus gained by consumer i by hypothetically resolving the uncertainty of thinking about j's quality level at Q_j^* in her decision process. Regarding cost of thinking, in VPI, consumer i's cognitive cost to explore is significantly smaller than evaluating optimal future component $\beta E[I_{it+1} \mid I_{jt}, j]$ wherein consumer i needs to bear an enormous cognitive cost to determine the expected gain at all nodes in an infinite depth decision tree.[16]

(b) Let $j = 1$. If consumer i hypothetically thinks of quality of the best myopic choice brand to be Q_1^*, then it will provide utility $u_{i1t}^* = -\exp(-r.Q_1^*) - w_p p_{i1t}$ for her. In case (b), the new knowledge indicates that the purchase of the best myopic choice brand is actually an inferior choice for her if $u_{i1t}^* < V_2^M(\theta \mid I_{it})$. Thus, consumer i expects to gain $V_2^M(\theta \mid I_{it}) - u_{i1t}^*$ by virtue of purchasing the *second best myopic choice* instead of the *best myopic choice*. Intuitively, consumer i thinks about consequences of exploring the superior alternative if its true quality is *low* enough to justify a switch. Similar to scenario (a), "the cost of thinking" to achieve VPI of Q_1^* is significantly smaller than the fully forward-looking dynamic programming approach.

[16]Note that, in Index Strategy, consumer i also needs to to determine the expected gain at an infinite depth decision tree which involves only brand j's stochastic process. This is still enormous cognitive cost in contrast to VPI; but significantly less complex than the optimal solution by breaking a J-dimensional DP to a J two DP optimal stopping problems.

Combining the above arguments, the *gain* from learning the value Q_j^* is defined as

$$Gain_{ijt}(Q_j^*) = \begin{cases} V_2^M(\theta \mid I_{it}) - u_{i1t}^* & \text{if} \quad j = 1 \quad \& \quad u_{i1t}^* < V_2^M(\theta \mid I_{it}) \\ u_{ijt}^* - V_1^M(\theta \mid I_{it}) & \text{if} \quad j \neq 1 \quad \& \quad u_{ijt}^* > V_1^M(\theta \mid I_{it}) \\ 0 \quad \text{otherwise} \end{cases} \tag{20}$$

Since consumer i does not know in advance what value will be revealed for Q_j^*, she needs to compute the expected gain, given her prior belief on qualities. Hence the expected value of perfect information about brand j is:

$$VPI_{ijt} = \int_{-\infty}^{+\infty} Gain_{ijt}(x) b_{ijt}(x) dx, \tag{21}$$

where $b_{ijt}(x) \sim N(Q_{ijt}, \sigma_{ijt}^2)$ is the prior density distribution of quality of brand j at time t for consumer i. We are calculating the VPI_{ijt} in the following section.

4.3 VPI-Formulation

We assume consumer i's (conditional indirect) utility of consuming brand j at quality level Q_j^* is $u_{ijt}(Q_j^*) = f(Q_j^*) - w_p p_{ijt} \quad \forall j \in A$. The function, $f(.)$, can be risk-neutral utility (e.g., $f(Q_j^*) = Q_j^*$), the CARA utility, or any weakly increasing function in Q_j^*.

We start to evaluate the VPI of the superior myopic choice, i.e., $j = 1$. If consumer i hypothetically assumes that the true quality is Q_1^*, then the consumer's utility $u_{i1t}^* = f(Q_1)^* - w_p p_{i1t}$. By definition of the gain function at (20), consumer i realizes a positive potential value of information Q_1^* if and only if:

$$\begin{aligned} u_{i1t}^* < V_2^M(\theta \mid I_{it}) &\iff f(Q_1^*) - w_p p_{i1t} < V_2^M(\theta \mid I_{it}) \\ &\iff f(Q_1^*) < V_2^M(\theta \mid I_{it}) + w_p p_{i1t} \\ &\iff Q_1^* < f^{-1}\left(V_2^M(\theta \mid I_{it}) + w_p p_{i1t}\right). \end{aligned}$$

Intuitively, this says consumer i needs only to pay attention to the consequences of deviating from her best myopic choice if its quality is lower than $\zeta_{i1t} = f^{-1}\left(V_2^M(\theta \mid I_{it}) + w_p p_{i1t}\right)$. Otherwise, she does not have an incentive to incur any cognitive cost to evaluate the other end of the quality distribution because those types of information won't change her myopic judgement. Briefly, if she invests her second thoughts to figure out what she can gain if her superior myopic choice turns out to be of low quality. So

$$
\begin{aligned}
VPI_{i1t} &= \int_{-\infty}^{+\infty} Gain_{i1t}(x) b_{i1t}(x) dx \\
&= \int_{-\infty}^{\zeta_{i1t}} Gain_{i1t}(x) b_{i1t}(x) dx \\
&= \left(V_2^M(\theta \mid I_{it}) + w_p p_{i1t}\right) \Phi\left(\beta_{i1t}\right) - \left(Q_{i1t} - \sigma_{i1t} \frac{\phi(\beta_{i1t})}{\Phi(\beta_{i1t})}\right),
\end{aligned}
\tag{22}
$$

where $\beta_{i1t} = \frac{\zeta_{i1t} - Q_{i1t}}{\sigma_{i1t}}$; and ϕ and Φ denote the density and cumulative function of standard normal distribution.

What about an inferior alternative? Let's $j \neq 1$. If consumer i hypothetically assumes that the true quality is Q_j^*, then the consumer's utility $u_{ijt}^* = f(Q_j)^* - w_p p_{ijt}$. Now, we have

$$
\begin{aligned}
u_{ijt}^* > V_1^M(\theta \mid I_{it}) &\iff f(Q_j^*) - w_p p_{ijt} > V_1^M(\theta \mid I_{it}) \\
&\iff f(Q_j^*) > V_1^M(\theta \mid I_{it}) + w_p p_{ijt} \\
&\iff Q_j^* > f^{-1}\left(V_1^M(\theta \mid I_{it}) + w_p p_{ijt}\right).
\end{aligned}
$$

Intuitively, a forward-looking consumer i needs to adjust her decision if the inferior choice j has a better quality than her average prior belief. This is a valuable information for her to go after. If j's quality is higher than $\zeta_{ijt} = f^{-1}\left(V_1^M(\theta \mid I_{it}) + w_p p_{ijt}\right)$, then she should switch to j to enjoy higher expected utility. Otherwise, she does not have incentive to pay any cognitive cost of rethinking about her best myopic choice since the expected gain will be zero. So

$$
\begin{aligned}
VPI_{ijt} &= \int_{-\infty}^{+\infty} Gain_{ijt}(x) b_{ijt}(x) dx \\
&= \int_{\zeta_{ijt}}^{\infty} Gain_{ijt}(x) b_{ijt}(x) dx \\
&= \left(Q_{ijt} + \sigma_{ijt} \frac{\phi(\beta_{ijt})}{1 - \Phi(\beta_{ijt})}\right) - \left(V_1^M(\theta \mid I_{it}) + w_p p_{ijt}\right)\left(1 - \Phi\left(\beta_{ijt}\right)\right),
\end{aligned}
\tag{23}
$$

where $\beta_{ijt} = \frac{\zeta_{ijt} - Q_{ijt}}{\sigma_{ijt}}$; and ϕ and Φ denote the density and cumulative function of standard normal distribution.

Empirically, we are going to evaluate VPI_{ijt} using a Monte-Carlo integration method. More precisely, we can make R random draws of qualities $\{Q^*_{jr}\}_{r=1}^R$ from the normal distribution $b_{ijt}(x) \sim N(Q_{ijt}, \sigma^2_{ijt})$. Then

$$\widehat{VPI}_{ijt} = (1/R) \sum_{r=1}^{R} Gain(Q^*_{jr}). \tag{24}$$

As $R \to \infty$, $\widehat{VPI}_{ijt} \to^p VPI_{ijt}$ by the law of large numbers. Therefore, $\widehat{VPI}_{ijt}$ provides us with a consistent estimator for VPI_{ijt}. From an empirical point of view, calculating VPI_{ijt} needs significantly lower computational burden compared with $\beta E[I_{it+1} \mid I_{jt}, j]$ or W_{ijt} (i.e., Index Strategy) because it only involves solving a one-dimensional integration problem.

Note that (24) not only empirically estimates VPI_{ijt}, it motivates consumer i's behavioral approach in Myopic-VPI. In sum, she first generates her primary judgement (i.e., $\{V_j^M(\theta \mid I_{it})\}_{j\in A}$). Then, she samples those information about j which she thought to have potential to change her myopic decision to a gain. Finally, she considers a weighted average of her possible gains as expected exploration bonus of j.

In summary, VPI approach has the following properties:

1. *If the variance of prior belief of j increases (i.e., higher uncertainty), VPI_{ijt} will increase.*

   The above feature simply says when consumers are more unfamiliar with an alternative, there will be a higher exploration value for it. So, consumers have a larger incentive to do an strategical exploration to learn about unfamiliar option.

2. *As variance of prior belief of j goes to zero, VPI_{ijt} also goes to zero.*

   Clearly, if the $\sigma_{ijt} \to 0$, the consumer i knows precisely what is the relative current position of j to all other alternatives, in terms of current utility. So, she does not need to figure out what will happen if j is either better or worse than she thought (The current expected utility is the true realized utility). So $VPI_{ijt} = 0$

The above features also show up in fully forward looking or index strategy. But VPI approach generates unique properties as follows

3. *The VPI generates an asymmetric learning process.*

- The VPI of superior Myopic choice, at time t, involves processing information at lower tail of its prior belief distribution.
- The VPI of an inferior Myopic choice, at time t, involves processing information at upper tail of its prior belief distribution.

The asymmetric learning process is a very interesting behavioral aspect of VPI approach. A VPI decision maker only analyzes information which incentivizes her to switch away from her myopic behavior. Therefore, she only investigates those information which either showing the superior choice can be worse than her other choices; Or showing inferior choices can be better than her superior choice. In contrast, fully forward-looking or index strategy decision maker will analyze all available information regardless the current expected utilities. Thus, VPI decision maker overcomes her cognitive limitation by rationally focusing on information which better off her than myopic counterpart. Therefore, VPI approach is a bounded rational heuristic descriptive model to choose in a dynamic uncertain environment.

4.4 Myopic-VPI Decision Process

Now, we are in a position to propose our Myopic-VPI decision process. The VPI_{ijt} provides the expected gain of exploration of brand j for consumer i at information state I_{ijt}. However, there is an expected *opportunity cost* incurred for exploring brand j. That the difference between the expected myopic values of j and the best expected myopic choice, i.e., $V_1^M(\theta \mid I_{it}) - V_j^M(\theta \mid I_{it})$. This suggests consumer i should choose the brand that maximizes

$$VPI_{ijt} - \left(\max_{j \in A} V_1^M(\theta \mid I_{it}) - V_j^M(\theta \mid I_{it}) \right). \tag{25}$$

That strategy is equivalent to choosing the alternative j which maximizes

$$V_j^{VPI}(\theta \mid I_{it}) := V_j^M(\theta \mid I_{it}) + VPI_{ijt}, \tag{26}$$

i.e., $j(t) = \arg\max_{j \in A} V_j^{VPI}(\theta \mid I_{it})$. Here, consumer i shows a forward-looking behavior in which she balances exploration vs. exploitation by evaluating VPI_{ijt} at information state I_{it}. Since VPI_{ijt} approximate the value of exploration of j, a Myopic-VPI should be regarded as a forward-looking agent; but she heuristically evaluates value of exploration of j instead of following Bellman principle of optimality which leads to $\beta E[I_{it+1} \mid I_{jt}, j]$.[17]

[17]Note that in Lin et al. (2015), consumer i's choice at t is $j(t) = \arg\max_{j \in A} V_j^M(\theta \mid I_{it}) + W_{ijt}(\theta \mid I_{it})$ where W_{ijt} represents Whittle index. So Whittle index of alternative j proxies $\beta E[I_{it+1} \mid I_{jt}, j]$.

VPI_{ijt} is more computational tractable and likely less cognitively demanding. In Myopic-VPI, consumer i reduces the infinite depth decision tree of evaluating $\beta E[I_{it+1} \mid I_{jt}, j]$ to a cognitive tractable problem in three steps. The first step is to cut the decision tree to a one-layer depth decision tree. In the second step, she only considers the branches that provide significant negative or positive quality signals with respect to her expected current payoff. *In this step, she only considers a subset of information which has potential to change myopic decision* (i.e., VPI). At the final step, the importance of a considered branch is proportional to its probabilistic occurrence. So Myopic-VPI can be interpreted as a generalization of directed cognition model by Gabaix and Laibson (2005) to learning dynamic models. Note that the cognitive cost to intuit VPI_{ijt} is significantly lower than the optimal future component. Moreover, it suffers significantly less computational burden too.

When consumer i is confident about true qualities (i.e., when the estimated value of $V_j^M(\theta \mid I_{it})$ is more accurate), the VPI_{ijt} of each brand is close to 0, and the consumer will always choose the brand with the highest expected value. Intuitively, consumer i wants to explore more at the beginning to gain more information on qualities, but over time, she will prefer to exploit her best myopic choices. When there is unobserved shocks e_{ijt}, a consumer i chooses brand j at time t if and only if

$$V_j^M(\theta \mid I_{it}) + VPI_{ijt} + e_{ijt} > V_k^M(\theta \mid I_{it}) + VPI_{ikt} + e_{ikt} \quad k \neq j, k = 1, \cdots J. \tag{27}$$

The *Myopic-VPI* strategy is:

$$j(t) = \arg\max_{j \in A} \{V_j^{VPI}(\theta \mid I_{it}) + e_{ijt}\}.$$

If e_{ijt} are independent and identically distributed extreme value, the choice probabilities have a simple multinomial logit form:

$$P_i(j \mid I_{it}) = \frac{\exp(V_j^{VPI}(\theta \mid I_{it})}{\exp(V_0^{VPI}(\theta)) + \sum_{k=1}^{J} \exp(V_k^{VPI}(\theta \mid I_{it}))}. \tag{28}$$

Researchers can construct the likelihood function of the brand choice model as in (11). Here, we need only to replace the the conditional choice probabilities by (28). The likelihood function in (11) can be estimated by SML approach. Since we do not observe the signals and unobserved contents, we draw D sets of signals as in:

$$S^d = \{Q^d_{jt-1}\}_{t=1}^T \quad \text{for} \quad d = 1, \cdots, D, \tag{29}$$

using the distributions defined in (2). Then, by the simulated probability, the conditional likelihood function can be approximated by

$$\frac{1}{D} \sum_{d=1}^{D} \prod_{t=1}^{T} P_i(j(t) \mid I_{it}) \prod_{t=1}^{T} P_i(j(t) \mid I_{i1}, D_{it-1}, S^d). \tag{30}$$

Finally, we sum the logs of these probabilities across individuals $i = 1, \cdots, N$. Since VPI_{ijt} at any I_{ijt} can be evaluated by a Monte-Carlo integration, without approximating value functions through DPs, the computational burden should be significantly lower than other approaches to forward-looking learning models. For example, in Erdem and Keane (1996)), a smooth approximation (e.g., Keane and Wolpin (1994)) was used to estimate the future component by an interpolation method. The details are provided in Ching et al. (2013). Also, Lin et al. (2015) use the Index Strategy to reduce the computational burden. Their approach makes use of a transformation formula (Proposition 2, page 8), which requires specific assumptions on the utility function and distribution of the unobserved preference shocks.

