Inference of a Firm's Learning Process from Product Launches

Léonce B. Ano[1] • Victor Martínez-de-Albéniz[2]

Abstract

In dynamic business environments, firms must make sequential decisions that account for changes in consumer interests. As consumer interests gradually evolve, firms need to be aware of available decision alternatives to identify promising options. This raises the question of how a firm's current knowledge allows it to make better decisions. To answer this question, we develop a framework in three steps to describe the mechanism for knowledge creation and rule formation when launching new products. First, we conceptualize the firm as an agent using a belief-based decision algorithm based on statistical and psychological literature. Second, we formulate hypothesis about the algorithm's knowledge update and choice rules components and propose a sleeping contextual bandit model of decision-making to solve the exploitation-exploration problem. Third, we use Bayesian hypothesis testing to structurally describe the firm's sequential choice process. We validate our procedure with product decision histories and performance data over 60 years from a leading toy product design firm. We find that a multinomial logit model yields predictions equivalent to a no-belief update Softmax model. However, unlike UCB and Greedy models, the multinomial logit model fails to accurately represent data generated by a firm operating in a changing environment. Finally, we illustrate how one aspect of the firm's adaptation process to changing environment occurs through leveraging product category heterogeneity and transferring knowledge across product features.

Submitted: 20 Juin 2023

Keywords: Innovation, exploration-exploitation, learning process, Bayesian inference, contextual bandit, product decisions, agent's cognition, organization learning.

1. Introduction

Most firms engage in the development and launch of new products in order to cater to evolving consumer preferences. Consumer interests often exhibit gradual shifts, and firms strive to identify the optimal combination of existing features to meet demand effectively. For instance, in 2017, the Coca-Cola company introduced 500 new beverages and variants, representing a 25% increase in product launches compared to the previous year (WSJ 2018a). This initiative was aimed at revitalizing stagnant sales by addressing the growing consumer demand for healthier drink options. As a result, sales experienced a 5% increase within a year (WSJ 2018b). Similarly, in 2015, following a series of disappointing sales performances for Barbie, Mattel's CEO was replaced (WSJ 2016a). The new management swiftly responded by successfully releasing Barbie dolls in new body sizes and colors (WSJ 2016b). For both Coca-Cola and Mattel, as well as numerous other firms, the ability

[1]IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain. Email: lano@iese.edu.

[2]Corresponding author. IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain. Email: valbeniz@iese.edu.

to select and launch new product variants is a crucial adaptive capability to meet the ever-changing preferences of consumers.

By developing and introducing more diverse and sophisticated product variants, firms engage in the exploration of product spaces as a means to foster growth and thrive amidst competition. This strategic approach can enable them to establish profitable market niches, affording them a temporary monopoly position (Tirole 1988, Kim and Mauborgne 2014). In industries characterized by horizontal differentiation, particularly within mature product categories, the process of exploration is often implicitly defined. Firms aim to uncover the elusive *formula* for effectively recombining existing features into successful new products. This entails a quest to better understand the underlying principles and patterns of product performance.

In the academic literature, this understanding is often studied as an exploration-exploitation problem. On the one hand, exploration involves trying new options and venturing into risky or unexplored spaces. On the other hand, exploitation refers to the act of repeating choices that have proven successful in the past (e.g. consistently betting on the winning team). Exploitation decisions are driven by the pursuit of predictable high rewards in the short term. By favoring what has already worked, decision-makers may benefit from immediate gains. However, this risk-averse behavior can also lead to a trap where the decision-maker becomes confined within a low-margin niche. Relying solely on exploitation can limit the potential for long-term growth and hinder adaptation to changing circumstances. To avoid a decline in performance, decision-makers often need to actively seek new opportunities while still capitalizing on their existing knowledge and successes. This requires striking a balance between exploring new possibilities and exploiting known ones. While the conquest of new and exciting regions can be enticing, it can also come with unfavorable consequences, such as bitterness and the loss of valuable resources.

In the field of new product development, it is well-established that extensive search efforts are undertaken when deciding to introduce new product categories that do not yet have a mental representation (Terwiesch 2008). In order to gain insight into the relationship between a firm's belief process and its search efforts, we frame these search efforts as a problem of experiential learning, where experiences are associations between tasks and their performance outcomes. While there is evidence that past experience plays a role in selecting tasks with predictable performance, our understanding of the factors driving ambiguous experiences, such as the launch of new products, is limited. Furthermore, the traditional empirical approach, which relies on reduced form models, does not offer a precise mechanism for understanding why unknown and novel tasks are often chosen.

The multiarmed bandit literature also studies experiential learning problems by considering task-related uncertainty. Several solution approaches exist, such as selecting the task with the highest expected performance while occasionally exploring randomly (e.g., Softmax Greedy algorithm) or incorporating an exploration performance bonus into task performance (e.g., algorithms like the Gittins Index, Upper Confidence Bound - UCB, Thompson Sampling). Although researchers have studied bandit models using real-world data, their focus has primarily been on evaluating algorithm performance (Li et al. 2010, Chapelle and Li 2011, Su et al. 2019) or comparing decision patterns and outcomes generated by algorithms with those made by humans (Li et al. 2020), or they have used experimental data (Ferecatu and De Bruyn 2022, Gans et al. 2007). However, we are not aware of previous research that uses bandit policies to explain actual product decisions.

In this work, our objective is to gain insights into how firms make sequential decisions under uncertainty by studying the exploration-exploitation problem in a real-world context. We aim to address the following question: how can we structurally describe a firm's learning process based on observed data? To answer this question, we adopt a Bayesian perspective, wherein the decision-maker is conceptualized as a single macro-agent who learns through interactions with the environment. We characterize this agent's learning process by describing her environment, her information structure as well as rules governing her knowledge management and choice processes.

The learning process is characterized by two main components: the knowledge management structure, which describes how information is used and updated by the agent; and the choice policy, which describes how product launches are decided given a certain knowledge state. On the one hand, the knowledge management strategy specifies the variables that determine the performance of a certain new product choice. We assume that the performance of any new item is parameterized by a vector of item features, such as a category identifier or aggregate category descriptors that affect the demand. Furthermore, the 'state' of knowledge may be updated after each product launch, by updating the coefficients relative to each of the features in a Bayesian manner; it is also possible that the state remains the same, when no update occurs. In other words, knowledge may (or may not) accumulate from the observation of the performance of past product launches, but this will depend on how the agent functions, i.e., what variables are considered and how they are updated.

On the other hand, launch choice rules describe, in probabilistic terms, how likely it is to launch a certain new product. At any launch moment, the agent can either launch a variant of an existing product – which we operationalize as them being in the same product category – or opt for a new-to-the-world category. The agent starts making decision with knowing only one category, whose number grows as she continues interacting and organizing her knowledge. Each category represents a bandit arm. The choice rule crystallizes whether the agent is employing an active learning policy, in which lower short-term rewards are selected more often, or follows a passive learning approach, in which short-term rewards are the main driver of launch decisions. By observing actual launch decisions together with product performance, we are able to infer what is the most likely knowledge and choice policy structure. These two components can be separately identified, by studying the effect of a successful performance in a certain category onto the likelihood of a next launch within the same category. Specifically, we develop an inference approach to reverse-engineer the agent's decision model (knowledge and choice), including the starting values – priors and uncertainty levels – of the model parameters.

To validate our methodology, we apply this novel inference approach with data from LEGO, a prominent toy firm with a unique product identity. We collect product data on 12,861 products released from 1949 to 2023. We classify products into LEGO categories, and measure how successful each one was, in terms of secondary-market price and volume. We then implement our inference approach to find which bandit policy best fits the data. This inference approach provides insights on how LEGO solves the exploration-exploitation problem and its internal learning process, as suggested by the revealed preference about launch decisions.

In addition to our novel methodological framework, our work yields two main empirical findings. Firstly, we provide compelling evidence that the predictions derived from a multinomial logit model are equivalent to those of a no-belief update model. This finding underscores the limitations of reduced-form models in illuminating the underlying process of knowledge update and fails to offer an explanation for the decision-maker's goals driving their choices. Consequently, in dynamic environments, reduced-form models prove inadequate in capturing the acquisition of new insights and the adaptation to evolving consumer preferences. Secondly, our analysis reveals that LEGO, as a case study, does not rely on simple, reduced-form heuristics for launching new products. Instead, our findings suggest that LEGO likely employs bandit learning policies, specifically active learning (UCB) or passive learning (Greedy) models, guided by performance-based rewards, such as revenue. The UCB model demonstrates better data fit over a short-term horizon (training period from year 1949 to 1968), while the Greedy model appears more plausible over longer horizons, as evidenced by our out-of-sample results.

Our paper makes significant contributions both in terms of methodology and managerial implications. Firstly, we bridge the gap between theoretical frameworks and data-driven research by developing Bayesian bandit models and leveraging real-world data to infer both the beliefs and decision-making behavior of agents. This approach challenges traditional modeling approaches and provides valuable insights into how firms adapt their decision-making processes in response to changing market conditions. By highlighting the use of bandit learning policies in real-world settings, we shed light on how firms navigate uncertainty and adjust their strategies based on performance feedback. Specifically, our research illustrates how a firm's adaptation process to a non-stationary environment occurs through product category heterogeneity and the transfer of learning across different product features. Our modeling framework enables us to quantify and assess the importance of various elements within this adaptation process, offering a comprehensive understanding of the dynamics at play. Furthermore, our contribution to proposing cognitive models that better explain agents' adaptive behavior aligns with the broader goal of developing explainable and safe artificial agents capable of making decisions in non-stationary environments (Da Costa et al. 2022).