But in Myopic-VPI approach, VPI_{ijt} can be evaluated at any I_{ijt} by Monte-Carlo integration. Therefore, we do not need any random grid (as in Rust (1997)) or value iteration to find VPI_{ijt}. Moreover, evaluating VPI does not require any restriction on the structural functional form of distributions and utility function. Therefore, our inner loop in the simulated maximum likelihood approach should be faster and more reliable than others. The above arguments show our VPI approach overcomes the computational burden drawback in estimating learning models.

Note that cognitive psychology shows that human minds/brains are finite systems with limited resources for computation (Van Rooij (2008)). So Human cognitive capacities are constrained by computational tractability. Since the computational complexity of Myopic-VPI is significantly lower than fully forward-looking or Index strategy, the Myopic-VPI could be regarded as a more cognitive tractable decision-making process if it fits into data appropriately. So it provides a more realistic behavioral approach to decision making process in learning dynamics model.

4.5 The Myopic-VPI vs Other Heuristics

We discuss how a Myopic-VPI decision-maker overcomes the drawbacks of myopic or fully forward-looking approaches in learning dynamic models. First, computing VPI_{ijt} only involves computing a one-dimensional integral and does not involve solving for any dynamic programming problems. The Myopic-VPI decision rule can be computed much faster than other fully structural approaches which involve dynamic programming (incl. index strategies). Moreover, it does not suffer the curse of dimensionality problem. Secondly, we show Myopic-VPI involves less cognitive operations. However, it provides a way to strategically consider value of exploration. Now we would like to posit Myopic-VPI respect two previous heuristic approaches which are applied in brand choice models: (1) the Index Strategy proposed by Lin et al. (2015) and (2) GK-approach proposed by Ching et al. (2014).

The main reason for using heuristic decision processes is to reduce the cognitive costs (i.e, cost of thinking). Clearly, consumers want to gain the value of exploration as long as it could help them to perform better than myopic decision process. Let's assume that the myopic decision process is the simplest cognitive decision process. In Myopic-VPI, a decision maker follows her intuition to restrict her attention to a subset of information which has potential to change her myopic judgement toward expected positive gains. This approach is significantly simpler than the Index Strategy, which still requires consumers to solve a series of optimal stopping problems. These optimal stopping problems are still DP problems, though much less complex than the original one. On the contrary, our VPI_{ijt} only requires a consumer to consider expectation based on a truncated distribution. On the other hand, in Index Strategy, we need to assume consumers understand decomposition of a J-dimensional DP to J-two dimensional DPs problems. Then, the index of each alternative should be evaluated in a separate DP process. But Myopic-VPI evaluates the exploration bonuses by comparison expected qualities among brands. Intuitively, consumers only need to intuit value of exploration of each alternative respect to their current best myopic choice throughout a sampling process.[18] Therefore, Myopic-VPI strategically considers the relative benefits of choices rather than evaluating the exploration bonus of a choice through a non-trivial mathematical decomposition as in Index Strategy.

There is also a connection between the Myopic-VPI to the directed cognition model by Gabaix and Laibson (2005). Gabaix et al. (2006) shows directed cognition type models

[18]It is obvious if alternatives' payoff distributions or prior beliefs are correlated, Index Strategy will break down to capture it since indexes are evaluated through separate DPs. On the other hand, Myopic-VPI can be adjusted to deal with correlated learning process. Simply, respect to any draw Q_j^*, expected current myopic utilities , V_j^Ms, should be updated; then, the VPI_{ijt} be evaluated accordingly.

match better laboratory evidence; while several experimental studies reject the existence of Index Strategy in a multi-bandit process (Meyer and Shi (1995), Gans et al. (2007)). These makes Myopic-VPI as a better match for consumers' behavior in an uncertain environment. In summary, we think the Myopic-VPI could resolve the cognitive tractability in a more efficient way than Index Strategy. However, a more precise experimental study is required to determine the cognitive difficulty of Myopic-VPI vs Index Strategy.

Ching et al. (2014) propose a clever identification argument, by leveraging exclusion restriction of consumers' prior uncertainty variance, to estimating learning model by applying GK approach. There, the expected future payoffs is written as a flexible functional form which be estimated along other parameters. However, it is not clear what structural model would generate the specific functional form for the expected future payoffs. So, We need to be careful to apply it to a counterfactual experiment which can change the assumed functional form. But, our VPI approach evaluates the expected exploration values in a structural way which includes all strategical tradeoffs. So, it will be invariant respect to counterfactual experiments.

4.6 The Myopic-VPI vs Dynamic Programming Approach

It is important to discuss the differences between the VPI approach and the expected future payoff generated from dynamic programming (or Bellman equation). One can view the VPI_j term as an analog of $EV[I_{it+1}|I_{it}, j]$ in Eq(12). For each alternative j, the calculation of the Gain function only involves comparing two alternatives, and only focuses on the potential values of the true Q_j which would change one's choice. To simplify the problem, the agent in the myopic-VPI model also assumes that he/she can learn the true quality of an alternative after one trial (but this can be easily relaxed). [The concept of VPI also entices the idea that an individual ultimately may just want to make decisions based on expected current payoffs. The VPI is asking the question, "What could I gain next period (assuming I just make my decisions based on expected current payoffs) if I try this alternative?" Figuring the value of exploration is costly, and hence one may just want to focus on some important aspects of the information, and ignores others.]

In the dynamic programming approach, $EV[I_{it+1}|I_{it}, j]$ also captures such potential gain, and much more. It recognizes that one cannot resolve all uncertainty about an alternative after one trial. More importantly, it recognizes the value of learning the true Q_j even when it does not change one's choice based on current expected payoffs. Such information will reinforce one's choice based on current expected payoffs and the potential upside (or downside) of those values that can make the best choice based on expected current payoffs even more attractive.

The fact that VPI ignores the potential upside (or downside) in the calculation can be viewed as a clever trick. Let's focus on the alternative which I rank highest based on expected current payoffs. The idea is that if the true value of Q_1 lies above a cutoff such that revealing it will not change my belief that brand 1 is still ranked the highest, I will benefit from this higher payoff next period anyway if I continue to choose simply based on V^M. So why bother to figure out the conditional expectation of Q_j above that cutoff if I am going to enjoy it anyway? This idea is intuitively appealing and may likely capture the first order importance of exploration value.

It is also worth highlighting some properties of VPI: (i) Everything else the same, the larger the current belief variance for alternative j, the larger VPI_j; (ii) VPI_j is not a smooth function of parameters – conceivably, when the value of a parameter moves across a cutoff such that it changes the ranking of the top two alternatives, there will be sudden discrete changes in VPI_j given the definition of the VPI functions; (iii) the computation of VPI_j ignores the fact that learning about the true quality usually takes more than one trial. Property (i) captures the idea that an alternative with high variance is attractive because of its potential upside. Property (ii) suggests that there should be some interesting testable empirical implications (which we will discuss more in the next section). It also suggests that the standard derivative-based method to search for parameter vector may not be optimal during estimation. Property (iii) suggests that when a bounded rational consumer weighs VPI_j, his/her utility weight on VPI_j may take into account the rate of learning (determined by his current prior variance and the variance of the signal noise). This suggests that we may want to model the utility weight as a function of their ratio.

5 Asymmetric Switching Pattern and IIA Property

In Marketing, asymmetric switching pattern among brands is well-documented (Blattberg and Wisniewski (1989), Kamakura and Russell (1989), and Allenby and Rossi (1991)). Based on consumers' purchase behavior in scanner datasets, this literature finds higher-price, higher quality (resp. low-price, low quality) brands steal more share from other brands in the same price-quality tier. Also, consumers prefer more strongly to switch to high quality tier (called "switch up") rather than to switch to low quality tier (called "switch down") in a response to price promotions. However, these papers assume that consumers know qualities and purchase without uncertainty. The above evidence can be restated, in a theoretical framework, that a consumers' choice model could more realistically estimate consumers' price elasticity if it breaks the independence of irrelevant alternatives (IIA) property and provides asymmetric switching reactions in response to price promotions. Ching et al. (2009) show in a static two-stage model— wherein a consumer first decides to purchase in a product category; and then she will choose a brand if product category is considered— the IIA can be broken only respect to outside option. This provides more flexibility to capture consumers' purchase incidence. We will show that our VPI approach breaks the independence of irrelevant alternatives (IIA) property and generating asymmetric switching reactions in response to price promotions. Therefore, we should attain a more sensible estimation result since the decision-making model could match with observed asymmetric switching pattern.

5.1 IIA Restriction in Myopic and Fully Forward-Looking Behaviors

Now, let us investigate the price elasticity at time t. If consumer i is a Myopic decision maker, then we have

$$\begin{aligned}\ln\left[\frac{P_i(k' \mid I_{it})}{P_i(k \mid I_{it})}\right] &= \ln\left[\frac{\exp(V^M_{k'}(\theta \mid I_{it}))}{\exp(V^M_{k}(\theta \mid I_{it}))}\right]\\ &= V^M_{k'}(\theta \mid I_{it}) - V^M_{k}(\theta \mid I_{it}).\end{aligned} \tag{31}$$

This implies that if $k' \neq k$,

$$\frac{\partial \ln\left[\frac{P_i(k'|I_{it})}{P_i(k,|I_{it})}\right]}{\partial p_{ijt}} = \begin{cases} -w_p & \text{if } k' = j \\ w_p & \text{if } k = j \\ 0 & \text{if } k' \quad \& \quad k \quad \neq j \end{cases} \tag{32}$$

The (32) simply states that if price of product j will change at time t, it *will* effect *only* the ratio of product j market share respect to other products; while j's price shock *won't* have any effect on the ratio of market shares of other products. This is the famous restriction called Independence of Irrelevant Alternatives (IIA). *More importantly, the absolute marginal effect of product j's price promotion, at time t, is equal and symmetric respect to all its rival.* Intuitively, it says that if j drops its price by a \$1, it will steal w_p-units of ratio of market share form all other brands; while the market share ratio of other products remain as before. Therefore, the Myopic descriptive model suffers to reflect the typical asymmetric switching pattern. This will arrive finally to an inaccurate estimation respect to price promotions. Note that when a descriptive choice model suffers from IIA and symmetric switching pattern, if either high quality or low quality products more frequently provide price promotion in a sample data, the model will generate a upward or downward bias estimation of price coefficient.

More interestingly, we face a similar IIA restriction if consumer i behaves "as if" fully-forward looking or index strategy. In learning literature, researchers usually assume that consumers have a rational expectation about future prices. That means consumers evaluate expected future component, $EV[I_{it+1} \mid I_{it}, j]$, based on exogenous price distributions $p_j \sim N(\overline{p_j}, \sigma^2_{p_j}) \quad \forall j \in \{1, \cdots, J\}$ where $\overline{p_j}$ and $\sigma^2_{p_j}$ denote the price mean and variance of product j in sample data. Since retailers choose price promotions in a random process in a competitive equilibrium (Narasimhan (1988) and Villas-Boas (1995)), the rational expectation about future prices is a reasonable and suitable assumption. Simply, consumers only predict future price based on each products' average price and its variance in the market.

Since prices are assumed i.i.d over time, so, $\forall s \geq t+1$ and $\forall j \in A$, p_{ijs} *will not* be a function of current prices $\{p_{ijt}\}_{j=1}^{J}$ at time t. Therefore, we have

Lemma 1 *According to the above setting, we have* $\frac{\partial EV[I_{it+1}|I_{it},j]}{\partial p_{ikt}} = 0$ *for any* $k \in \{1, \cdots, J\}$, *i.e., the expected future payoff of any product is not a function of current prices of products.*

Lemma 1 says that if product j changes its price at time t to $p_{ijt} + \eta$, where $\eta \in (-p_{ijt}, +\infty)$, the consumer i's utility of consumption j will *only* change through its current expected utility; while the discounted exploration bonus $\beta EV[I_{it+1} \mid I_{it}, j]$, won't be different. The expected future component is only a function of informational quality signals at time t not prices. For example, if consumer i receives a significant promotion from familiar product, she still have a same exploration incentive toward more unfamiliar product as before although her relative current expected utility of consumption the known brand increases.

The above features in either fully forward looking or index strategy[19] imply

[19] Note that in the Index Strategy, the exploration bonus of each alternative is evaluated separately than other choices at time t. So the expected future exploration value of alternative j is only a function of $\forall s \geq t+1$, p_{ijs}

$$\ln\left[\frac{P_i(k' \mid I_{it})}{P_i(k \mid I_{it})}\right] = \ln\left[\frac{\exp(V^F_{k'}(\theta \mid I_{it}))}{\exp(V^F_k(\theta \mid I_{it}))}\right]$$
$$= V^F_{k'}(\theta \mid I_{it}) - V^F_k(\theta \mid I_{it}) \tag{33}$$
$$= \left[V^M_{k'}(\theta \mid I_{it}) - V^M_k(\theta \mid I_{it})\right] + \beta\left[EV[I_{it+1} \mid I_{it}, k']) - EV[I_{it+1} \mid I_{it}, k])\right]$$

According to Lemma 1,

$$\frac{\partial \ln\left[\frac{P_i(k',t|I_{it})}{P_i(k,t|I_{it})}\right]}{\partial p_{ijt}} = \begin{cases} -w_p & \text{if } k' = j \\ w_p & \text{if } k = j \\ 0 & \text{if } k' \quad \& \quad k \quad \neq j \end{cases} \tag{34}$$

Th similarity of (32) and (34) implies that even if we assume consumer i is fully-forward looking, respect to a price promotion, the consumers' switching pattern suffers from IIA and a symmetric market share marginal effect.

5.2 Myopic-VPI and Breaking IIA

Let's consumer i sorts the products according expected current utilities, i.e.,

$$V^M_1(\theta \mid I_{it}) > V^M_2(\theta \mid I_{it}) > \cdots > V^M_J(\theta \mid I_{it}). \tag{35}$$

The ratio of Myopic-VPI choice probabilities can be written as

$$\ln\left[\frac{P_i(k' \mid I_{it})}{P_i(k \mid I_{it})}\right] = \ln\left[\frac{\exp(V^{VPI}_{k'}(\theta \mid I_{it}))}{\exp(V^{VPI}_k(\theta \mid I_{it}))}\right]$$
$$= V^{VPI}_{k'}(\theta \mid I_{it}) - V^{VPI}_k(\theta \mid I_{it})$$
$$= \left[V^M_{k'}(\theta \mid I_{it}) - V^M_k(\theta \mid I_{it})\right] + \left[VPI_{ik't} - VPI_{ikt}\right]. \tag{36}$$

We are interested to $\frac{\partial \ln\left[\frac{P_i(k'|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{ijt}}$. Note that

$$\frac{\partial \ln\left[\frac{P_i(k'|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{ijt}} = \frac{\partial\left[V^M_{k'}(\theta \mid I_{it}) - V^M_k(\theta \mid I_{it})\right]}{\partial p_{ijt}} + \frac{\partial\left[VPI_{ik't} - VPI_{ikt}\right]}{\partial p_{ijt}} \tag{37}$$

The first term (37) is identical to (34). However, the second term in (37) will provide us new insights about consumer i's switching pattern respect to j's price variation. For simplicity, we are focusing on price variations which won't change the ordering at (35).

- $k = 1$.

  Note that ζ_{i1t} denotes the threshold that consumer i expects a positive gain by switching her choice if consumption of the best myopic choice ends up at quality lower than ζ_{i1t}. By inverse function theorem, we have $\frac{\partial \zeta_{i1t}}{\partial p_{i1t}} = \frac{w_p}{f'(\zeta_{i1t})} > 0$ since f is an increasing function. So if p_{i1t} increases then ζ_{i1t} will increase. Here, at a higher p_{i1t}, consumer i's expected current utility of consumption $j = 1$ decreased. So even a lower quality signals of other products will incentivize her to switching to attain a positive gain at (20). Therefore, VPI_{i1t}, at Eq(21), will be a truncated integral over a *larger* domain at the lower tail of consumer i's prior belief b_{i1t} about the superior choice. In sum, the uncertainty assessment of the superior product will generate a *higher* expected exploration bonus if p_{i1t} increases. A mirror effect will be accrued if p_{i2t} increases. However, price variation of p_{ijt}, where $j \neq 1, 2$, has no effect on VPI_{i1t}. Precisely, we have

$$\frac{\partial VPI_{i1t}}{\partial p_{ijt}} = \begin{cases} w_p\left(\Phi(\beta_{i1t}) + f(\zeta_{i1t})\left(\frac{\phi(\beta_{i1t})}{\sigma_{i1t} f'(\zeta_{i1t})}\right) + \left(\frac{\phi(\beta_{i1t})\Phi(\beta_{i1t}) - \phi^2(\beta_{i1t})}{f'(\zeta_{i1t})[\Phi(\beta_{i1t})]^2}\right)\right) > 0, & j = 1 \\ -w_p\left(\Phi(\beta_{i1t}) + f(\zeta_{i1t})\left(\frac{\phi(\beta_{i1t})}{\sigma_{i1t} f'(\zeta_{i1t})}\right) + \left(\frac{\phi(\beta_{i1t})\Phi(\beta_{i1t}) - \phi^2(\beta_{i1t})}{f'(\zeta_{i1t})[\Phi(\beta_{i1t})]^2}\right)\right) < 0, & j = 2 \\ 0 & j \neq 1, 2. \end{cases} \tag{38}$$

  This shows that a Myopic-VPI's exploration bonus of the superior myopic choice (i.e., VPI_{i1t}) is *actually* a function of its own price and the price of the second best myopic choice. Intuitively, consumer i's incentive to explore the superior myopic choice is dependent to its relative current position to the second best myopic choice. Whenever, expected current utilities be closer, exploration value will increase. In contrast, a fully forward-looking (resp. index strategy) agent will have a constant exploration bonus for a product regardless of its current expected utility position to its rivals.

- $k \neq 1$.

  Note that ζ_{ikt} denotes the threshold that consumer i expects a positive gain by switching her superior myopic choice if consumption of the kth-rank choice ends up at quality higher than ζ_{ikt}. By inverse function theorem, we have $\frac{\partial \zeta_{ikt}}{\partial p_{ikt}} = \frac{w_p}{f'(\zeta_{ikt})} > 0$ since f is an increasing function. So if p_{ikt} increases then ζ_{ikt} will increase. Here, at a higher p_{ikt}, consumer i's expected current utility of consumption $j = k$ decreased. So much more higher quality signals of product k will be needed to incentivize her to switching away, form superior myopic choice, to attain a positive gain at (20). Therefore, VPI_{ikt}, at (22), will be a truncated integral over a *smaller* domain at the upper tail of consumer i's prior belief b_{ikt} about the kth-rank choice. In sum, the uncertainty assessment of the superior product will generate a *lower* expected exploration bonus if p_{ikt} increases. A mirror effect will be accrued if p_{i1t} increases. However, price variation of p_{ijt}, where $j \neq 1, k$, has no effect on VPI_{ikt}. Precisely, we have

$$\frac{\partial VPI_{ijt}}{\partial p_{ijt}} = \begin{cases} w_p\left(1 - \Phi(\beta_{ikt}) - f(\zeta_{ikt})\left(\frac{\phi(\beta_{ikt})}{\sigma_{ikt} f'(\zeta_{ikt})}\right) - \left(\frac{\phi(\beta_{ikt})\Phi(\beta_{ikt}) - \phi^2(\beta_{ikt})}{f'(\zeta_{ikt})[\Phi(\beta_{ikt})]^2}\right)\right) > 0, & j = 1 \\ -w_p\left(1 - \Phi(\beta_{ikt}) - f(\zeta_{ikt})\left(\frac{\phi(\beta_{ikt})}{\sigma_{ikt} f'(\zeta_{ikt})}\right) - \left(\frac{\phi(\beta_{ikt})\Phi(\beta_{ikt}) - \phi^2(\beta_{ikt})}{f'(\zeta_{ikt})[\Phi(\beta_{ikt})]^2}\right)\right) < 0, & j = k \\ 0 & j \neq 1, k. \end{cases} \tag{39}$$

This shows that a Myopic-VPI's exploration bonus of the inferior kth-rank choice (i.e., VPI_{ikt}) is *actually* a function of its own price and the price of the superior myopic choice. Intuitively, consumer i's incentive to explore the kth-rank choice is dependent to its relative current position to superior choice. Whenever, expected current utilities be closer, exploration value will increase. In sum, the exploration bonus in fully forward-looking or index strategy are irrelevant to price variation at time t; while VPI-exploration bonus, i.e., VPI_{ijt}, is a function of relative expected current utilities which are functions of current prices at time t.

Now we are in a position to revisit $\frac{\partial \ln\left[\frac{P_i(k'|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{ijt}}$. Based on (37) and (38), we have, for all $k \neq 1$,

$$\begin{aligned} \frac{\partial \ln\left[\frac{P_i(1|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{i1t}} = & -w_p \\ & + w_p\left(\Phi(\beta_{i1t}) + f(\zeta_{i1t})\left(\frac{\phi(\beta_{i1t})}{\sigma_{i1t} f'(\zeta_{i1t})}\right) + \left(\frac{\phi(\beta_{i1t})\Phi(\beta_{i1t}) - \phi^2(\beta_{i1t})}{f'(\zeta_{i1t})[\Phi(\beta_{i1t})]^2}\right)\right) \\ & - w_p\left(1 - \Phi(\beta_{ikt}) - f(\zeta_{ikt})\left(\frac{\phi(\beta_{ikt})}{\sigma_{ikt} f'(\zeta_{ikt})}\right) - \left(\frac{\phi(\beta_{ikt})\Phi(\beta_{ikt}) - \phi^2(\beta_{ikt})}{f'(\zeta_{ikt})[\Phi(\beta_{ikt})]^2}\right)\right) \end{aligned} \tag{40}$$

According (40), if the superior myopic choice decreases its price, it will steal market share from all other products; *but the marginal effect on other products market share will be asymmetric in contrast to Myopic or Forward-Looking models.* (i.e., $\frac{\partial \ln\left[\frac{P_i(1|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{i1t}} \neq \frac{\partial \ln\left[\frac{P_i(1|I_{it})}{P_i(k'|I_{it})}\right]}{\partial p_{i1t}}$ when $k \neq k'$). This provides a more realistic pattern as documented in marketing literature Moreover, the IIA restriction will be relaxed by applying Myopic-VPI descriptive model. Note that we have, for all $k', k \neq 1$,

$$\begin{aligned}\frac{\partial \ln\left[\frac{P_i(k'|I_{it})}{P_i(k|I_{it})}\right]}{\partial p_{i1t}} &= w_p\left(1 - \Phi(\beta_{ik't}) - f(\zeta_{ik't})\left(\frac{\phi(\beta_{ik't})}{\sigma_{ik't} f'(\zeta_{ik't})}\right) - \left(\frac{\phi(\beta_{ik't})\Phi(\beta_{ikt}) - \phi^2(\beta_{ik't})}{f'(\zeta_{ik't})[\Phi(\beta_{ik't})]^2}\right)\right) \\ &- w_p\left(1 - \Phi(\beta_{ikt}) - f(\zeta_{ikt})\left(\frac{\phi(\beta_{ikt})}{\sigma_{ikt} f'(\zeta_{ikt})}\right) - \left(\frac{\phi(\beta_{ikt})\Phi(\beta_{ikt}) - \phi^2(\beta_{ikt})}{f'(\zeta_{ikt})[\Phi(\beta_{ikt})]^2}\right)\right) \neq 0 \end{aligned} \tag{41}$$

Basically, (41) says, any price variation by product will have effect on ratio of market share of other products. This generates a more flexible and realistic pattern. In sum, we summarize the above observations as follows

Proposition 1 *Let's assume consumer i information state at time t is $I_{it} = \{(Q_{ijt}, \sigma^2_{ijt})\}_{j=1}^J$ which denotes the prior belief of consumer i at time t toward each product. Consumer i also receives exogenous price p_{ijt} at time t for product $j \in \{1, \cdots J\}$. The $p_{ijt} \sim N(\overline{p_j}, \sigma^2_{p_j})$ which are i.i.d over time. Here, $\overline{p_j}$ and $\sigma^2_{p_j}$ denote the average price of product j and the variance of price of product j.*

1. *If consumers behave "as if" Myopic or Fully Forward-Looking, the absolute marginal effect of product j's price variation, on ratio of market shares at time t, is symmetric. Also, the Independence of Irrelevant Alternatives (I.I.A) property holds.*

2. *If consumers behave "as if" Myopic-VPI, the absolute marginal effect of product j's price variation, on ratio of market shares at time t, is asymmetric. Also, the Independence of Irrelevant Alternatives (I.I.A) property does not hold. More precisely, a price variation by the superior myopic choice will have a more extensive effect on market shares than inferior choices.*

Proposition 1 shows our VPI approach is capable to fit into asymmetric switching pattern in consumers' purchase behavior in scanner datasets. This means the price elasticity should be estimated more accurately by considering VPI approach. Therefore, form a managerial point of view, VPI approach provides more realistic estimation of demand parameters by regarding (1) consumers' bounded rationality in purchase behavior and (2) well-documented asymmetric switching pattern. We will provide evidence through simulations to show that the Myopic-VPI can perform as well as Index Strategy in terms of accumulated payoffs, and yet it is much more superior in terms of computational time (which was assumed as the surrogate of cognitive tractability).

6 Simulated Data Performance

First, we provide evidence that Myopic-VPI is a "fast and frugal" way to balance exploitation and exploration incentives (Gigerenzer and Goldstein (1996)). Our result shows that the VPI approach provides a significant improvement in terms of long-run payoffs with slightly more computational time relative to myopic approach; while the improvement level of long-run payoffs by following the near to optimal approach will be negligible if computational time is taken into account. This is in the line of Kenneth j. Arrow' comment: "boundedly rational procedures are in fact fully optimal procedures when one takes into account the cost of computation" (Arrow (2004)).

We demonstrate it in two dimensions: (1) Accumulated Long-Run Payoffs and (2) Performance Time. When simulate the performance, we use identical signals and shocks. Here, the goal is to demonstrate that Myopic-VPI can achieve a reliable long-run payoffs regarding its cost of performance. As in footnote 6, the computational time is regarded as a surrogate for cost of thinking.

We design our simulation in terms of introduction of a new high quality product in the market. If consumers do exploration, they will be able to figure out the advantage of consuming the new alternative. We simulate the models for $T = 50$ periods. We set the discount factor, $\beta = 0.9$, and choose $J = 2$ for simplicity. We use O to denote an old brand, and N to denote a new brand, and assume the "true" qualities of O and N be 1 and 2, respectively. We normalize the "experience variability" of both products to 1. The initial prior belief standard deviation for brands O and N are set to $\sqrt{1/10}$ and 1, respectively. This set of parameters can be interpreted as an introduction of a new high quality product N to market. Since it is new, consumers are more uncertain about its quality than the older brand O.

Consumers observes retail prices. We assume that the price of older brand follows a binomial distribution wherein the regular price is 0.6 with probability 4/5. There is a promotional price 0.3 with probability 1/5. This is a common pricing strategy among retailers in reality.

We assume that the new product N are going to use the *penetration strategy* for its pricing. For the first 10 weeks, the regular price is 1 with probability 4/5, and the promotional price is 0.5 with probability 1/5. For other weeks, the regular price is 1.5 with probability 4/5, and the promotional price 1 with probability 1/5. We set the price sensitivity coefficient, w_p, be equal to 0.15 which is an average of estimations respect to our filed data result in diaper category. Finally, the measurement error is distributed as extreme value Type I with mean 0 and standard deviation 0.05. For each experiment, we run it for 10000 times.