Secondly, from a managerial perspective, our work offers practical implications that can help managers enhance the quality of their decision-making processes. For example, while *System 1* thinking is known for its speed, adopting a practice within a firm that involves pausing and explicitly formulating hypotheses to reverse-engineer beliefs and models in sequential decision-making can be a valuable practice to reduce noise of decisions made by various agents within an organization (Daniel 2017, Kahneman et al. 2021). Hence, our work can help managers in better understanding and organizing how knowledge is created, applied, and maintained within their organization.

The rest of this paper is organized as follows. In §2 we review the literature. We then present in §3 theoretical framework. Section 4 describes the setting and the data used as an application to our proposed inference process. In §5 we present the results. We conclude in §6.

2. Literature Review

2.1 Innovation and Dynamic Learning Models

In the innovation literature, the exploration-exploitation problem is often studied by focusing on how risks, complexity, and uncertainty in the environment relate to the search process for better product decisions. For example, Chao and Kavadias (2008) define risk as the probability that a product launch decision will achieve its expected performance. They argue that a decision is considered less risky when managers possess knowledge about which product launch will yield quick rewards, and such decisions are more likely to be made in less complex environments. Furthermore, environmental complexity can hinder the ability to infer how interactions among product decisions may affect performance. To measure environmental complexity, Sommer and Loch (2004) propose an autocorrelation matrix of neighboring product decisions and suggest that lower correlation implies higher complexity, as it complicates the prediction of product performance based on these decisions. They emphasize the importance for managers to recognize and articulate the relevant variables and their functional relationships to assess the level of unforeseeable uncertainty and make well-informed decisions. Consequently, traditional risk management techniques can only be implemented when both environmental complexity and uncertainty are low. In other cases, managers must acquire information through trial and error learning or a parallel search approach (Sommer and Loch 2004, Sommer et al. 2009). In our work, instead of framing the problem solely in terms of risk and complex environments, we focus on the agent's behavior in relation to her perception of a noisy and non-stationary environment. We employ the multi-armed bandit framework to model the exploration-exploitation problem.

In the operations research literature, the use of bandit problems to analyze the exploration-exploitation problem is prevalent. Gittins (1979) provided the first computable optimal solution to a general Markovian formulation of this problem, where a firm facing an infinite period has access to a prior Bayesian distribution of beliefs over opportunities and updates the distribution after each period. The solution consists of indices that incorporate immediate reward (value of exploitation) and a measure of the reduction in uncertainty (value of exploration), with the option having the highest index selected (Gittins et al. 2011). However, the assumptions made by the Gittins index are hard to sustain in a real-world changing environment. Caro and Gallien (2007) provide a reformulation of the Gittins allocation rule for dynamic assortment decisions by studying the multiarmed bandit problem (see Cesa-Bianchi and Lugosi 2006 or Slivkins 2019 for a review on bandit). They derive a heuristic policy comparable to policies for restless bandits. Loch and Kavadias (2002) study dynamic budget allocation policies for product lines and find that the optimal policy is that of a restless bandit, for which there is no optimal policy. Rusmevichientong et al. (2010) address a dynamic assortment problem with capacity constraints and propose a multinomial logit (MNL) bandit approach to design a policy by modeling customer preferences. Similarly, Sauré and Zeevi (2013) use Bayesian learning of demand to design an MNL bandit policy. In contrast to these papers that assume known demand, Besbes et al. (2014) provide regret bounds for several policies in non-stationary environments, and Keskin and Zeevi (2017) design a dynamic pricing policy for such environments. Despite the availability of various algorithms for specific bandit problems, the literature focuses on providing performance guarantees in terms of regret minimization. However, little is known about the policies' sampling behavior, which is crucial to understanding the decision patterns recommended by these algorithms and to undertaking ex-post causal inference. Recently, Kalvit and Zeevi (2021) have shown that the UCB policy satisfies an asymptotically balanced sample split among arms, regardless of the problem complexity. In contrast, Thompson Sampling does not, particularly when the gap between the top two arms is small (Kalvit and Zeevi 2021). Our work uses bandit policies to reverse engineer a decision-maker's sampling behavior and her cognitive representation of the environment.

There are various formulations of bandit problems that correspond to different settings. When the environment is non-stationary, one can assume that the decision-maker observes contexts in each period, resulting in a stationary environment (Lattimore and Szepesvári 2018). Another scenario involves bandit problems where the number of arms is finite but can vary over time (Kanade et al. 2009, Huang et al. 2020). A more general formulation is found in restless bandits, where the state of each arm evolves independently of the chosen action, and the learner can only perceive a limited number of arms at a time (Ortner et al. 2014). Bayesian approaches are often employed to study restless bandit problems. For instance, Speekenbrink and Konstantinidis (2015) use this framework with a four-arm bandit in an experimental setting and find that humans reduce uncertainty through exploration decisions. In our work, we focus on describing the structure of decision-making behavior by considering a bandit problem with an increasing number of arms. In a similar vein, Gans et al. (2007) utilize a two-arm Bernoulli bandit to examine whether it is a reasonable representation of consumer choices in an experimental setting. In contrast, we concentrate on studying firm decisions where multiple options are available. Moreover, we estimate decision models with or without Bayesian belief updates using real-world data. Finally, our work also relates to the off-policy learning literature, as we replay various policies from logged data (Levine et al. 2020). However, we do not aim to conduct a counterfactual analysis of policies by comparing their empirical performance, as done in studies such as Li et al. (2010), Chapelle and Li (2011), or by benchmarking them against human decisions (Li et al. 2020, Anderson et al. 2017). Instead, our focus is on delineating the structural dynamics of the decision-maker's learning process. In this regard, our paper casts decision-making as an inference problem, wherein one can derive the decision-maker's beliefs and cognitive model from her actual decision data (Friston et al. 2013, Smith et al. 2022).

2.2 Experiential and Organizational Learning

There is a significant body of empirical research focused on quantifying the average behavioral response resulting from the acquisition of new information through past experience. These studies provide evidence that experience generally improves performance. For example, Kc and Staats (2012) demonstrate that a surgeon's focal experience has a greater impact on performance compared to experience in other related tasks. Staats et al. (2015) show that individuals learn not only from their own successes but also from others' failures. Staats et al. (2018) provide nuances to these findings by highlighting that learning may not occur consistently in response to new information and may be influenced by factors such as the attribute of experience, which can lead decision-makers to discount negative news over time. However, recent methodological contributions have identified limitations in past model specifications, particularly in addressing issues related to unit root tests and self-selection bias (Bennett and Snyder 2017, Anand et al. 2016). Furthermore, there is limited research on learning in the context of ambiguous experiences or experiences where performance cannot be immediately evaluated as success or failure. Musaji et al. (2020) propose a reduced-form model to analyze how learning occurs with respect to recent information and task ambiguity. They find that tasks with higher variance reduce the cost of learning. However, their study is based on cross-sectional data, limiting their ability to study the dynamics of the learning process. Overall, these empirical studies contribute to our understanding of how individuals learn and improve their performance through experience. However, methodological innovations, such as the one we present in this paper, can enable researchers to study learning processes in settings characterized by ambiguous experiences

Finally, our work is also connected to the literature on organizational learning, which explores the learning processes within organizations in relation to various contextual factors (Argote et al. 2020). Some studies investigate how innovative capabilities are influenced by factors such as the pace of change in the industry (Teodoridis et al. 2019), firm performance (Chao et al. 2012), or goal-related tasks (Clark et al. 2018). Others highlight the importance of managers' foreign experience (Godart et al. 2015) or the potential benefits of being a generalist for fostering innovation (Custódio et al. 2019). While these works provide valuable insights into the factors that shape organizations' engagement with innovative tasks, they often lack a comprehensive framework that fully accounts for knowledge creation processes within organizations.

There is an emerging body of literature that adopts a process-oriented perspective on learning tasks and emphasizes the significance of cognitive skills and knowledge representation in decision-making (Posen et al. 2018). This line of research builds upon the tradition of the Carnegie School, which challenges the notion of the firm as a rational economic agent and highlights the role of performance feedback in the firm's adaptive processes. It advocates for opening the "black box" of decision-making to observe and measure the internal mechanisms at work. For instance, Csaszar and Levinthal (2016) investigate the interaction between mental representations and search strategies. They underscore the importance of the decision context, which includes factors such as available choices, firm performance, and profit. Another psychological model proposed by Csaszar (2018) draws upon the lens model framework developed by Brunswik (1952) in psychology, which describes how representations are formed in response to environmental cues. Our modeling framework shares similarities with the work of Denrell et al. (2004), who propose a reinforcement learning model to describe knowledge acquisition. However, their findings are based on simulation studies and do not explicitly address how uncertainty reduction leads to improved decision-making or the specific processes that drive high-quality information acquisition. In a similar vein, our work, like that of Keil et al. (2022), presents a belief-based model of decision-making. However, we propose a different modeling approach, which we compare against their multinomial choice model specification. Furthermore, our work aligns with the perspective of organizations as artificial intelligences (Csaszar and Steinberger 2022). We conceptualize decision-making as the outcome of computational processes carried out by individual decision-makers, referred to as agents, whose decisions and performance can be observed.

3. Theoretical Framework

In this study, we consider a firm that sequentially launches new products. We observe launched products and their performance. At the organizational level, there is a decision-making structure, embodied into a macro agent. Our aim is to infer this agent's beliefs and cognitive model that guides her action. In this section, we provide a theoretical framework in four steps: (i) the agent's setting and environment, (ii) her knowledge management model, (iii) her choice model, and (iv) an approach to infer her model structure from observable outcomes.

3.1 Agent's Environment

Launch decision. The agent's task is to select which new product to launch in the market. Making this decision is akin to select the features that go into a new product. Features known to the agent are both observable functions (e.g., shape, color or themes) and subjective elements (e.g., an abstract concept). We consider discrete periods $t = 1, \ldots, T$, and assume without loss of generality that only one product is launched in each period. We do not need to specify the calendar behind these periods, and can hence accommodate various launch schedules in a flexible way, including simultaneous launches (although this implicitly requires that information revelation occurs instantaneously). For instance, in the toy industry –the context from which we apply our theory later–, periods can have a shorter duration prior to the holiday season, and a longer one during the rest of the year. In other words, we just needs to specify T, which corresponds to the total 'budget' that the firm can spend on new product launches.

Partially observable and stochastic environment. The agent cannot accurately predict the outcome of her decisions. She must learn about her environment by launching products and observing performance. At inception, the firm's environment is unknown. As the firm acquires experience, she may better identify market trends, seasonality, competition dynamics and progress on her learning curve. She may encode her knowledge about the environment with feature variables.
Product category as the unit of action. We consider a product within a category i as a potential item to launch. We thus assume that agent's product knowledge organization occurs at the category level (Russell 2010). Specifically, products within one category share the same knowledge representation (product features and performance). It is characterized by the agent's knowledge structure K, which defines it by a feature vector X_{it}. This vector may capture aggregate measures of the category, such as the amount of products launched so far, or their average price. We consider that the launch can be made either from a finite set of existing categories $\mathcal{I}_t := \{1, \ldots, I_t\}$, where I_t denotes the number of categories launched up to t; or from a new category $i = 0$, for which this will be the first experience. When the agent chooses to launch a new category, $I_{t+1} = I_t + 1$; otherwise $I_{t+1} = I_t$.
Information structure and performance. The agent makes its launch decision using C, a choice policy. This choice is driven by comparing the potential performance of each option, given product features X_{it} for different existing options $i \in \{1, \ldots, I_t\}$ or a new category. For every potential choice i, performance R_{it} is a random variable characterized as a function of the agent's current knowledge and a random shock. We assume that performance R_{it} follows a distribution that depends on Ω_t, which includes all beliefs about the potential of each category, in the form of a vector $\tilde{\theta}$, that follows a normal distribution with mean effect μ_t and variance-covariance matrix Σ_t. In other words, $\Omega_t := (\mu_t, \Sigma_t)$. If category $i \in \mathcal{I}_t$ is selected, then product performance follows a normal distribution with average $\tilde{\theta}^T X_{it}$ (T to transpose vector $\tilde{\theta}$) and a Gaussian shock ϵ_{it} with standard deviation $\sigma_\epsilon \geq 0$. Formally, product performance is written as

$$R_{it} = \tilde{\theta}^T X_{it} + \epsilon_{it} \tag{1}$$

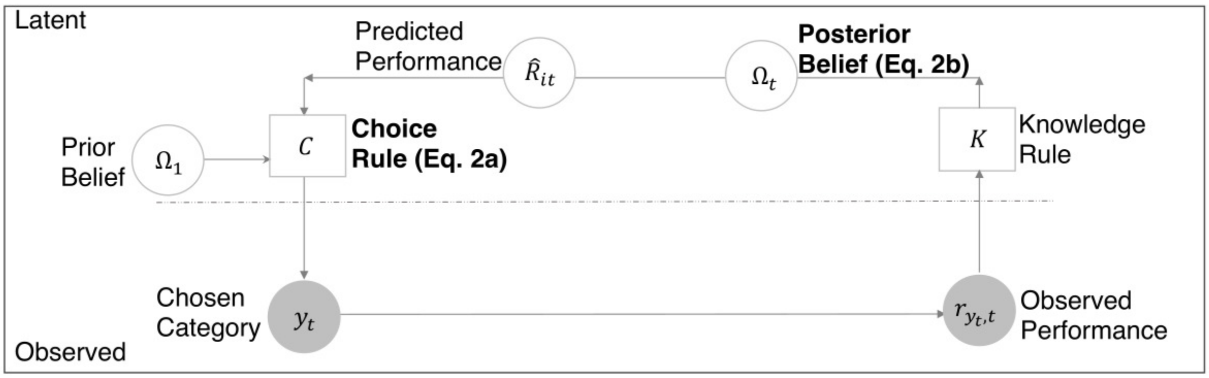
Note that these sufficient statistics X_{it} may overlap across categories. X_{it} could include general features such as average price of products launched so far, as well as category-specific identifiers. For example, if the first two entries are generic and the rest are category-specific, category 1 could be represented by $X_{1t} = (1.5, 5.2; 0, 1, 0)^T$, while in category 2 by $X_{2t} = (3.1, 2.6; 0, 0, 1)^T$. Category 0 would then be represented as $X_{0t} = (0, 0; 1, 0, 0)^T$.
Sequence of events. The agent seeks to maximize her long-term performance. Since the agent sequentially makes decision under uncertainty, she faces the classical exploration-exploitation trade-off. She behaves in an iterative manner, starting with her knowledge vector Ω_t. This allows her to simulate tuples $(i, R_{it})_{i=0,\ldots,I_t}$, then choose one category to launch, which we denote by Y_t (a random variable at this point), which realizes into an observed value y_t. Finally, the realized reward $r_{y_t,t}$ is observed and the agent can learn from it to form a new Ω_{t+1}. The agent's model M is thus made of a knowledge rule K and a choice rule C, i.e., $M = (K, C)$. Formally, the agent's learning process for $t \geq 1$, starting from a given Ω_1, is illustrated in Figure 1 and can be structured as follows:

$$Y_t \sim Choice(C, \Omega_t) \tag{2a}$$

$$\Omega_{t+1} = Update(K, \Omega_t, y_t, r_{y_t,t}) \tag{2b}$$

First, Equation (2a) reveals the chosen outcome. Note that the choice mechanism is driven by Equation (1), which depends on the information structure Ω_t which affects the distribution of $\tilde{\theta}_t$, but also on the choice rule C with which the agent selects out of the options $i = 0$ or $i \in \mathcal{I}_t$. Second, in Equation (2b), the agent uses her knowledge management policy K to update her former belief Ω_t into Ω_{t+1}. She leverages both the outcome of her choice y_t and performance $r_{y_t,t}$ to compute a posterior belief Ω_{t+1}.

Figure 1: The agent's learning process is represented as a loop in which variables are updated.



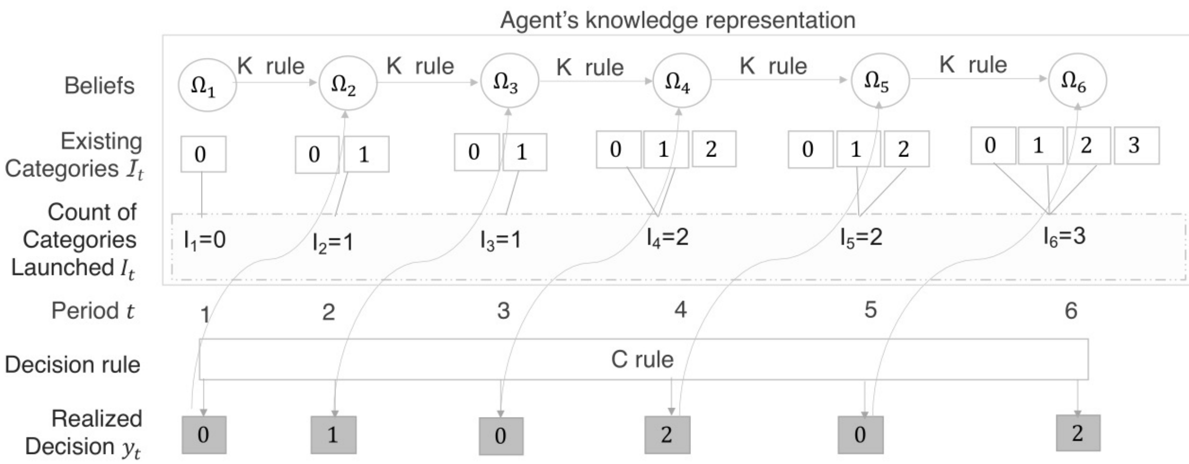
3.2 Knowledge Management Rules: Representation and Belief Update

The agent's information structure Ω_t contains information about the belief vector $\tilde{\theta}$ which determines the potential for each category. We model the agent's perception of her environment with a sleeping linear bandit and describe various methods through which belief update may occur.