In all experiments, we vary the consumer's prior mean values to capture three possibilities in the market: (a) a pessimistic consumers' prior belief towards the higher quality product, (b) an indifferent consumers' prior belief about the true quality, and (c) an optimistic consumers' prior belief towards the higher quality product. We fix the prior mean quality belief of brand O as $Q_{O0} = 0.8$ and vary the prior mean quality belief for brand N as $Q_{N0} \in \{0.6, 0.8, 1\}$. The prior belief of brand O is chosen close to its true quality since it has been in the market for a while.
The case $Q_{N0} = 0.6$ should be interpreted as a scenario that consumers are pessimistic toward the new product N. That could happen because of very weak marketing campaigns or a poor packaging of N which failed to inform consumers about its advantages. The case $Q_{N0} = 0.8$ can be interpreted as a scenario that consumers transfer their belief on quality of O to the new product N's quality. Finally, the case $Q_{N0} = 1$ can be interpreted as a scenario that consumers are optimistic toward the new product N. That could happen if the brand carries out a strong marketing campaign.

We did our simulation by assuming consumers are either risk neutral or risk averse. People show risk-aversion specially in an uncertain environment. Simply, a risk-averse consumer prefers to choose more familiar brand rather than exploring new brands in the market. The learning dynamic literature finds evidence of risk-aversion in purchasing frequently experience goods (Erdem and Keane (1996), Ching et al. (2014)). To parametrize our simulation, we assume that consumers exhibit CARA utility: $u_{ijt}(Q^E_{ijt}) = 1 - \exp(r.Q^E_{ijt}) - w_p p_{ijt} + \epsilon_{ijt}$.[20] This is the most common functional form have been used in learning dynamics (Crawford and Shum (2005), Ching et al. (2014)).

Let's compare four descriptive models to make purchase decision in the market: (1) No-Learning, (2) Myopic, (3) Myopic-VPI, and (4) Index Strategy respectively.[21] These set-

[20] Clearly, the expected utility at (8) is strictly decreasing in σ_{ijt}. This captures the consumer i's preference to avoid exploration of more uncertain alternatives.

[21] In No-Learning model, a consumer always myopically makes decision only based on her prior belief at

tings are in favor of Index Strategy.[22] Clearly, purchasing the new product provides higher expected payoffs. Because forward-looking models do exploration, they should generate a higher long-run payoffs. The consumer's accumulated rewards, through 50 periods respect to each descriptive model, are presented in Table 2.

Table 2: The Average Performance of Descriptive Models

	The expected accumulated reward (standard errors in parentheses)			
	No Learning	Myopic Learning	Myopic-VPI Strategy	Index Strategy
Computation time for per decision (surrogate for cognitive complexity)[a]	$\approx 10^{-7}$	$\approx 10^{-4}$	$\approx 10^{-3}$	≈ 10
The Risk Neutral Utility		$r = 0$		
Mean of prior quality belief (brand 1, brand 2)				
$(Q_{O0}, Q_{N0}) = (0.8, 0.6)$	11.374	14.552	17.059	17.952
	(0.025)	(0.034)	(0.299)	(0.0277)
$(Q_{O0}, Q_{N0}) = (0.8, 0.8)$	13.833	16.673	17.739	18.104
	(0.0256)	(0.0311)	(0.0285)	(0.026)
$(Q_{O0}, Q_{N0}) = (0.8, 1)$	16.261	17.589	18.008	18.169
	(0.0247)	(0.0296)	(0.0278)	(0.025)
The Risk Aversion Coefficient		$r = 0.9$		
Mean of prior quality belief (brand 1, brand 2)				
$(Q_{O0}, Q_{N0}) = (0.8, 0.6)$	3.2095	3.4134	4.7041	5.8708
	(0.0168)	(0.0167)	(0.0149)	(0.011)
$(Q_{O0}, Q_{N0}) = (0.8, 0.8)$	3.6758	4.3658	5.4682	6.0101
	(0.0160)	(0.0161)	(0.014)	(0.0102)
$(Q_{O0}, Q_{N0}) = (0.8, 1)$	4.4666	5.2804	5.8239	6.0323
	(0.0153)	(0.0143)	(0.0120)	(0.009)
The Risk Aversion Coefficient		$r = 1.2$		
Mean of prior quality belief (brand 1, brand 2)				
$(Q_{O0}, Q_{N0}) = (0.8, 0.6)$	2.9588	2.9588	3.1431	3.3314
	(0.026)	(0.026)	(0.025)	(0.024)
$(Q_{O0}, Q_{N0}) = (0.8, 0.8)$	2.9807	3.0328	4.3727	5.5682
	(0.026)	(0.025)	(0.023)	(0.019)
$(Q_{O0}, Q_{N0}) = (0.8, 1)$	3.3431	3.9055	5.7782	6.3995
	(0.025)	(0.023)	(0.0182)	(0.015)

[a] These are seconds required to make a decision at each time period respect to the consumer's prior belief using a Dell-PC with a processor Intel(R) Core(TM)i7-3770 CPU.

For each decision process, we evaluate the computational time to make a decision at an information state at a time step t. In a No-Learning model, a consumer only needs to find her utility of each alternative given her prior belief at $t = 1$. A myopic consumer should also spend time to update her prior belief at time t after receiving signals at time $t-1$. In forward-looking models, Myopic-VPI and Index Strategy, consumers need to spend additional time to evaluate $\{VPI_{ijt}\}_{j \in A}$ and $\{W_{ijt}\}_{j \in A}$ for each alternative before making purchase decision.

time $t = 1$. She does not even update her beliefs over time.

[22] Note that Proposition 2, at page 8 in Lin et al. (2015), requires location and scale parameter distributions and risk-neutrality. Otherwise, researchers need either to evaluate indexes at each information state or the interpolation will be as a proxy for the Whittle Index. In Myopic-VPI, we are not facing these restrictions.

These computational times can capture the cognitive complexity of descriptive models. The computational time per decision for No-learning and Myopic are negligible. The computational time per decision for Myopic-VPI are slightly higher than Myopic model; while it is significantly faster than Index Strategy.[23] In sum, a Myopic-VPI consumer needs to pay higher cost of thinking to enjoy the exploration value. But cognitive complexity of Index strategy (or approximately optimal) is much higher than that of Myopic-VPI.

First, we are looking at the risk neutral case. Table 2 shows that learning is useful since the performance of No-Learning strategy is significantly lower than other decision-making processes. For example, a pessimistic risk neutral consumer gains 26% less if she ignores learning process (compare No-Learning with Myopic Learning). Also, it is clear that a myopic risk neural consumer gains a lower accumulated reward since she did not consider the value of exploration at all. The losses are even more serious if consumers have more negative prior attitude towards high quality product. For example, when prior belief is pessimistic, the accumulated reward for Myopic-VPI is 12.7% better than Myopic behavior. Under Myopic-VPI and Index Strategy, consumers consider exploring the new brand more frequently. Similar arguments apply when consumers have more optimistic initial prior towards the new brand. But note that optimistic consumers have a more efficient heuristic clue such that No-learning or Myopic behaviors do not cause significant losses. *So if the prior belief is directed correctly towards the high quality product, learning value will decrease.*

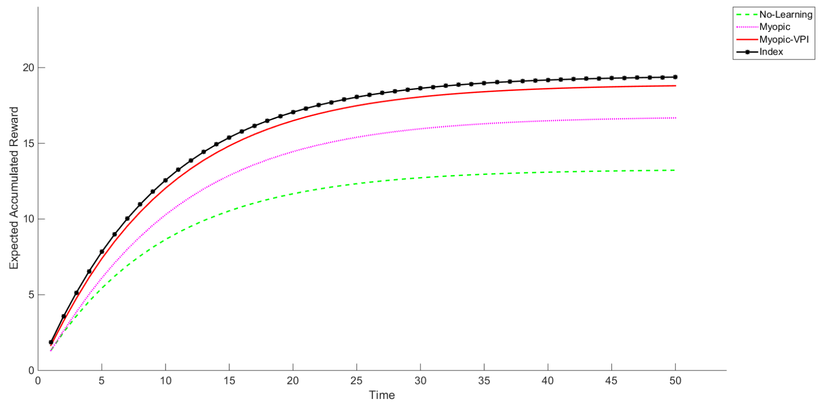
When prior belief is pessimistic (i.e., $Q_{0N} = 0.6$), the Index Strategy is 5% better than the Myopic-VPI. However, if we consider the computational time per decision, the loss under Myopic-VPI is not really significant. The Index Strategy is only 5% better than Myopic-VPI; while she needs to spend a massive cognitive cost in Index strategy compare to Myopic-VPI (i.e., 10^4 more seconds at each time step). So, Consumers can make a "fast and frugal" decision by using Myopic-VPI. Therefore, we could predict that Myopic-VPI can be more tractable by consumers. Under other scenarios, the performance of Myopic-VPI is as good as Index Strategy. Regarding the cost of thinking to make a decision, the Myopic-VPI is again more tractable than Index Strategy. Figure 1 depicts the performance of all decision processes over time. *In summary, Table 2 provides evidence that the Myopic-VPI is more compatible with bounded rationality of consumers in the market.*

[23]Lin et al. (2015) provides a clever interpolation technique in Index Strategy to evaluate the Whittle indexes. However, researchers have to assume more assumptions on observed, unobserved shock distributions, and utility functional form as in footnote 22. Otherwise, the Index Strategy computational burden increases significantly. But there does not exist such a restriction in Myopic-VPI.

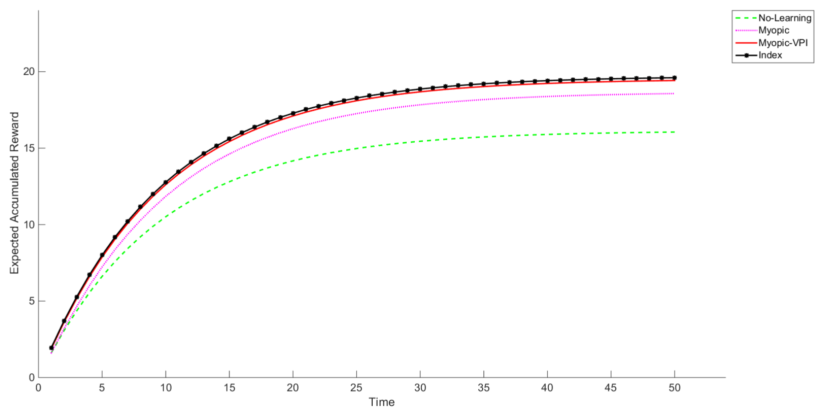
Figure 2 depicts the choice probabilities of choosing high quality product over time. The choice probabilities of higher quality brand is always higher under Myopic-VPI and Index Strategy than Myopic behavior. Consumers are more uncertain about the new brand. Since the myopic consumer does not consider the exploration benefits, for her, the choice probability of higher quality product (i.e., new brand) is lower. We can see that, under Index Strategy, consumers prefer to choice the new product more than Myopic-VPI at initial purchase periods. But the choice probability of higher quality brand, under Myopic-VPI, increases faster than the choice probability under Index Strategy over time. Note that, as time passes, the choice probabilities converge and become stable. This is because as consumers learn more about a product, its value of exploration declines. Under Index Strategy, the incentive to try the high quality brand is slightly higher (reflected in its slightly higher choice probabilities), consumers therefore learn the benefits of new brand faster. Consequently, the Index Strategy choice probabilities converges slightly faster than Myopic-VPI. However, it should be noted that the Myopic-VPI rule is much easier to compute. Given that the gain of using Index Strategy is very small, Myopic-VPI can be a more realistic heuristic approach. The choice probabilities under Myopic-VPI and Index Strategy are almost similar to each other.

Although our results are robust when consumers are risk averse, there are two interesting points to emphasize here. **First**, if consumers are seriously risk-averse (i.e., $r = 1.2$), the value of exploration will be negligible in comparison to disutility of experiencing risky alternatives. This provides a significant negative attitude towards exploration. The above effect is even more pronounce if consumers have negative prior belief to unfamiliar product. Therefore, Myopic-VPI or Index Strategy are almost determining identical choices as in No-Learning or Myopic models (Figure 4). **Secondly**, when exploration is not disincentivized because of consumers' negative attitude to risk or new product, exploration is much more important. The difference between the performance of Mypoic consumer with Myopic-VPI and Index Strategy is significantly larger than risk neutral case (even if consumers be optimistic). For example, Myopic-VPI and Index consumers respectively achieve 38% and 70% higher accumulated reward relative to their Myopic counterpart, when prior belief is pessimistic at $r = 0.9$. Index strategy endures 10^4 times computational time to achieve the 25% additional improvement of accumulated reward relative to Myopic-VPI; while 38% Myopic-VPI improvement will be achieved only by 10 times computational time. This is consistent with experimental evidences in the literature that consumers prefer to avoid behave myopically; while their choice behavior is faraway from Index Strategy (or the optimal solution). When Myopic consumer is risk averse, she prefers the low quality brand more since she won't want to take the risk to try uncertain new product. But Myopic-VPI and Index Strategy are able to determine the benefit of exploration which increases the accumulated rewards over time.