Knowledge parameterization with a bandit model. We choose the multi-armed bandit framework to model the agent's knowledge primitives. In each period, the agent's task consists in pulling a category, the bandit arm, from which a new product will be selected for launch. Bandit problems traditionally work under the assumption of a finite known number of independent arms. In a context of new product introductions, the assumption is not valid since the agent does not know ex ante the number and type of categories available.

Product categories in a sleeping linear bandit model. To overcome the aforementioned shortcoming, we recall the assumption stated in §3.1: product performance is driven by features X_{it} and the agent faces a linear bandit. Furthermore, we allow a growing number of categories as the agent acquires more experience. Some arms are thus in a *sleeping* state and are revealed after a product has been launched within that category. Namely, the contextual variable X_{it} is revealed and the specific arm 'awakens'. In other words, the agent faces a sleeping linear bandit model, illustrated in Figure 2. In each period t, the agent can choose from existing categories, which are already awake, and hence have well-defined covariates X_{it}. It can also choose to launch a product from category $i = 0$, which is a placeholder for all developed products not yet launched. The performance of this category is driven by a generic $\tilde{\theta}_t X_{0t}$, and results in $Y_t = 0$ if launched. However, after the product is launched, it opens up category $I_{t+1} = I_t + 1$, which now has its own covariates. Since there is no existing categorical parameter for the new category I_{t+1}, we assume that it inherits its value (mean and variance-covariance) from category $i = 0$'s posterior, as suggested by Gilboa and Marinacci (2016) or Harsanyi (1967). In the meantime, 'unborn' categories stay in a *sleeping* state.

Figure 2: Representation of the sleeping bandit model for 6 decision-periods.



Knowledge dynamics: posterior update. The agent manages her knowledge update K either through a Bayesian update, or no update at all.
Bayesian update. This method provides the normative way according to probability theory to update the agent's beliefs sequentially. The agent's prior is her starting point about the true effect of the environment's performance factor associated with a product features and it shapes the agent's data generating process. This prior must be learnt from observed data generated by the agent's behavior in real-world setting (Gelman et al. 2017). Since we assume a normally-distributed prior, the posterior update is a tractable equation (Powell and Ryzhov 2012). It is written as follows:

$$\mu_{t+1} = \mu_t + \frac{\Sigma_t \cdot X_{y_t t}}{\sigma_\epsilon^2 + X_{y_t t}^T \cdot \Sigma_t \cdot X_{y_t t}} (r_{y_t t} - \mu_t^T X_{y_t t}) \tag{3a}$$

$$\Sigma_{t+1} = \Sigma_t - \frac{\Sigma_t X_{y_t t} (X_{y_t t})^T \Sigma_t}{\sigma_\epsilon^2 + X_{y_t t}^T \cdot \Sigma_t \cdot X_{y_t t}} \tag{3b}$$

In Equation (3a), μ_{t+1} denotes the posterior update of the mean belief μ_t. $(r_{y_t t} - \mu_t^T X_{y_t t})$ is the prediction error and describes the amount of surprise about observed performance. A big surprise leads to a large mean posterior update and reveals a big uncertainty about μ_t. The weight for the prediction error is tuned by an uncertainty ratio. A lower weight will discount the value of information surprise, implies a lower learning rate and occurs either when variance values are low or when Gaussian shock is high.

In Equation (3b), Σ_{t+1} is the posterior update of the variance-covariance of agent's belief. The posterior variance equation is driven by an uncertainty ratio. A higher observation noise σ_ϵ impedes making progress into updating the state of knowledge. Note that it is possible to include a memory factor to 'forget' learning from the distant past data Powell and Ryzhov (2012). However, in our empirical study such memory factor did not improve results much, so we retain the simpler formulation for simplicity.

For example, imagine there are two existing categories, i.e., $\mathcal{I}_t = \{1, 2\}$ with $X_{1t} = (1.5, 0.2; 0, 1, 0)^T$, and $X_{2t} = (3.1, 2.6; 0, 0, 1)^T$, while the new category is $X_{0t} = (0, 0; 1, 0, 0)^T$. At the current state of information $\mu_t = (1.2, -3.0; 0.4, -0.2, 1.4)^T$ and $\Sigma_t = diag(2, 2; 0.8, 0.6, 1.0)$, and $\sigma_\epsilon = 1$. In period t, the product launched is in category 1, i.e., $y_t = 1$, and a reward of $r_{1t} = 4$ is obtained. Since $\mu_t^T X_{1t} = 1.8 - 0.6 + 0 - 0.2 + 0 = 1.0$, applying Equations (3a)-(3b), we obtain

$$\mu_{t+1} = \begin{pmatrix} 1.2 \\ -3.0 \\ 0.4 \\ -0.2 \\ 1.4 \end{pmatrix} + \begin{pmatrix} 3.0 \\ 0.4 \\ 0 \\ 0.6 \\ 0 \end{pmatrix} \frac{3}{1 + 6.75 + 0.016 + 0 + 0.6 + 0} = \begin{pmatrix} 2.27 \\ -2.85 \\ 0.4 \\ 0.02 \\ 1.4 \end{pmatrix}$$

and

$$\Sigma_{t+1} = \begin{pmatrix} 0.92 & -0.14 & 0 & -0.22 & 0 \\ -0.14 & 1.98 & 0 & -0.03 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ -0.22 & -0.03 & 0 & 0.56 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

As we can see, experimenting with category 1 allows the agent to update the value of the parameters that affect that category (entries 1, 2 and 4 in the vector μ_t), while nothing is learnt in the other parameters (entries 3 and 5 which correspond to the category-specific shifts for category 0 and 2). Similarly, the variance-covariance matrix is adjusted, and in particular the variance for the first dimension is strongly reduced from 2 to 0.92, while the rest does not change much. Note that while Σ_t is a diagonal matrix, Σ_{t+1} is not, meaning that we can generally learn from the correlation structure too.

Model-free and heuristic updates. Several empirical studies suggest that the Bayesian update rule does not apply in real-world settings (Benjamin 2019). The learning literature provides many simpler rules to compute posterior update for mean belief parameter, such as the Delta learning rule (also known as Rescorla-Wagner, see Gronau et al. 2017), or the Decay learning rule (Speekenbrink and Konstantinidis 2015). In our study, we do not consider these rules and focus two extreme cases: standard Bayesian updating and no updating (discussed next). It would nevertheless be possible to extend the model by considering these alternative updating rules.

No belief update. In contrast with the above heuristics, we consider a more extreme update, in which the belief remains constant: the agent maintains her prior over every decision period.

3.3 Agent's Choice Rules and Models

Of course, the quality of the learning, in the form of knowledge refinement, depends on the agent's choices over time. This is determined by the choice model (C), which we describe here. An agent seeking to improve her knowledge about the environment would actively experiment to reduce uncertainty, while a exploitation-oriented agent would select actions maximizing her belief of correct action. We can organize these policies into two groups: (i) passive learning and (ii) active learning. We present them next and include their pseudo-code in Appendix A.

Passive learning policies (greedy). A passive learning policy selects actions greedily with respect to mean belief parameters. The randomized choice rule is operationalized by a Softmax function, which determines the probability that the new product launch falls in category i. Using $\gamma \geq 0$, it can be written as follows:

$$p_{it} = \text{Softmax}_\gamma(\mu_t) = \frac{e^{\gamma \mu_t^T \cdot X_{it}}}{\sum_{k=0}^{I_t} e^{\gamma \mu_t^T \cdot X_{kt}}} \tag{4}$$

The probability of selecting an action p_{it} is always between zero and one, leading to a finite log-likelihood, as opposed to possibly $-\infty$ likelihood when the policy does not randomize over all possible choices. Unlike another passive learning policies, the *Softmax* policy explores arms in a directed fashion. Instead of randomly exploring an arm with ϵ probability and making greedy choices otherwise (as in ϵ-greedy), *Softmax* selects arms in descending preference order, tuned by parameter γ. A value $\gamma = 0$ yields a uniform probability of selecting each available actions whereas a positive γ leads agent to select only high performance actions.

Interestingly, as we discuss below, our inference is set to maximize the likelihood of the observed launch patterns. Due to this choice, it is not possible to separately identify γ and μ_t. In other words, our inference problem has one degree of freedom due to the structure of the choice probability (adding to one by construction), so we can set $\gamma = 1$ without loss of generality. Having said that, if one chooses to also consider the likelihood of the observed rewards, which depend on μ_t but not γ, then we can identify it.

Active learning policies. With active learning rules, the agent attempts to balance exploration-exploitation decisions. There are two classes of policies: randomized policies (Thompson sampling - TS) and non-randomized policies which are again smoothed with a *Softmax* layer to make them randomize over the entire choice set (Upper Confidence Bound - UCB, Gittins Index - GI). Note that there are other active learning policies such as Knowledge Gradient or Value of Information Ratio, but we focus on the former two policies in our study.

Randomized Upper Confidence Bound. With a UCB policy, the agent selects action with the highest upper confident bound performance, which is computed by taking into account the mean performance and adding an additional value for the uncertainty bound, which is tuned with a parameter β. A higher value of β indicates a higher exploration rate. We implement the randomized UCB by applying the *Softmax* function on the product performance with exploration bonus.

Randomized Gittins. The agent uses the optimal solution to the exploration-exploitation problem formulated as a Markov Decision Process (MDP), formulated in a setting of infinite horizon, with a known prior about a finite number of categories and a known state-transition probability matrix $p(\theta'|\theta, X_i)$. With this solution, the agent selects category i, whose performance makes her indifferent between exploring and exploiting. In other words, for each category, a threshold is computed so that the arm should be pulled when the expected reward plus an exploration bonus is higher than the threshold value (Chick et al. 2010, Han and Powell 2020). Again, to generate a randomized policy, we also augment the Gittins index with the *Softmax* rule.

Thompson-sampling. The agent selects her choice through a simplified Bayesian optimization mechanism, that picks the option with the highest predicted performance from a single draw of her belief posterior distribution, instead of maximizing with respect to the full distribution. Exploration occurs through the non-zero probability of selecting every actions, even unknown ones. Unknown actions with high mean and higher variance would be explored often, and belief update would lead to reduce uncertainty. Unknown actions with lower mean than other actions but higher variance would also be explored occasionally and beliefs would be updated to learn their parameters. Recall that since rewards across arms are correlated through product features, uncertainty reduction will still take place by improving the precision of beliefs associated with these features. We compute the choice rule probability p_{it} by running $n = 1,000$ simulations of the agent's decision and measuring the frequency in which arm i delivers that highest simulated outcome R_{it}.

A Reduced-Form Benchmark. In addition to passive and active learning models, we specify a multinomial logit (MNL) model to account for the fact that available categories are time-variant (Cameron and Trivedi 2010). The MNL model controls for the number of alternatives available in each decision period. It sets the attractiveness of a category as μX_{it}, where μ is a static parameter that can be estimated from actual launch decisions y_{it} with standard econometric packages (The inference of the multinomial logit is undertaken with the cox proportional hazard package in R).

Once choice policies are defined, the agent's strategy is shaped by the interaction of her knowledge and choice rules. We provide a description of each of several candidate models and take an engineering perspective on her launch decisions to retrieve both her prior and data generating process. We benchmark a reduced-form model against other models summarized in Table 1. Next, we describe the inference procedure.

Table 1: Agent's candidate models

Model number	Class	Knowledge update	Choice rule
0	Reduced form		
1	No belief update	None	Greedy, Softmax
2	Passive learning	Bayesian	Greedy, Softmax
3	Active learning	Bayesian	UCB, Softmax
4	Active learning	Bayesian	Gittins, Softmax
5	Active learning	Bayesian	Thompson sampling

3.4 Inference Procedure

In this section, we describe our method to retrieve the decision model best supported by the agent behavior data. We undertake a Bayesian inference that takes place in two steps: parameter estimation first for a given model in Table 1, i.e., estimate the starting values for the prior and compute the highest likelihood; and then model selection among the five options.

Parameter Estimation. Unlike most Bayesian analysis that specifies a default prior, the understanding of agent's behavior in real-world setting requires to set her prior in the context of the likelihood function (Gelman et al. 2017). Indeed, agent's prior is associated with her data generating process (Kruschke and Liddell 2018). Hence, the prior distribution must be selected from the most likely model (Vanpaemel and Lee 2012). To estimate parameters from bandit models, we simulate the agent decision with each available policy and the logged data. Then, in each period, we can compute the one-period log-likelihood as follows:

$$LLH_t = \ln\Big(p(y_t|M, X_{\cdot t}, \mu_t, \Sigma_t, \sigma_\epsilon^2)\Big) \tag{5}$$

which yields the model log-likelihood $LLH = \sum_{t=1}^{T} LLH_t$. Finally, we search for agent's belief parameters contained in information structure Ω_1 by using a grid-search approach to solve the following optimization problem:

$$\max_{\substack{\mu_1 \\ \Sigma_1 \\ \sigma_\epsilon^2}} \sum_{t=1}^{T} \ln \left(p(y_t | M, X_{\cdot t}, \mu_t, \Sigma_t, \sigma_\epsilon^2) \right) \tag{6}$$

Model comparison. The likelihood value $p(y_t|M)$ is an initial starting point to compare models. However, it would favor complex models that would overfit the data and fail at predicting new data. Thus, we evaluate model strength with the Watanabe-Akaike Information Criteria (WAIC), a fully Bayesian information criteria that measures the goodness-of fit as well as the model complexity.

The lower the WAIC, the better the data supports the model. In addition, we compute the WAIC variance and the Akaike weight to validate the model selection decision. The higher the WAIC variance, the lower the confidence in the WAIC estimates and the lower the confidence in the estimation decision. Finally, the Akaike weight puts the WAIC value for each models on a probability scale and provides evidence for model plausibility. It is computed as follows:

$$w_{M_k} = \frac{exp(-\frac{1}{2} dWAIC_{M_k})}{\sum_{q=1}^{K} exp(-\frac{1}{2} dWAIC_{M_q})} \tag{7}$$

where $dWAIC$ is the difference between the WAIC value for model M_k and that of the model with the lowest $WAIC$. The exponential factor is the WAIC value transformed on a probability scale.

Note that WAIC metric has been shown to be asymptotically equal to Bayesian leave-one-out cross-validation. However, cross-validation requires data to be conditionally independent, whereas the agent's data is a time series. Thus, the WAIC metric may have its limitations. To alleviate this problem, we partition the dataset in two samples (in-sample and out-of-sample), we estimate parameters from the first sample and evaluate predictions from the out-of-sample dataset. We use both log-likelihood and WAIC as accuracy metrics to compare models.

4. Empirical Setting

In this section, we present an application of our framework on data from LEGO products. First, we provide background on the setting, then present some descriptive statistics and finally describe how we operationalize the inference procedure.

4.1 The Agent: LEGO

LEGO, a renowned toy company, introduced its first plastic construction product in 1949. As of 2022, the global revenue of the toy industry reached 9.27 billion USD, exhibiting a remarkable compound annual growth rate (CAGR) of 16% over the past four years (LEGO Group 2022). The industry's expansion has been primarily propelled by the rise of video and computer games in the last three decades. Among the toy companies, LEGO holds the leading position, surpassing the top four industry leaders that generate 25% of the industry's total revenue. The remaining 75% market share is fragmented among various players. LEGO stands out with its unique product concept, which has effectively prevented the emergence of similar competitive products. This characteristic of LEGO's offerings makes it an intriguing context for studying new product launch decisions, without the need to account for competitive responses. In recent years, the company has experienced remarkable growth rates surpassing the industry's average. However, LEGO's journey began as a small family-owned firm and encountered several existential challenges over the past three decades. Adapting to changes in the industry and facing various contingencies, the company ventured into different product decisions to foster growth. For this study, our focus is on LEGO's mainstream construction brick products. We do not consider other new product segments such as movies, theme parks or electronic games.

4.2 Data

We collected 20,908 products data from Rebrickable, a platform that provides building instructions both for LEGO and user-generated products. We also gathered 17,000 product-performance data from Bricklink[3], the largest online community and marketplace for adult fans of LEGO, with about 1.1 million members in 70 countries, 600 million items for sale (new and second-hand) and 10,000 resellers. The data contains information about products available at resellers, offering newly launched and *antique* LEGO products in high demand. We verify the accuracy of product category data with information from LEGO website and BrickInsights, a website that offers analysis on LEGO products.

After cleaning the data, we were left with 12,861 observations at product-level, from year 1949 to year 2023. Table 2 displays an overview of the raw data. Each year has at least one product launch, except years 1950 to 1953, for which we do not observe product launches in our dataset. Moreover, we do not observe the exact date in which a product was first launched, but we do know the year. This means that we may miscalculate the order of launches within one year. Fortunately, we do have a product number ID that grows in time, so, within each year, we sort launches by their product ID. In our robustness section, we replicate the analysis with a random order within each year, and results are extremely similar (cf Appendix E.1).

Figure 3 illustrates the growth trend of category launches over the years. The number of category launches experienced growth from 1960 to 1970, remained stable for a decade, then saw a resurgence in the 1980s, and finally exhibited exponential growth after year 1990. In Figure 4, we can observe the growth of the top hexadecile categories, which refers to those categories with the top 5% of total products launched over multiple years. Notably, categories such as LEGO Duplo and Universal Building set have maintained their status as best-selling categories for over 50 years since their initial launch. Other categories like Educational and Dacta (including Mindstorms and NXT products), Town, and Technic have also consistently dominated the product category launch pattern since their inception. More recently, LEGO Ninjago emerged as a top-performing category in the 2010s. However, LEGO Starwars has only achieved top-category status in this current decade. Finally, categories such as Marvel and Harry Potter fall into hexadecile 2 and hexadecile 3, respectively, and are grouped under the category *Other Hex*.