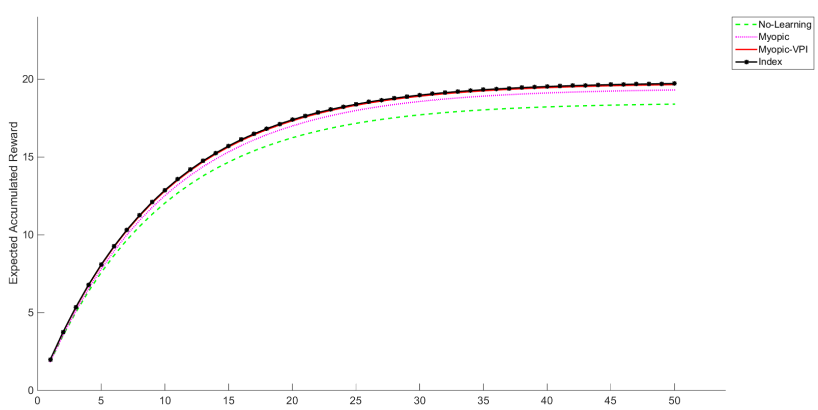
Figure 3 and 4 depict the choice probabilities of choosing high quality product over time at $r = 0.9$ and $r = 1.2$ respectively. Following figures depict and results, regarding cost of thinking, Myopic-VPI is able to balance the tradeoff of exploitation vs exploration better than Index Strategy when the agent is risk averse. Also, our result shows that the exploration is much more valuable for risk averse consumers. Here, even myopic consumers, who believe that new product should be better, they still prefer to stick with the old product to avoid taking the risk of purchasing an uncertain product. On the other hand, forward-looking descriptive models will help consumers to consider the exploration value of uncertain products in their decision process. *In summary, we can conclude that if consumers are more risk averse: (1) exploration is much more valuable to prevent losses in a sequential choice model, and (2) Myopic-VPI is a better heuristic way to make a decision regarding cognitive complexity.*



(a) The Pessimistic Consumer

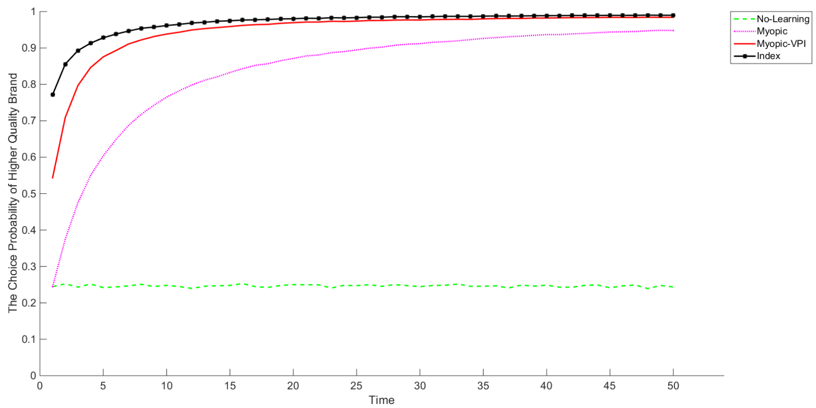


(b) The Indifferent Consumer

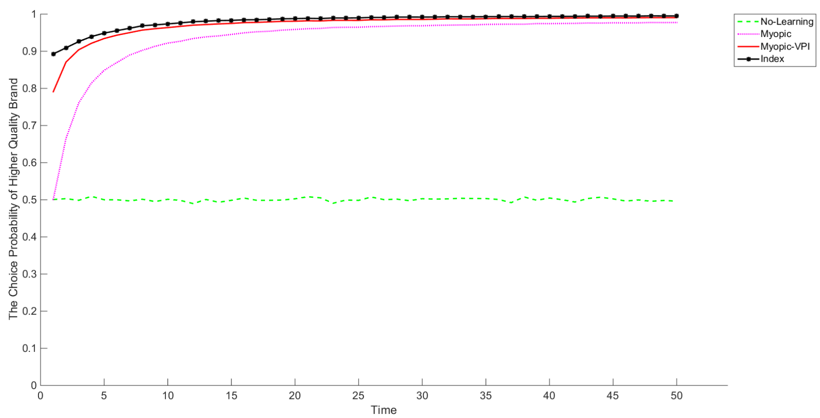


(c) The Optimistic Consumer

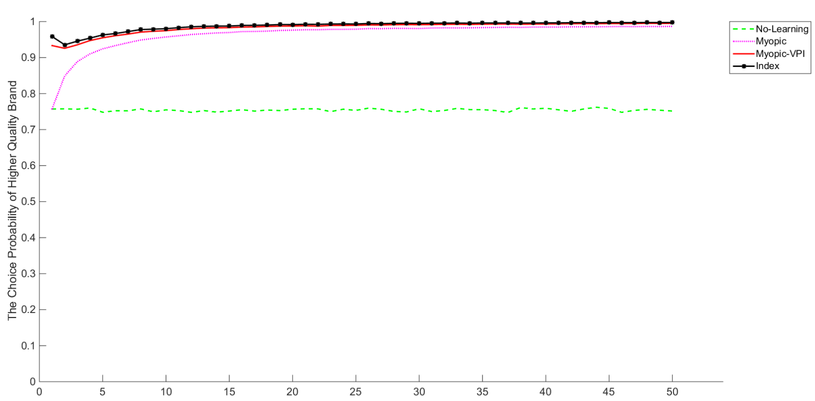
Figure 1: The Expected Acumulated Reward: Risk Neutral



(a) The Pessimistic Consumer

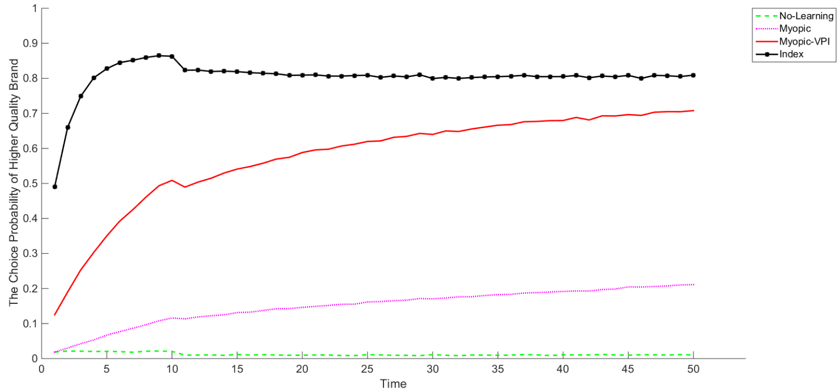


(b) The Indifferent Consumer

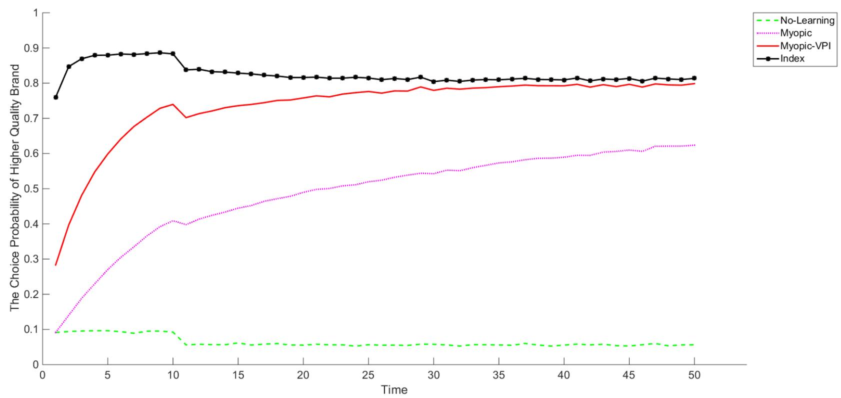


(c) The Optimistic Consumer

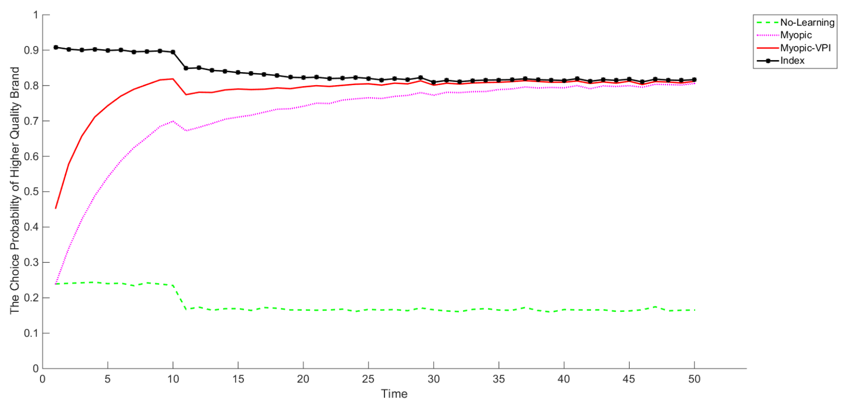
Figure 2: The Choice Probability of Higher Quality Brand: Risk Neural



(a) The Pessimistic Consumer

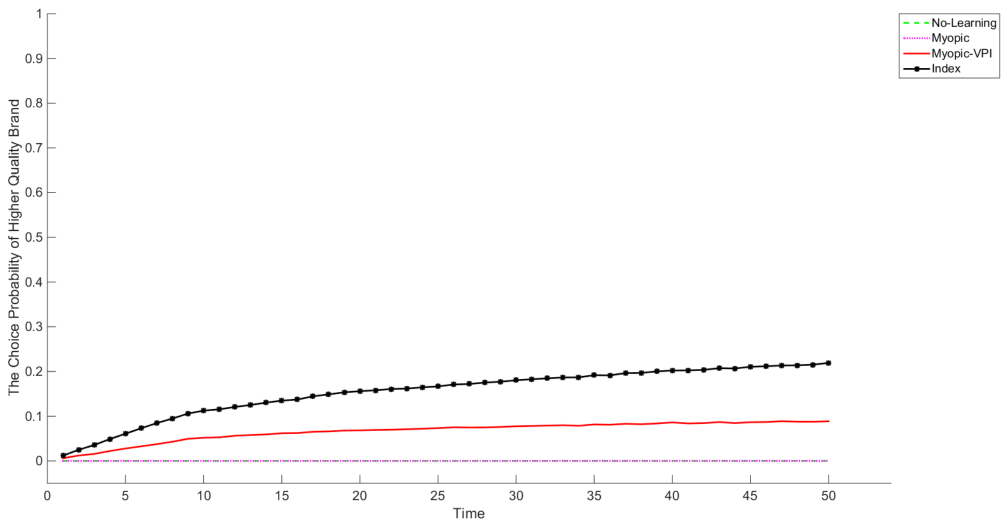


(b) The Indifferent Consumer

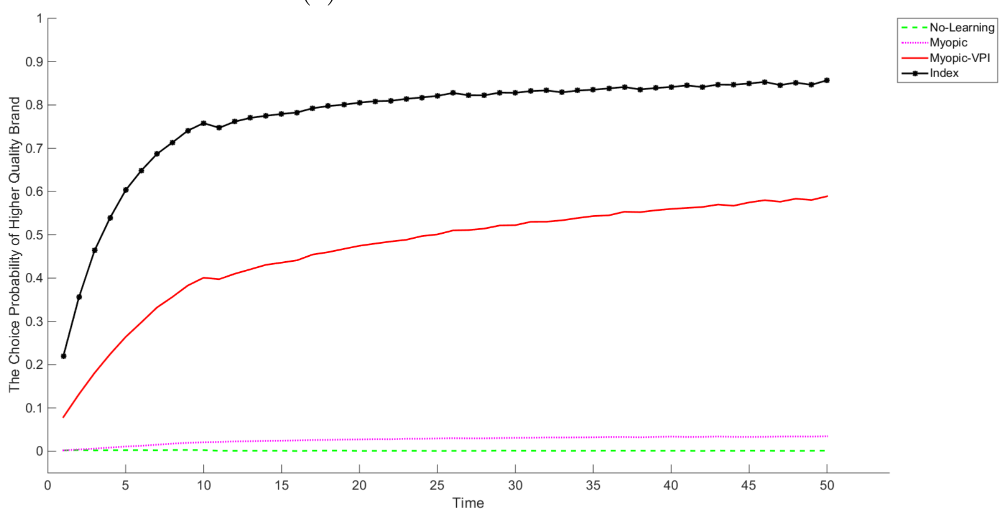


(c) The Optimistic Consumer

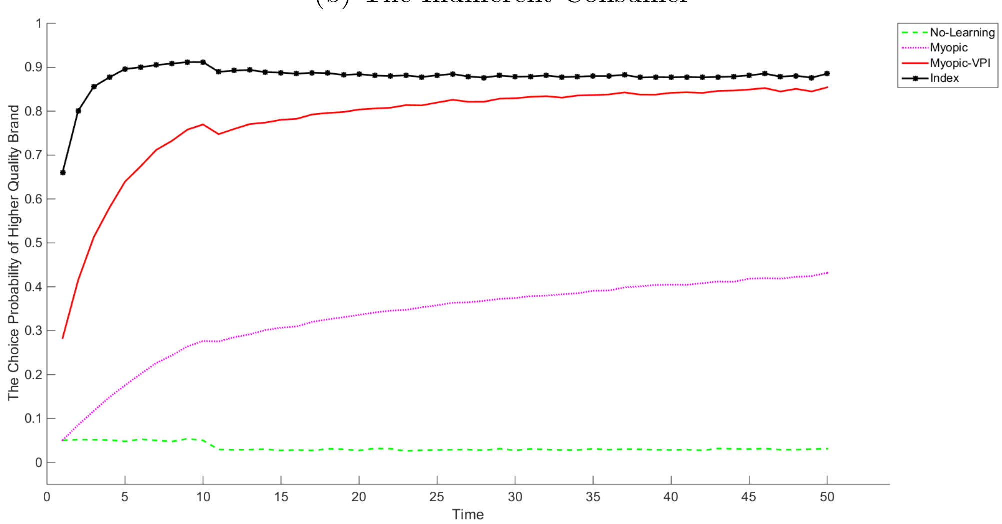
Figure 3: The Choice Probability of Higher Quality Brand: Risk Averse at $r = 0.9$



(a) The Pessimistic Consumer



(b) The Indifferent Consumer



(c) The Optimistic Consumer

Figure 4: The Choice Probability of Higher Quality Brand: Risk Averse at $r = 1.2$

7 Field Data Estimation

From a theoretical perspective, we show that VPI approach provides new insights toward forward-looking behavior. It generates a more cognitive tractable “as if” choice model which performs as a reliable fast and frugal decision process in simulation settings. Now, we examine how a Myopic-VPI estimation fits and predicts behaviors compared with a myopic learning and index strategy.[24] Here, we consider product category and sample where consumers are likely to be forward looking. Note that if a Myopic-VPI does no better than an approximately optimal solution and Index Strategy, we consider the result promising because a Myopic-VPI solution is cognitively simpler and less restrictive to functional forms of utilities and distributions. We expect Myopic-VPI learning strategy to outperform no-learning strategies and, because we focus on a situation that favors forward-looking behavior, we expect forward-looking aspect of Myopic-VPI strategy to outperform Myopic learning.

Our data is based on the diaper category from the IRI Marketing Data Set for academic research (Bronnenberg et al. (2008)), and our sample data set is the same as the one used in Lin et al. (2015).[25] One advantage of using the diaper category is that parents typically begin purchasing diapers based on a discrete birth event, and their entry to the category is arguably exogenous (Ching et al. (2014)). Ching et al. (2014) find that diaper consumers conduct strategic trials of various brands. There are observable shocks due to price promotions and shocks due to unobservable events. Lin et al. (2015) wrote “even if the birth is a second or subsequent child, diaper quality may have changed. Informal qualitative interviews suggest that parents learn about whether diaper brands match their needs through experience (with often more than one purchase), that diapers are sufficiently important that parents take learning seriously, and that parents often try multiple brands before settling on a favorite brand.”

Lin et al. (2015) selected the sample based on three criteria: First, households are selected only if their first purchase occurs 30 weeks after the start of data collection (73% of the entire sample); such a sample likely contains first-time buyers. Second, households need to make at least five purchases during the observation. So, such a sample likely contains regular

[24]Lin et al. (2015) show that the index strategy and near optimal method provide similar goodness-of-fit and estimation results on this data set. Since the estimation of the model with near to optimal method is time consuming, we only do our comparison with No-Learning, Myopic, and Index Strategy to show that Myopic-VPI is able to capture the forward-looking behavior by fitting into data in a better way. However, by using simulation, we already show that the choice probabilities under Myopic-VPI and Index Strategy are significantly close to each other.