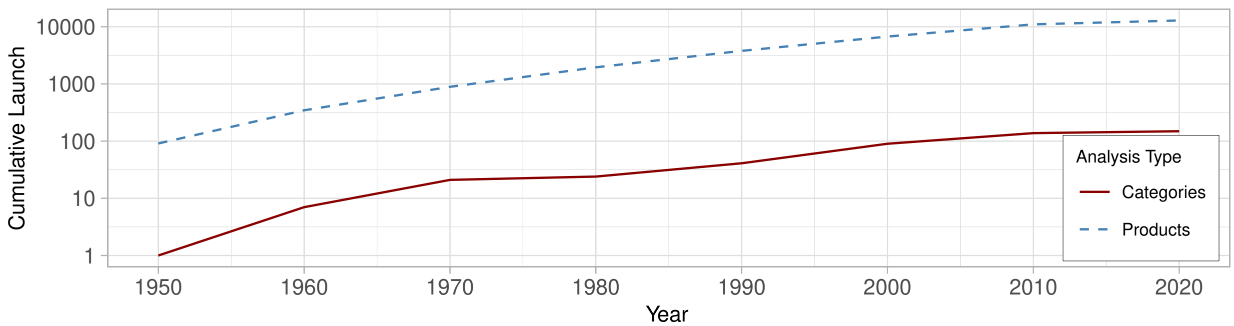
We aggregate data at the category level to have a finite set of product groups from which new products are selected, thereby transforming the raw data set into a pseudo-panel (Cameron and Trivedi 2010). This panel has 1,074,281 observations, many for each of the 12,861 launch events,

[3] LEGO acquired Bricklink in 2019 and justify this purchase by its willingness to consolidate its engagement with its adult fan community.

Table 2: A sample of raw data

Year	Category	Product Name	Product ID	Volume	Mean Price
1969	System	Lighting Device	050-2	1	35.00
1969	System	Two Garage Door	065-1	4	30.95
1969	Universal Building	Super Set	088-1	3	1220.96
1969	Legoland	Super Value	102-3	2	247.22
1969	Universal Building Set	4.5V Motor with Tracks	103-1	16	50.05

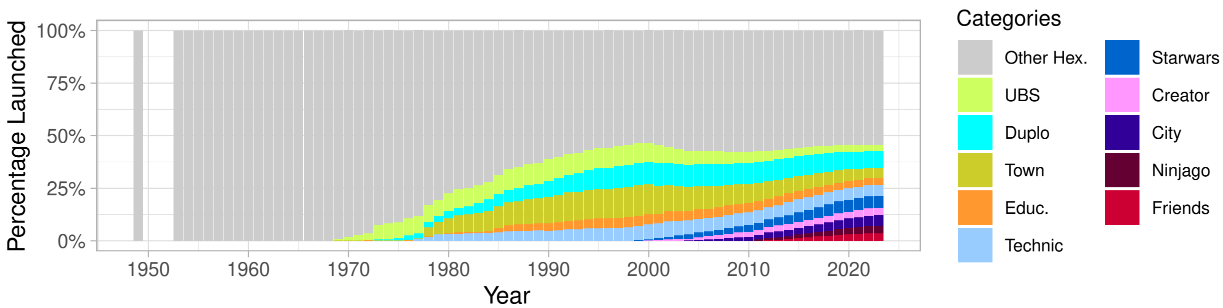
Figure 3: Products and Categories Growth Curve



which we call periods (fewer in the 1950s and many more in the 2010s). For each category-period, category features are computed. Categories selected for launch are observed in each period and the outside option of unobserved never launched categories are lumped under the *Category0* label. Before partitioning the dataset, we have 150 unique categories, with 149 launched categories and *Category0*. After a category is launched for the first time, it is labelled with the actual name of that category, as discussed in §3.

To undertake our work, we focus on task selection variable, as well as its theoretical predictors, which are past experience and performance. Variable $Selected_{it} = \mathbf{1}_{y_t=i}$ denotes whether category i is observed for product launch. Experience is the cumulative count of products launched with a category. This metric is transformed into $MarketShare_{it}$, by dividing the count by the number of periods since a category is first launched. Variables $Volume_{it}$, $Price_{it}$ and $Revenue_{it}$ are proxies for product performance. $Volume_{it}$ is the wholesale volume of units available for each product at resellers around the world, plus total units sold over the past 6 months, measured at the time of data collection – averaged across all products launched within the category up to period $t-1$. This is measured in units available for sale in secondary markets. (We replicate the study using only available units or only sold units, and obtain nearly identical results, cf. Appendix E.2 and E.3.) $Price_{it}$ is the average quoted unit product price in Euros, again averaged across all products in the category up to $t-1$. Finally, $Revenue_{it}$ is equal to $Price_{it}$ multiplied by $Volume_{it}$. Finally, we use the variable $Novel_{it} = \mathbf{1}_{y_t=0}$ to indicate whether *Category0* is selected for launch. Note that we need both variables in the model to separate the effect of the first launch vs. the later ones, captured by $Novel$, and the effect of the difference between a new category and the rest, through the inheritance process by which the new category takes the value of $Category0$ (these two distinct

Figure 4: Total Products Launched in Top Hexadecile Categories in each year.



roles are apparent when comparing Tables 19 and 20 in the Appendix).

Table 3: A sample of the pseudo panel data

Period	*Category*	I_t	$Selected_{it}$	$Novel_{it}$	$Market\_Share_{it}$	$Volume_{it}$	$Price_{it}$	$Revenue_{it}$
289	Category0	0	1	1	0.00	0.00	0.00	0.00
289	System	1	0	0	0.92	19.70	105.44	888.69
289	Samsonite	2	0	0	0.01	11.50	40.35	362.28
289	Train	3	0	0	0.08	34.23	127.39	1339.87
290	Category0	0	1	1	0.00	0.00	0.00	0.00
290	System	1	0	0	0.91	19.70	105.44	888.69
290	Samsonite	2	0	0	0.01	11.50	40.35	362.28
290	Train	3	0	0	0.08	34.23	127.39	1339.87
290	Universal Building Set	4	0	0	0.00	3.00	1220.96	3662.88
291	Category0	0	0	1	0.00	0.00	0.00	0.00
291	System	1	0	0	0.91	19.70	105.44	888.69
291	Samsonite	2	0	0	0.01	11.50	40.35	362.28
291	Train	3	0	0	0.08	34.23	127.39	1339.87
291	Universal Building Set	4	1	0	0.00	3.00	1220.96	3662.88
291	Legoland	5	0	0	0.00	2.00	247.22	494.45

Table 3 provides a sample of data with four categories from periods 289 to 291, corresponding to year 1969. In $Period = 289$, $Category0$ is selected for launch. Thus variable $Selected_{it} = 1$. For this new category, $Novel_{it} = 1$ and the rest of covariates (market share, volume, price and revenue) are set to zero, by construction. Category $System$ has the highest market share value of 0.92. Products in this category sells at an average of 19.7 units and an average price of 105.44 EUR, which yields a category revenue of 888.69 EUR. In $Period = 290$, it is revealed that the new category launched in the previous period is named $UniversalBuildingSet$. Thus, the number of available categories grows from $I_t = \{0, 1, 2, 3\}$ to $I_t = \{0, 1, 2, 3, 4\}$. In addition, $Category0$ is selected for launch for a second consecutive time. We observe in period 291 that new category launched the previous period is named $Legoland$. Hence, we have a total of 6 categories in this period.

4.3 Descriptive Statistics

To undertake our analysis, since variables are highly skewed, we apply a log-transformation so that its distribution is bell-shaped, i.e., we transform them with $\log(1 + \cdot)$ to avoid problems with zeros. We report the descriptive statistics for the log-transformed variables in Table 4. Note that each variable has a minimum value of zero because of Category0. The variable $log\_MarketShare_{it}$ has a slight positive skewness, and further analysis regarding its distribution can be found in Appendix B. A visualization of the density of variables and their correlation matrix can be found in Appendix C.

Table 4: Descriptive Statistics

	Mean	Std Dev.	Minimum	pc25	Median	pc75	Max
$log\_MarketShare_{it}$	0.012	0.026	0	0.00088	0.0028	0.0096	0.69
log_Volume_{it}	3.757	1.068	0	3.12888	3.6376	4.2790	8.76
log_Price_{it}	4.112	0.916	0	3.50288	4.0835	4.6936	8.22
$log\_Revenue_{it}$	7.869	1.508	0	6.96864	7.7993	8.7043	12.17

4.4 Structural model estimation

We undertake the parameter estimation with the same specification as above, by simulating candidate models in Table 1. Maximum Likelihood Estimation is undertaken with a differential evolution algorithm (Mullen et al. 2011), in which we look for improvements on the likelihood until convergence. Hyperparameters values are constrained as follows: mean $\mu \in [-5, 5]$; variance $\sigma_\theta \in [0, 326]$, observation noise $\sigma_\epsilon \in [0, 326]$; $\beta \in [0, 2]$. Models are trained in-sample over $0 \leq t \leq 350$ launch periods from the beginning. We chose 350 periods because the ground truth search is computationally intensive, with several thousands of points evaluated. In the robustness section, we train the model over 1,000 periods and obtain similar results.

The simulation requires to specify a product performance function, which is the function the agent seeks to maximize. We chose $Revenue_t$ as the dependent variable for product performance. Indeed, the agent is most likely to seek to increase both price and inventory sold than price or inventory only. Note that each simulation starts with its prior being one of the point in the grid. When simulation for each model is completed, we select parameters from the best model to compute predictions both in-sample and out-of-sample for period 351 to $T = 12,861$. We compare models all candidate models against our benchmark Model 0, the MNL model with the same covariates. Our main model incorporates past experience and performance variables, with current revenue serving as the predicted performance variable. Next, we present the estimation results of the various models.

5. Analysis

5.1 Estimation results

Our inference results are shown in Table 5. Note that the variable $log\_Revenue$ is dropped from the models because it is just the sum of log_Volume and log_Price. Instead, $log\_Revenue$ is used as the objective to maximize when evaluating decisions in all Bayesian models.

Table 5: Inference results. Training sample includes product launches in periods 1-350

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	3.17	3.17	4.87	-3.93	4.95	4.94
	(0.29)	[3.17, 3.17]	[3.85, 5.83]	[-40.10, 30.98]	[4.71, 5.20]	[4.51, 5.39]
log_Volume	0.29	0.29	-0.07	-0.84	0.17	0.37
	(0.15)	[0.29, 0.29]	[-2.50, 2.41]	[-3.26, 1.72]	[-0.02, 0.37]	[-0.29, 1.05]
log_Price	0.7	0.71	2.96	4.89	2.04	2.37
	(0.17)	[0.71, 0.71]	[-34.76, 37.42]	[-2.60, 12.90]	[-31.07, 33.20]	[-9.93, 15.63]
Novel	2.74	2.74	-0.82	4.85	-2.43	-4.68
	(1.15)	[2.74, 2.74]	[-16.37, 14.10]	[-4.18, 13.93]	[-8.68, 3.82]	[-13.65, 4.44]
Noise (σ_ϵ)			18.91	7.68	13.72	18.12
Bonus (β_{UCB})				0.31		
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-181.94	-175.95	-183.09	-186.43
AIC		380.86	371.88	361.9	376.18	382.86
Observations	1290	1290	1290	1290	1290	1290
Simulations per period						1,000

Our first observation is that, in sample, Bayesian models provide a slightly better fit, and hence are more plausible than reduced-form and no-belief-update models, as shown by the lower Akaike Information Criterion (AIC), except for Thompson Sampling; we discuss later goodness of fit out of sample. Two of the models perform particularly well: Greedy (model 2) and UCB (model 3) provide the two highest likelihoods in the training sample (periods from 1 to 350).

5.2 Parameter estimates

Static predictions (models 0/1). Interestingly, the reduced-form model (model 0) and the no-belief update model (model 1) result in identical parameter estimates and log-likelihood values. This implies that a non-Bayesian Softmax model keeps the same preference structure for the different arms over time, and is hence equivalent to the MNL in terms of choice probabilities. Note that while the MNL is able to provide confidence intervals for the estimated coefficients, which are obtained from the Fisher Information matrix, in the Bayesian approach, the uncertainty about the parameter is set to zero because we force the parameters to stay the same given that there are no belief updates. Moreover, model 1's fit remains unaffected by the choice of predicted performance variable (revenue in this case), which is never used in this model. This observation implies a crucial insight: the MNL model only estimates static parameter values, thereby ignoring the underlying mechanisms driving the decision-making process and hence limiting its effectiveness in capturing actual launch patterns. This cannot be mitigated by including temporal trends or interactions, as discussed in Appendix D.5.

Regarding estimates, first, we observe a strong and positive parameter estimate for past experience $log\_MarketShare$. Specifically, a 1 percentage point increase in the category market share leads to about an increase of 3% ($3.17 \times [\log(1 + ms + 0.01) - \log(1 + ms)] \approx 0.0286$ when ms is small) in the category attractiveness, hence approximately 3% in the probability of being selected for launch; the effect is smaller for larger categories, i.e., when ms is large. In other words, there is a strong path dependency where past launches are an indicator of future launches. This finding is consistent with prior research, which has highlighted the importance of past experience as a critical determinant of performance.

Second, both volumes and past prices are also good predictors of category selection, albeit with a smaller effect size. The findings indicate that higher volume or price are associated with a greater likelihood of category selection. Since these are products launched before 1970, this suggests that in the early days, LEGO was increasing variety in categories that had shown potential to generate higher profits, via volume or contribution margin.

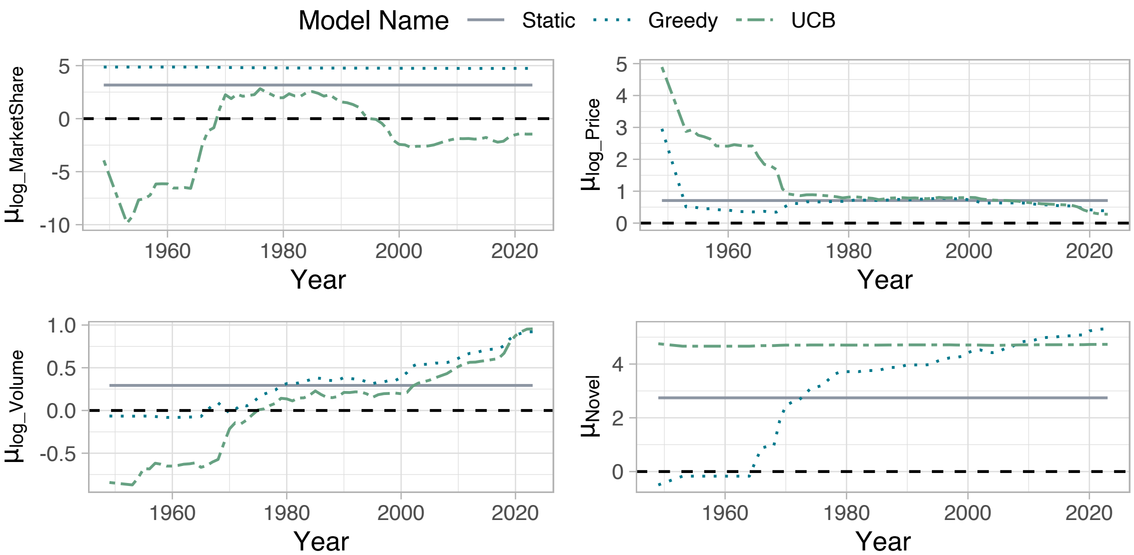
Third, the coefficient of $Novel$ shows that category novelty is a significant predictor of category launch, with a new category being 15 times ($e^{2.73} = 15.48$) more attractive than an existing category with a small market share, volume and price (so that $log\_MarketShare = log\_Volume = log\_Price = 0$).

If this model was correct, this finding would imply that, in the early days, there was a premium associated with introducing new categories, which is intuitive given that the toy industry rewards differentiation in general, and uniquely positioned firms like LEGO in particular. Overall, our results for models 0 and 1 indicate that past experience, performance, and category novelty are reliable predictors for category launch.

Greedy (model 2). Model 2 ranks as the second best model in terms of in-sample fit. This is coherent with recent theoretical results (Bayati et al. 2020 show that greedy algorithms can be asymptotically optimal). In comparison with static models, the Greedy model has similarly large parameter estimate for $log\_MarketShare$, but puts much more weight on price. This result underscores that greedy actions occur through variables such as $log\_MarketShare$ and log_Price. Moreover, the model penalizes categories with high volume sales (log_Volume) and novelty. Large volumes imply that a category is no more niche and is less likely to carry a performance bonus. A penalty for novelty means that it is costly not to be greedy. Most importantly, prior values for log_Price and $Novel$ carry a lot of uncertainty, in the form of a large variance in the Bayesian hyperparameters, especially for price (starting standard deviation of about 35). In other words, the model recognizes that parameters are highly uncertain initially, and will wait for more experiments to refine the parameter value. As a matter of fact, when simulating the model, we see that the mean value for log_Price decreases, moving from 2.96 in period $t = 1$ to 0.55 in period $t = 350$, while that of $Novel$ increases from -0.82 to 2.46, see Figures 5. Therefore, though log_Price seems to pay off immediately, its contribution to performance wears off rapidly. Similarly, albeit it is costly to create new categories, this decision is also rewarded in the future. Furthermore, though Category0 fixed effect seems to start high at value 1.14 (given that the effect of $Novel$ is negative), it receives a positive boost, with a value of 1.45, after new categories reveal that novelty generates a positive reaction from the market. Since the greedy model has a better fit than the static ones, passive learning seems to be capturing the changing preferences of LEGO for new products prior to year 1970.

UCB (model 3). UCB provides the best fit for interpreting the LEGO launch patterns prior to year 1970. This suggests that LEGO is more likely to use a Bayesian knowledge management rule and a choice rule guided by the active learning policy UCB. In other words, it operates on the principle of optimism under uncertainty. We find that β_{UCB}, the exploration bonus, has a value of 0.31, meaning that increasing hyperparameter's standard deviation by one is equivalent to increasing its average by about 1/3. This is especially important given that model uncertainty

Figure 5: Evolution of average belief value



is quite high for $log\_MarketShare$, and results in having LEGO explore categories with higher uncertainty. These include all those with few launches, because the starting category fixed effect uncertainty is very high $\sigma_{Category0,t} = 18.93$ and $\sigma_{Novel,t} = 4.52$ at $t = 1$. Note that parameter estimates for $log\_MarketShare$ and log_Volume are lower than Static and Greedy models, but one should keep in mind that an exploration bonus is added to these average parameters when predicting launch probabilities, e.g., at $t = 1$, while the average belief for $Novel$ is 4.84, in the prediction we use $4.85+0.31\times 4.85 = 6.35$. In other words, in the early days, LEGO experimented in 'young' categories more than what a Greedy approach would suggest, thereby launching products in categories with lower expected revenue, for the sake of gaining knowledge on the revenue distribution.