[25]We would like to thank Song Lin for sharing the data set used in Lin et al. (2015) with us.

Table 3: Descriptive Statistics of Sample

Brands	**Pampers**	**Huggies**	**Luvs**	**Other Brands**
Number of Purchase	547	608	347	36
Market Share	0.3557	0.3953	0.2256	0.0234
Price mean	12.43	12.31	10.44	8.55
Price Standard Deviation	1.80	1.61	1.99	1.99

buyers. Third, households who bought private label are excluded. Private label buyers may not care about quality much, and hence also not likely to learn. The sample data consists of 131 households, and their choices are dominated by Pampers, Huggies, and Luvs. We aggregate all other branded purchases as "other brands" and do not model the no-purchase option. Here, we focus on consumers' choice behavior and exploration-exploitation tradeoffs rather than purchase incidence. So, we model consumers' choice behavior at each purchase occasion. We observe, on average, 13 purchase occasions per consumer.

The Table 3 shows a summary statistics of our sample. The Huggies has the largest market share. Pampers has the the second largest market share with 4% less than Huggies. However, the average prices of Huggies and Pampers are quit similar. Luvs has the third market share; and its price, on average, \$2 lower than the market leaders. The other brands only have 2.3% market share. Their price are significantly lower than Huggies, Pampers, and Luvs. Figure 5 depicts the percentage of consumers who switch to another brand at purchase occasion t form their choices at $t-1$. We observe that the switching pattern is decreasing over time. This provides a consistent evidence about our exploration-exploitation story. Consumers did more switching at initial weeks when they entered into diaper market. Since they were uncertain about quality of each brand, consumers did exploration to learn about true qualities of different brands. When they resolved their uncertainty about true quality of brands, they started to repurchase their favorite brands more frequently. Also, in Table 4, we show consumers' switching matrices from purchase occasion 1 to 2, purchase occasion 5 to 6, and from week 10 to 11 respectively. By comparing tables 4a, 4b, and 4c, we observe the decreasing and increasing patterns among off-diagonal and on-diagonal entries. This provides further evidence of exploration-exploitation in our sample.

However, it appears that consumers are heterogeneous in their "brand loyalty." Figure 6 provides some examples. Figure 6a shows a consumer who always bought the same brand

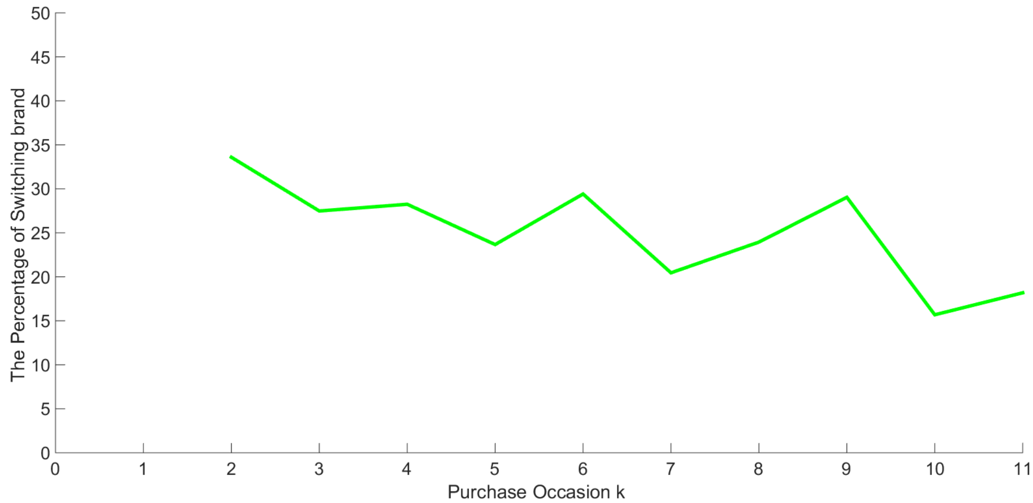


Figure 5: Consumers' Switching Pattern Over Purchase Occasions

Table 4: The Switching Pattern Matrix at Purchase Occasion k

Brand	Pampers	Huggies	Luvs	Other Brands
Pampers	0.2443	0.0916	0.0458	0
Hugggies	0.0687	0.2137	0.0229	0.0076
Luvs	0.0305	0.0305	0.1832	0.0153
Other Brands	0.0229	0	0	0.0229

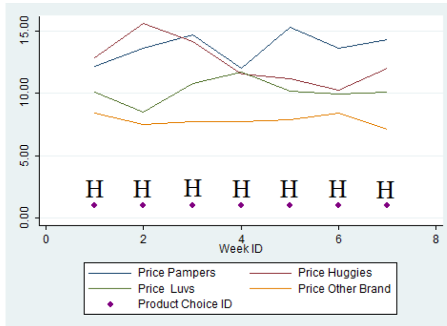
(a) Switching Matrix form purchase occasion **1** to purchase occasion **2**

Brand	Pampers	Huggies	Luvs	Other Brands
Pampers	0.3039	0.049	0.0588	0.0098
Hugggies	0.0588	0.2451	0.0392	0
Luvs	0.0098	0.0588	0.1373	0
Other Brands	0.0098	0	0	0.0196

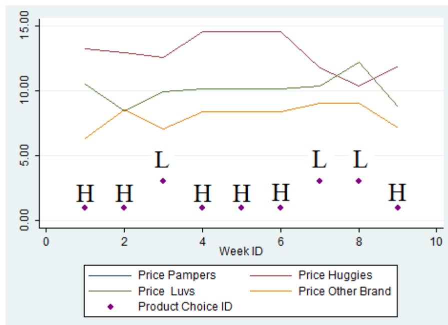
(b) Switching Matrix form purchase occasion **5** to purchase occasion **6**

Brand	Pampers	Huggies	Luvs	Other Brands
Pampers	0.2955	0.0455	0.0227	0
Hugggies	0.0455	0.3409	0.0227	0
Luvs	0.0455	0	0.1818	0
Other Brands	0	0	0	0

(c) Switching Matrix form purchase occasion **10** to purchase occasion **11**



(a) High Brand Loyalty



(b) Low Brand Loyalty

Figure 6: Heterogeneity in Brand Loyalty

throughout our sample. Figure 6b shows a consumer who did exploration-exploration by switching among brands. To be able to address the brand loyalty in our empirical implementation, the expected current utility of consumer i from choosing brand j at time t is written as,

$$V_j^M(\theta \mid I_{it}) := -exp\left(-r\left(Q_{ijt} - \frac{r}{2}\left(\sigma_{ijt}^2 + \sigma_\xi^2\right)\right)\right) - w_p p_{ijt} + \lambda GL(H_{ijt}, \gamma),$$

where

$$GL(H_{ijt}) = \gamma GL(H_{ijt-1}) + (1-\gamma)d_{ijt}. \tag{42}$$

The term $GL(H_{ijt}, \gamma)$ was defined by Guadagni and Little (1983) to capture the idea that a consumer who bought a brand frequently in the past is likely to buy it again. This term enables us to address the brand loyalty in our estimation. Here, H_{ijt} is consumer i's purchase history for brand j prior to time t, and γ is the exponential smoothing parameter; λ is the coefficient for GL.

Table 5 summarizes the fit statistics for the 1,538 diaper purchases in the in-sample estimation. To measure the goodness-of-fit, we report AIC and BIC here. The "No-Learning Without Brand Loyalty" and "Myopic Learning Without Brand Loyalty" denote the models wherein $\lambda = 0$. Adding the GL increases significantly the Log likelihood in "No-Learning With Brand Loyalty". However, it is clear that "Myopic Learning Without Brand Loyalty" improves the goodness-of-fit relative to "No-Learning With Brand Loyalty". This shows that consumers' learning process plays a significant role to explain part of the state dependency in data. Also, adding the GL increases the fitness of "Myopic Learning With Brand Loyalty" relative to "Myopic Learning Without Brand Loyalty". This shows that it is important to control simultaneously for learning and brand loyalty to provide a precise comparison among models.

It is clear that "Myopic-VPI" and "Index Strategy" fit the data better than "Myopic" and "No learning". This suggests that consumers are forward-looking and consider the tradeoffs between exploration vs exploitation. Note that the improvement of log likelihood of forward-looking models are few units better than myopic model. However, "Myopic-VPI" and "Index Strategy" beat "Myopic" model in terms of AIC and BIC. It is worthy to notice that this level of improvement is consistent with literature. For example, Erdem and Keane (1996), Ackerberg (2003), and Ching et al. (2014) find similar results. According to AIC, there are only 0.135 and 0.06 chance that "Myopic Learning With Brand Loyalty" minimize the information loss relative to the "Myopic-VPI" and "Index Strategy" respectively. This shows a significant evidence that consumers involve active learning in the market by doing strategic exploration rather than only passive learning by updating beliefs.

Moreover, "Myopic-VPI" and "Index Strategy" produce similar goodness-of-fit. To provide a sense of how much computing Myopic-VPI is faster than Index Strategy, we note that the times required to compute one likelihood function for Myopic-VPI and Index Strategy are 2 sec and 12 sec per likelihood, respectively. Our estimation results suggest that the Myopic-VPI model is at least as plausible as the Index Strategy model (i.e., near optimal) in capturing consumer behavior in balancing exploration vs. exploitation. From the practical viewpoint, Myopic-VPI reduces the computational burden significantly. In addition, from the behavioral viewpoint, Myopic-VPI should be more cognitively tractable than the Index Strategy or near optimal models. Note that Myopic-VPI is more compatible with experimental evidence on exploration-exploitation trade-off; while the Index strategy have been rejected by those experiments. Thus, we argue that Myopic-VPI can be considered as a more plausible "as if" forward-looking model to explain the exploration-exploitation trade-off.

Furthermore, the index strategy is applicable only if the decision making problem is proved to be indexable. Lin et al. (2015) shows indexability of brand choice models by adding new assumption. The other advantage of Myopic-VPI approach is its capability to be implemented in any multi-armed bandit process. Without using the index strategy, the computational time to estimate near to optimal model by full dynamic solution will increase to 65 sec per likelihood if we choose only 5 points to discretize the state spaces. If we increase the number of discretization points to 20, the estimation process will be intractable because of memory constraint. Therefore, this shows Myopic-VPI will be extremely more practicable if the index strategy is not applicable.

We also consider more flexible versions of Myopic-VPI by allowing the utility weight on VPI to be a function of $GL(H_{ijt}, \gamma)$. That is, we model $V_{ijt}^{VPI} = V_{ijt}^{M} + \zeta.GL(H_{ijt}, \gamma).VPI_{ijt}$. In Table 5, we report its results under "Myopic-VPI Symmetric". This specification performs better than the other models in terms of goodness-of-fit. As we can see in Table 6, the estimate of ζ is positive, it shows that if $GL(H_{ijt}, \gamma)$ increases, the VPI_{ijt} will have more significant effect on consumers' choice behavior. On the other hand, the VPI approach imposes an asymmetric learning process. The VPI of the superior myopic choice will be driven from types of information those are lower than a threshold; but the VPI of other inferior myopic choice will be driven from types of information those are higher than a threshold. To address this asymmetric feature, we consider a specification wherein $V_{ijt}^{VPI} = V_{ijt}^{M} + \zeta_{jt}.GL(H_{ijt}, \gamma).VPI_{ijt}$ and call it "Myopic-VPI Asymmetric". Here, $\zeta_{jt} = \zeta_1$ if brand j is the superior myopic choice at time t; otherwise $\zeta_{jt} = \zeta_0$. It turns out that "Myopic-VPI Asymmetric" provides the best Log-Likelihood, but its goodness-of-fit is worse than the symmetric version in terms of AIC and BIC. According to AIC, there is only 0.02 chance that "Index Strategy" minimizes the information loss relative to the "Myopic-VPI Symmetric" and "Myopic-VPI Asymmetric".

In sum, Our results show that Myopic-VPI can be considered as one viable way to explain the exploration-exploitation trade-off such that it (1) reduces significantly the computational burden to make it more plausible for practitioners and (2) provides a more cognitively simple decision process which is compatible with experimental evidence of consumers' bounded rationality.

Table 6 summarizes the estimated parameter values. As expected, the price sensitivity coefficient is negative in every model. As in Lin et al. (2015),[26] the standard errors in Index

[26]Note our model is slightly different with Lin et al. (2015). First of all, our utility functional form is the standard CARA form which have been commonly used in dynamic learning literature. Lin et al. (2015) considered $u(w) = 1 - \exp(w)$ where $w = Q_{ijt} - w_p p_{ijt} + \epsilon_{ijt}$. They did not find consumers are risk averse. However, our functional form shows consumers are significantly risk averse in this data set which is a consistent result as in Erdem and Keane (1996) and Ching et al. (2014). Also, we add the $GL(H_{ijt}, \gamma)$

Table 5: In-Sample and Out-of-Sample Fit Statistics for Diaper Data

	No Learning Without Brand Loyalty	No Learning With Brand Loyalty	Myopic Learning Without Brand Loyalty	Myopic Learning With Brand Loyalty	Index Strategy	Myopic-VPI	Myopic-VPI Symmetric	Myopic-VPI Asymmetric
Calibration sample								
Log likelihood	-1759.8	-1040.7	-1024.9	-1015.9	-1013.1	-1013.9	-1008.2	-1007.7
AIC	3551.6	2093.4	2081.8	2067.8	2062.2	2063.8	2054.4	2055.4
BIC	3637.01	2125.42	2167.21	2163.88	2158.2	2159.8	2155.8	2162.16
No. of parameters	4	6	16	18	18	18	19	20
No. of Observation	1,538	1,538	1,538	1,538	1,538	1,538	1,538	1,538
Computation time in seconds	Negligible	0.02(s)	0.14(s)	0.15(s)	12(s)	2(s)	2.2(s)	2.3(s)

Strategy are large and that most parameter estimates not significant; in contrast, this is not the case in Myopic-VPI approaches. Across all models, λ is positive and statistically significant. Thus, consumers show a brand loyalty in their purchase behavior. Consumers initially believe that Huggies is slightly better than Pampers. Meanwhile, they believe Luvs should have a lower quality than Huggies and Pampers; while they treated other brands as the worst product. Across forward-looking models, all four brands increase in mean quality relative to prior beliefs, which implies that diaper buyers learn to appreciate these brands more through experiences. Learning models identify the magnitude of utility of the inherent quality uncertainty when the Pampers parameter normalized to 1 for identification. Across all learning models, Huggies show has the highest inherent quality uncertainty; while Luvs shows a more consistent inherent quality than Pampers. Interestingly, the other brands has the lowest inherent quality uncertainty. Finally, Myopic-VPI models show that consumers are more risk averse than Index Strategy; but the standard deviation of prior beliefs are lower in Myopic-VPI models. Therefore, the overall dis-utility because of taking risk to explore a more unfamiliar product are similar in Myopic-VPI and Index Strategy.

More interestingly, the flexible versions of Myopic-VPI approach give us new behavioral insights toward consumers' learning process. As we discussed before, the estimation of coefficients of VPI are positive either in symmetric version or asymmetric. It shows that if $GL(H_{ijt}, \gamma)$ increases, the VPI_{ijt} will have more significant effect on consumers' choice behavior. Since $GL(H_{ijt}, \gamma)$ captures the loyalty level to brand j by consumer i, she desires to consider a higher weight on its value of perfect information (VPI_{ijt}). Note that the VPI_{ijt} term is always positive since it was defined as the expected gain to explore j. So, increasing weight of VPI_{ijt} in $GL(H_{ijt}, \gamma)$ will increase the chance to repurchase brand j. This can be interpreted as bias toward a product you had an experience about it. Note that, in the "Myopic-VPI Asymmetric" model, people have a larger weight on VPI of inferior choices, 1.078, than the superior choice, 0.327, at each time step. The VPI of the superior choice is based on why a consumer should switch away from the brand which myopically provides the highest current expected utility; while the VPI of an inferior choice is based on why a consumer should choose it than her best myopic choice. This can be interpreted as an evidence of loss aversion. Consumers considers the value of information of inferior alternatives more than the superior myopic choice at time t into their decision process. Here, consumers prevent losses which occurred by ignoring exploration of inferior choices.

into the utility function to capture consumers' brand loyalty which is observed in our sample.

Table 6: Parameter Estimates for Diaper Data

	No Learning Without Brand Loyalty	No Learning With Brand Loyalty	Myopic Without Brand Loyalty	Myopic With Brand Loyalty	Index	Myopic-VPI	Myopic-VPI Symmetric	Myopic-VPI Asymmetric
Relative mean of prior beliefs								
Pampers	0	0	0	0	0	0	0	0
	—	—	—	—	—	—	—	—
Huggies	0.035	0.025	0.027	0.041	0.111	0.087	0.114	0.093
	(0.0066)	(0.03)	(1.41)	(1.25)	(7.8)	(0.082)	(0.74)	(0.114)
Luvs	-0.233	-0.18	-0.437	-0.183	-0.245	-0.31	-0.274	-0.305
	(0.011)	(0.03)	(0.351)	(3.28)	(1.45)	(0.223)	(0.692)	(0.171)
Other brand	-0.782	-0.608	-1.237	-0.1.308	-1.478	-1.153	-1.243	-1.26
	(0.017)	(0.03)	(0.108)	(1.57)	(1.86)	(0.056)	(0.054)	(0.064)
Standard deviation of prior beliefs								
Pampers	—	—	0.353	0.786	0.918	0.389	0.368	0.381
	—	—	(0.0331)	(0.255)	(0.765)	(0.093)	(0.093)	(0.12)
Huggies	—	—	0.254	0.729	0.33	0.157	0.131	0.153
	—	—	(0.144)	(0.468)	(1.325)	(0.07)	(0.178)	(0.154)
Luvs	—	—	0.432	1.267	1.18	0.48	0.45	0.461
	—	—	(0.099)	(2.258)	(1.537)	(0.131)	(0.188)	(0.137)
Other brand	—	—	0.077	0.653	0.62	0.0044	0.0068	0.0053
	—	—	(0.34)	(1.68)	(2.774)	(0.005)	(0.092)	(0.025)
True mean quality								
Pampers	—	—	0.664	0.448	0.597	0.884	0.763	0.848
	—	—	(0.141)	(0.481)	(0.654)	(0.283)	(0.278)	(0.311)
Huggies	—	—	1.688	0.906	3.695	4.704	5.978	5.194
	—	—	(2.0007)	(1.62)	(11.434)	(4.185)	(8.987)	(8.86)
Luvs	—	—	-0.042	-0.382	-0.124	0.27	0.256	0.262
	—	—	(0.3805)	(0.467)	(0.701)	(0.35)	(0.918)	(0.31)
Other brand	—	—	-0.868	-0.69	-0.921	-0.622	-0.797	-0.69
	—	—	(0.161)	(0.787)	(0.658)	(0.189)	(0.218)	(0.238)
Inherent quality variance								
Pampers	—	—	1	1	1	1	1	1
	—	—	—	—	—	—	—	—
Huggies	—	—	1.044	1.089	1.345	1.092	1.125	1.117
	—	—	(0.713)	(0.85)	(5.71)	(0.056)	(0.416)	(0.64)
Luvs	—	—	0.726	0.508	0.757	0.825	0.849	0.837
	—	—	(0.35)	(0.221)	(0.457)	(0.178)	(0.457)	(0.12)
Other brand	—	—	0.07	0.511	0.303	0.003	0.0059	0.0054
	—	—	(0.322)	(0.936)	(0.778)	(0.0054)	(0.079)	(0.028)
Price sensitivity	-0.126	-0.143	-0.149	-0.145	-0.154	-0.133	-0.155	-0.16
	(0.014)	(0.02)	(0.023)	(0.022)	(0.024)	(0.016)	0.023	(0.025)
Risk aversion	—	—	1.807	0.994	0.971	1.475	1.518	1.499
	—	—	(0.139)	(0.086)	(0.42)	(0.072)	(0.118)	(0.107)
γ	—	0.661	—	0.69	0.639	0.395	0.333	0.469
	—	(0.023)	—	(0.05)	(0.075)	(0.094)	(0.175)	(0.118)
λ	0	3.54	0	2.314	2.05	1.074	0.889	1.0441
	—	(0.15)	—	(0.352)	(0.487)	(0.192)	(0.324)	(0.369)
Coefficient of VPI for the Superior Choice							0.182	0.327
							(0.368)	(0.505)
Coefficient of VPI for Inferior Choices							0.182	1.078
							(0.368)	(0.764)

- All estimations are driven by 100 draws purchase history signals and 200 draws to evaluate the VPI by using Monte-Carlo integration Method.
- All standard errors are estimated by applying *OPG* (Other Product Gradient) estimator.
- As in (Lin et al., 2015), a grid of size ($M = 200$ and $N = 75$) to estimate the Whittle indexes.
- As in (Lin et al., 2015), the discount factor β is fixed at 0.9 to estimate the Whittle indexes.

8 Further Analyses

As a further visualization of model fit, we investigate more evidence by measuring each model in terms of explaining consumers' purchase pattern beyond the standard goodness-of-fit criteria. In this section, we compare the five models— Myopic with Brand Loyalty, Index Strategy, Myopic-VPI, Myopic-VPI Symmetric, and Myopic-VPI Asymmetric— which have shown viable performances in Table 5, based on the predicted switching patterns and predicted actual choices.

8.1 Predicting Switching Pattern

To have a better understanding about the model fit, we compare models in terms of predicting switching patterns. From a managerial perspective, it is very important to predict how many consumers are switching form a brand to another brand at a purchase occasion. So, our statistical approach provides us more accurate evidence about explanatory aspects of each model to predict the consumers' switching pattern. First of all, we determine consumers switching patterns over 10 first purchase occasions at data. For purchase occasion k, we determine a 4×4 matrix as in Table 7a. Entry d_{ij}^k determines the fraction of consumers who purchased brand i at purchase occasion $k-1$, and switch to brand j at purchase occasion k. As Tables 4a-4c show, consumers switched less over time. This is consistent with our exploration-exploitation story.

Table 7: The Switching Matrices

Brand	Pampers	Huggies	Luvs	Other Brands
Pampers	d_{11}^k	d_{12}^k	d_{13}^k	d_{14}^k
Hugggies	d_{21}^k	d_{22}^k	d_{23}^k	d_{24}^k
Luvs	d_{31}^k	d_{32}^k	d_{33}^k	d_{34}^k
Other Brands	d_{41}^k	d_{42}^k	d_{43}^k	d_{44}^k

(a) Actual data switching matrix at purchase occasion k

Brand	Pampers	Huggies	Luvs	Other Brands
Pampers	m_{11}^k	m_{12}^k	m_{13}^k	m_{14}^k
Hugggies	m_{21}^k	m_{22}^k	m_{23}^k	m_{24}^k
Luvs	m_{31}^k	m_{32}^k	m_{33}^k	m_{34}^k
Other Brands	m_{41}^k	m_{42}^k	m_{43}^k	m_{44}^k

(b) Predicted switching matrix at purchase occasion k Under model m

For each descriptive model, we generate the average switching pattern matrix at purchase occasion k based on 100 draws of purchase history signals (Table 7b). Then, we calculate the overall mean absolute error (MAE) and the overall weighted mean absolute error (WMAE) as follow:

$$E_m = \frac{1}{10} \sum_{k=1}^{10} \sum_{i,j} \mid d_{ij}^k - m_{ij}^k \mid \tag{43}$$

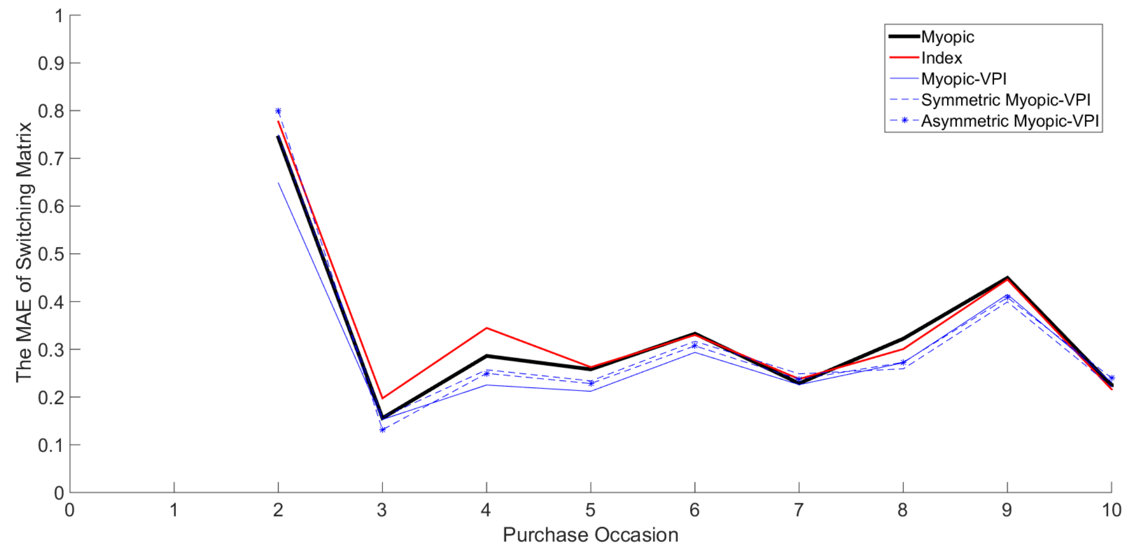
$$WE_m = \frac{1}{10} \sum_{k=1}^{10} \sum_{i,j} d_{ij} \times \mid d_{ij}^k - m_{ij}^k \mid \tag{44}$$

This provides us a clear criteria to determine which model has a better explanatory power to generate correctly consumers' switching pattern. The WMAE determines each cell's weighted absolute error based on its importance. For example, the diagonal entries represent the market shares. Arguably, predicting market shares correctly can be more important for a brand manager. The WMAE will reflect the above importance.

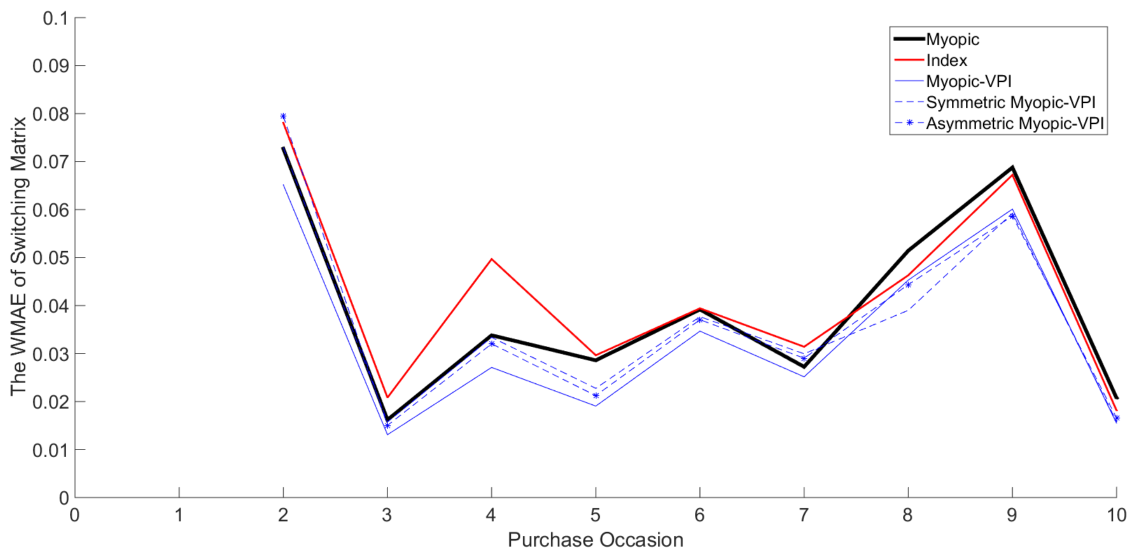
According to Table 8, the Myopic-VPI models have better performances than "Myopic With Brand Loyalty" and "Index Strategy". On average, Myopic-VPI models provide approximately 10% and 17% lower MAE and WMAE relative to Index Strategy respectively. Also, it is important to track these prediction errors over purchase occasions. Figure 7 depicts the predicted switching patterns absolute errors at 10th first purchase occasion. Myopic-VPI models dominate the "Myopic With Brand Loyalty" and "Index Strategy" at both levels of MAE and WMAE. These results show that Myopic-VPI is a better viable way to explain the exploration-exploitation trade-off. Thus, our VPI approach can be considered as a superior forward looking model to explain exploration and exploitation trade-off in a more practical and pragmatic way.

Table 8: Fit Based on Switching Matrices

	Myopic Learning With Brand Loyalty	Index Strategy	Myopic-VPI	Myopic-VPI Symmetric	Myopic-VPI Asymmetric
MAE	3.26	3.37	2.92	3.1	3.12
WMAE	0.4	0.42	0.34	0.363	0.367
No. of Observation	1,538	1,538	1,538	1,538	1,538



(a) MAE of Predicted Switching Patterns at purchase occasion k



(b) WMAE of Predicted Switching Patterns at purchase occasion k

Figure 7: Predicted Switching Patterns Absolute Errors

8.2 Predicting Actual Choices

We can compare models in terms of predicting consumers' actual choices. It is important to be able to predict a consumer's choice at purchase occasion k at the specific observed prices. Here, our statistical approach provides us further evidence about explanatory aspects of each model to match with the consumers' choices.

We draw 100 purchase history signals for each consumer. Then, we compare the simulated choice with the actual choice at each purchase occasion. Precisely, for each draw d, we determine her simulate choice at purchase occasion k under each model (with respect to observed prices at purchase occasion k by consumer i). Let's C^m_{ikd} denote a boolean variable. $C^m_{ikd} = 1$ if and only if the simulated consumer i's choices, at draw d at purchase occasion k under model m, is identically matched with the actual choice of consumer i at time k in data. When N and D denote the number of consumers and number of draws of purchase histories respectively, we define the average hit rate, denoted by AHR^k_m, as follow:

$$AHR^k_m = \frac{1}{N.D} \sum_{i=1}^{N} \sum_{d=1}^{D} C^m_{ikd}. \tag{45}$$

The AHR^k_m shows the average percentage of times which model m correctly predict consumers' choices at purchase occasion k. We are interested in knowing how far forward-looking models improve the average hit rate relative to AHR^k_M which denotes the Myopic model's average hit rate at purchase occasion k. Figure 8 depicts the level of improvement of the average hit rate relative to AHR^k_M at the first 10th purchase occasions. Again, the Myopic-VPI models (blue lines) are superior than the Myopic model (the base line). However, the Myopic model has the best performance than all forward-looking models in the first purchase occasion. Also, the Myopic model dominates index strategy over purchase occasions. In sum, the intuitive idea of value of perfect information shows a more realistic and doable approach to model exploration-exploitation in the market. Our statistical types of evidence show that Myopic-VPI matches better with consumers' choice patterns than the near to optimal solution. Its performance is consistent and robust with respect to different aspects of consumers' choice data.

9 Conclusion

In this paper, we propose that consumers use cognitively simple heuristics to solve a forward-looking learning problem. Based on a reinforcement learning method in the artificial in-

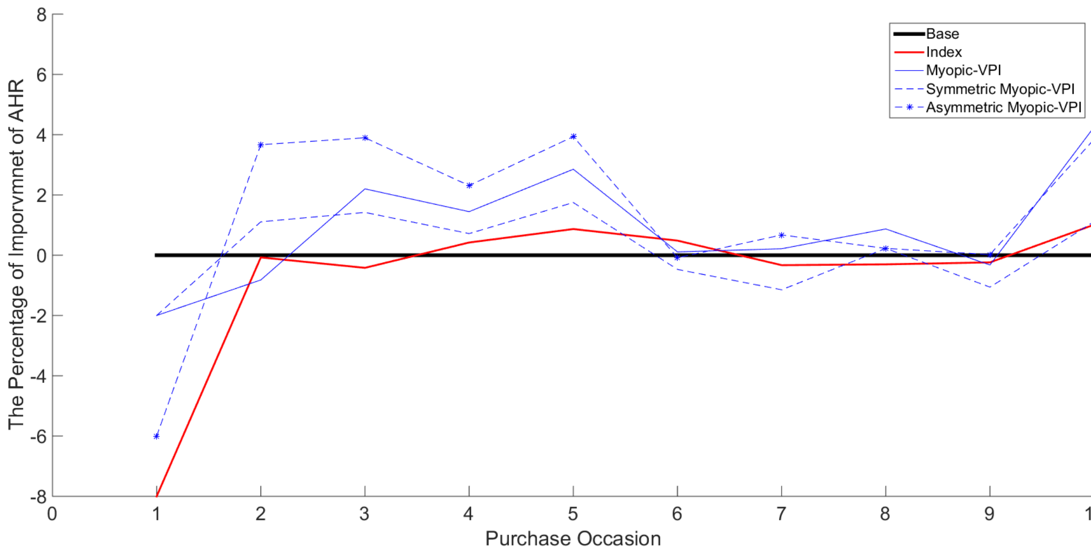


Figure 8: The Percentage of Improvement of AHR_m^k relative to AHR_M^k

telligence literature, we propose Myopic-VPI as a new heuristic approach to balance the exploration vs exploitation tradeoff. We show that its cognitive tractability is significantly higher than than near-optimal approach (i.e., the Index Strategy approach). Also, we argue that consumers can intuit the value of perfect information easily. In VPI, they consider their cognitive limitations to process all informational nodes; so a decision maker rationally focuses on those types of information that have potential to change myopic behavior.

Using simulated data, we demonstrate that a well-defined Myopic-VPI performs better than the myopic behavior regardless of prior beliefs. Moreover, we show that its losses of accumulated utility over time, in a comparison to Index Strategy, is insignificant. We determine the computational time for per-decision respect to each descriptive choice model. Our result shows that Myopic-VPI should be regarded as a better "fast and frugal" decision-making process since it provides a significant long-term payoffs by spending less cognitive efforts.

Using field data on diaper purchases, we show that Myopic-VPI can capture consumers' choice under uncertainty well. It fits the data significantly better than a model without learning or a myopic learning model. Compared with an approximately optimal solution, it requires significantly lower computational costs and it is much easier to implement for structural estimation. Our results provide evidence to support Myopic-VPI as a better "as if" heuristic descriptive learning model which (1) reduces significantly computational burden to estimate forward-looking learning model; and (2) presents a more tractable "as if" model for consumers to use and can explain their choice behavior under uncertainty. We also find some evidence that consumers' brand loyalty have a positive effect on the exploration value of their favourite brands. The asymmetric aspect of VPI learning process shows that people pays more attention to the VPI of inferior myopic choices more than the superior one. This can be interpreted as a loss aversion evidence among consumers in our sample. In sum, Myopic-VPI has a viable and robust explanatory aspects to match with consumers' choice patterns.

We address many issues and advantages of VPI approach; but there are other aspects of VPI which should be considered in future research. We do not model multi-dimensional learning models. When there are several uncertain attributes, consumers need to rely more on heuristic approaches. VPI can be implemented in a multi-dimensional learning model. It would be interesting to study how consumers weigh the value of perfect information in each attribute based on its importance. It would also be interesting to model correlated learning and extend the VPI approach to capture spill-over effects in learning process. One of the advantages of VPI approach is its flexibility to adjust properly to model decision makers' behavior in the above environments without indexability or computational burden restrictions which occur in near to optimal approaches.

References

Daniel A Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040, 2003.

Greg M Allenby and Peter E Rossi. Quality perceptions and asymmetric switching between brands. *Marketing science*, 10(3):185–204, 1991.

Kenneth J Arrow. Is bounded rationality unboundedly rational? some ruminations. *Models of a man: Essays in memory of Herbert A. Simon*, pages 47–55, 2004.

James R Bettman, Mary Frances Luce, and John W Payne. Constructive consumer choice processes. *Journal of consumer research*, 25(3):187–217, 1998.

Robert C Blattberg and Kenneth J Wisniewski. Price-induced patterns of competition. *Marketing science*, 8(4):291–309, 1989.

Bart J Bronnenberg, Michael W Kruger, and Carl F Mela. Database paper-the iri marketing data set. *Marketing Science*, 27(4):745–748, 2008.

Nuno Camacho, Bas Donkers, and Stefan Stremersch. Predictably non-bayesian: quantifying salience effects in physician learning about drug quality. *Marketing Science*, 30(2):305–320, 2011.

Colin Camerer. *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press, 2003.

Andrew Ching and Matthew. Osborne. Identification and estimation of forward-looking behavior: The case of consumer stockpiling. *(April 7, 2017). Rotman School of Management Working Paper No. 2594032.*, Available at SSRN: https://ssrn.com/abstract=2594032 or http://dx.doi.org/10.2139/ssrn.2594032, 2017.

Andrew Ching, Tülin Erdem, and Michael Keane. The price consideration model of brand choice. *Journal of Applied Econometrics*, 24(3):393–420, 2009.

Andrew T Ching and Masakazu Ishihara. Measuring the informative and persuasive roles of detailing on prescribing decisions. *Management Science*, 58(7):1374–1387, 2012.

Andrew T Ching, Susumu Imai, Masakazu Ishihara, and Neelam Jain. A practitioners guide to bayesian estimation of discrete choice dynamic programming models. *Quantitative Marketing and Economics*, 10(2):151–196, 2012.

Andrew T Ching, Tülin Erdem, and Michael P Keane. Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, 32(6):913–938, 2013.

Andrew T Ching, Tülin Erdem, and Michael P Keane. A simple method to estimate the roles of learning, inventories and category consideration in consumer choice. *Journal of choice modelling*, 13:60–72, 2014.

Andrew T. Ching, Tulin Erdem, and Michael P. Keane. Empirical models of learning dynamics: A survey of recent developments. *Chapter 8 in Handbook of Marketing Decision Models (New Edition)*, Berend Wierenga and Ralf van der Lans (eds.):223–257, 2017.

Pradeep Chintagunta, Tülin Erdem, Peter E Rossi, and Michel Wedel. Structural modeling in marketing: Review and assessment. *Marketing Science*, 25(6):604–616, 2006.

C-S Chow and John N Tsitsiklis. An optimal one-way multigrid algorithm for discrete-time stochastic control. *Automatic Control, IEEE Transactions on*, 36(8):898–914, 1991.

Gregory S Crawford and Matthew Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

Richard Dearden, Nir Friedman, and Stuart Russell. Bayesian q-learning. In *AAAI/IAAI*, pages 761–768, 1998.

Richard Dearden, Nir Friedman, and David Andre. Model based bayesian exploration. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 150–159. Morgan Kaufmann Publishers Inc., 1999.

Michael J Dickstein. Efficient provision of experience goods: Evidence from antidepressant choice. *Harvard, unpublished*, 2014.

Zvi Eckstein, Dan Horsky, and Yoel Raban. An empirical dynamic model of optimal brand choice. *Foerder Institute of Economic Research, Working Paper*, (88), 1988.

Tülin Erdem and Michael P Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1): 1–20, 1996.

Xavier Gabaix and David Laibson. A boundedly rational decision algorithm. *American Economic Review*, pages 433–438, 2000.

Xavier Gabaix and David Laibson. Bounded rationality and directed cognition. *Harvard University*, 2005.

Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg. Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, 96(4):1043–1068, 2006.

Noah Gans, George Knox, and Rachel Croson. Simple models of discrete choice and their performance in bandit experiments. *Manufacturing & Service Operations Management*, 9 (4):383–408, 2007.

John Geweke and Michael P Keane. Bayesian inference for dynamic discrete choice models without the need for dynamic programming. *Simulation-Based Inference in Econometrics: Methods and Applications, eds. RS Mariano, T. Schuermann and M. Weeks, Cambridge: Cambridge University Press*, pages 100–131, 2000.

Gerd Gigerenzer and Daniel G Goldstein. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650, 1996.

John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 148–177, 1979.

John C Gittins and David M Jones. A dynamic allocation index for the discounted multi-armed bandit problem. *Biometrika*, 66(3):561–565, 1979.

Peter M Guadagni and John DC Little. A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3):203–238, 1983.

Daniel Houser, Michael Keane, and Kevin McCabe. Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica*, 72(3):781–822, 2004.

Ronald A Howard. Information value theory. *Systems Science and Cybernetics, IEEE Transactions on*, 2(1):22–26, 1966.

J Wesley Hutchinson and Robert J Meyer. Dynamic decision making: Optimal policies and actual behavior in sequential choice problems. *Marketing Letters*, 5(4):369–382, 1994.

Susumu Imai, Neelam Jain, and Andrew Ching. Bayesian estimation of dynamic discrete choice models. *Econometrica*, 77(6):1865–1899, 2009.

Leslie Pack Kaelbling. *Learning in embedded systems*. MIT press, 1993.

Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *arXiv preprint cs/9605103*, 1996.

Wagner A Kamakura and Gary Russell. A probabilistic choice model for market segmentation and elasticity structure. 1989.

Michael P Keane and Kenneth I Wolpin. The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics*, pages 648–672, 1994.

Song Lin, Juanjuan Zhang, and John R Hauser. Learning from experience, simply. *Marketing Science*, 34(1):1–19, 2015.

Nitin Mehta, Surendra Rajiv, and Kannan Srinivasan. Role of forgetting in memory-based choice decisions: A structural model. *Quantitative Marketing and Economics*, 2(2):107–140, 2004.

Nitin Mehta, Xinlei Chen, and Om Narasimhan. Informing, transforming, and persuading: Disentangling the multiple effects of advertising on brand choice decisions. *Marketing Science*, 27(3):334–355, 2008.

Nicolas Meuleau and Paul Bourgine. Exploration of multi-state environments: Local measures and back-propagation of uncertainty. *Machine Learning*, 35(2):117–154, 1999.

Robert J Meyer and Yong Shi. Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem. *Management Science*, 41(5):817–834, 1995.

Robert A Miller. Job matching and occupational choice. *The Journal of Political Economy*, pages 1086–1120, 1984.

Chakravarthi Narasimhan. Competitive promotional strategies. *Journal of Business*, pages 427–449, 1988.

Sridhar Narayanan and Puneet Manchanda. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, 28(3):424–441, 2009.

John W Payne, James R Bettman, and Eric J Johnson. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):534, 1988.

John W Payne, James R Bettman, and Eric J Johnson. *The adaptive decision maker*. Cambridge University Press, 1993.

John H Roberts and Glen L Urban. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science*, 34(2):167–185, 1988.

John P Rust. A comparison of policy iteration methods for solving continuous-state, infinite-horizon markovian decision problems using random, quasi-random, and deterministic discretizations. *Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations (April 1997)*, 1997.

Herbert A Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

Herbert A Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129, 1956.

WT Teacy, Georgios Chalkiadakis, Alex Rogers, and Nicholas R Jennings. Sequential decision making with untrustworthy service providers. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 755–762. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pages 141–162. Springer, 1975.

Iris Van Rooij. The tractable cognition thesis. *Cognitive science*, 32(6):939–984, 2008.

J Miguel Villas-Boas. Models of competitive price promotions: Some empirical evidence from the coffee and saltine crackers markets. *Journal of Economics & Management Strategy*, 4 (1):85–107, 1995.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

Martin L Weitzman. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654, 1979.

Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, pages 287–298, 1988.