We can also observe that parameter uncertainty is much larger than that of the Greedy model for only $log\_MarketShare$, hence revealing the ambiguous value of past experience. In contrast, the noise standard deviation σ_ϵ is lower, showing that under UCB, LEGO attributes variability in revenue to hyperparameter uncertainty, which can be reduced via experimentation and learning. Unexplained uncertainty is then captured in σ_ϵ. In contrast, greedy has less 'faith' in the model, i.e., it starts with high model uncertainty for several variables, and admits a larger uncontrollable uncertainty.

Gittins Index (Model 4). This model returns estimates close to the Greedy model, with the same coefficient signs and a slightly worse fit. In addition, it also has lower hyperparameter uncertainty for all coefficients but for log_Price. As a result, the model has the second lowest value for noise variance σ^2_ϵ, after UCB. Finally, fit values and coefficients for Gittins Index and UCB are different. However, Russo (2021) shows that there is an asymptotic equivalence between the Gittins Index and Bayesian UCB. We find that fit values for both models are approximately the same in models without context-dependent covariates (cf. table 19 and 20). Nonetheless, parameter values are never the same. Note that the Gittins Index has been computed with a discount factor close to 1 (cf. Appendix A), thus assuming a sufficiently patient agent. This implies that LEGO is far from being a sufficiently patient agent.

Thompson Sampling (Model 5). Finally, Thompson Sampling is the worst-performing of the Bayesian models. Parameter estimates uncertainty are of the same magnitude as those of the Gittins Index model, except for the $Novel$ covariate, whose parameter value is strongly negative. Kalvit and Zeevi (2021) shows that Thompson Sampling exhibits an "incomplete learning" effect, in which it continues to sample suboptimal performing arms and fails to generate an asymptotic normal sampling distribution, especially in settings where the rewards of the arms are not very different (non-separated arms). This result raises question on the possibility of doing inferential work with Thompson Sampling (Kalvit and Zeevi 2021). Furthermore, it is worth mentioning recent theoretical analyses showing that settings with context-dependent information require better algorithms than Thompson Sampling or UCB (Van Parys and Golrezaei 2020). Our results provide evidence that, unlike UCB, Thompson Sampling is less likely to use structural information effectively. Finally, (Qin and Russo 2022) show that the performance of Thompson Sampling is negatively impacted in settings where the sequence of context influences arm performance. These results support our evidence that Thompson Sampling has the worst goodness of fit.

5.3 Model Comparison

We next evaluate the models with out-of-sample predictions and discuss model selection using the Watanabe information criterion (WAIC).

Out of sample predictions. We use the evaluation dataset, covering the launching period from 351 to 12,861, to simulate models with priors from the UCB model since this is the most likely learning model of LEGO; in the robustness section, we replicate the results with priors obtained with Greedy (model 2) and Static (models 0/1) and obtain similar results.

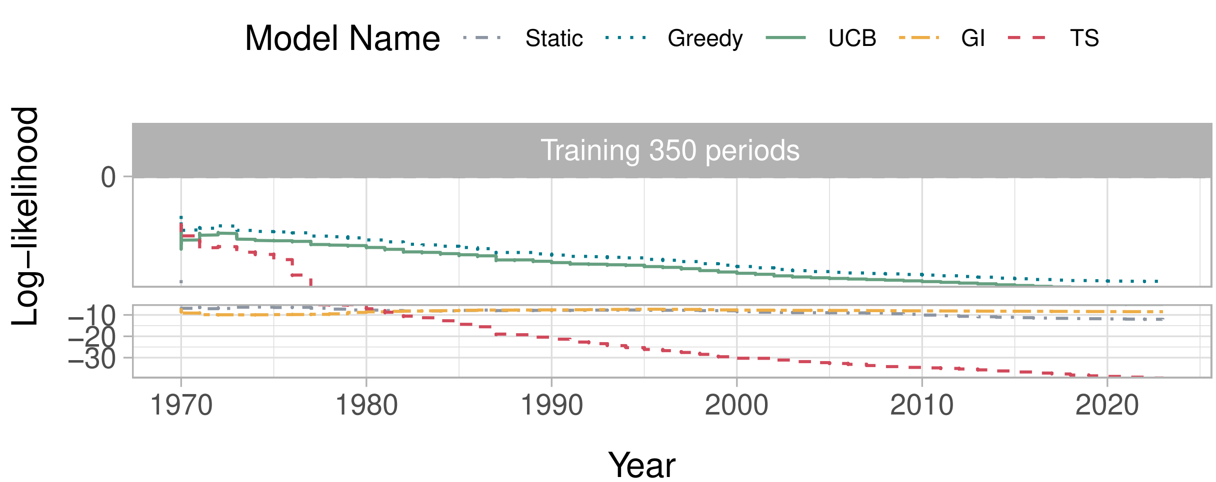
Figure 6 illustrates the dynamics of log-likelihood of model predictions. We report cumulative log-likelihood, defined as total log-likelihood between 1 and t, divided by the number of periods t. We observe that the Greedy and UCB models provide superior fit, with Greedy outperforming UCB. The Static model ranks third in terms of prediction accuracy. On the other hand, the Gittins Index model shows a very poor out-of-sample fit. Finally, Thompson Sampling initially demonstrates competitive predictive performance compared to all other models except Greedy. However, its performance rapidly deteriorates, and the model eventually performs worst among all models.

To better understand the behavior of these out-of-sample fit performance, we consider the number of existing categories in different periods and compare the summary statistics of model probability predictions to LEGO's. Tables 6 and 7 provide details of this comparison by focusing on periods 351 to 1,000 (where there are 22 or fewer categories) and periods 12,000 to the last period (where there are up to 150 categories), respectively. The minimum and 25-percentile probabilities show the predicted frequency of least likely options, whose values are close to zero. As more categories are available, these frequencies decrease further. The 50-percentile and maximum probabilities on the other hand are larger, up to 97% (for model 1) in the early periods, and also decrease as more options become available (see Table 7). Mean and median predictions reflect the equi-probable prediction over the existing choices ($1/n$ probability: 0.111 when there are 9 categories in period 351, 0.045 when there are 22 categories in period 1000, and 0.0066 when there are 150 in the last period).

We can make the following observations based on the information presented in the tables. Firstly, the maximum probability and standard deviation for the Static (model 0/1) closely match LEGO's actual choices summary statistics. This indicates that the Static model attempts to replicate the decisions made, although it does not accurately represent the underlying probability distribution generating those choices. Secondly, the summary statistics for Greedy and UCB probabilities are the most similar. This is another way of saying that predictions from both models are close, as displayed in Figure 6. Thirdly, both the Gittins Index and Thompson Sampling exhibit very low probability values below the 50th percentile. These values decrease to zero as the number of arms increases, as shown in Table 7. Consequently, both policies, compared to UCB and Greedy models, seem inadequate in settings with growing number of options. Additionally, the summary statistics for the Gittins Index closely resemble those of the Static model, showing that this model does not effectively learn from the data. Lastly, predictions above the 50th percentile for Thompson Sampling fall within the same range as those of UCB and Greedy policies. This confirms the favorable performance of the Thompson Sampling policy when the number of arms is low.

In conclusion, the out-of-sample predictions provides evidence that the Greedy model is the most plausible model to explain LEGO's data.

Figure 6: Out of Sample Predictions



WAIC. Table 8 displays the computed Watanabe information criteria on the same evaluation dataset as before. To provide a metric per period, we present the mean WAIC, calculated by dividing the total WAIC by 12,511, which represents the length of the validation sample (12,861 minus 350). Alongside the mean-period WAIC, we provide the number of effective parameters ($pWAIC$), its standard error, and the Akaike weight. The results demonstrate that the Greedy model outperforms all other models, exhibiting an Akaike weight of 86.7% and the lowest $pWAIC$ value among the effective parameters. It is worth noting that the Model 1 to Model 5 estimate each category with a fixed effect, resulting in 150 fixed effect parameters when all categories are revealed. When combined with other covariates and non-belief parameters, the maximum number of estimated parameters reaches 156. Interestingly, the Greedy model excels in reducing model complexity, as evidenced by its lower $pWAIC$ value of 49.77. This indicates that, with the same 156 parameters, the Greedy model effectively exploits the correlation structure between product categories, leading to more stable posteriors. The UCB model ranks as the second-best model with

Table 6: Predictions for simulated out-of-sample decisions in periods 351-1,000

	Actual	Model 1	Model 2	Model 3	Model 4	Model 5
	LEGO	Non-Bayesian Softmax	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
Min. probability	0.0000	0.0000	0.0016	0.0007	0.0000	0.0000
Pc25 probability	0.0000	0.0000	0.0253	0.0203	0.0001	0.0070
Median probability	0.0000	0.0005	0.0495	0.0396	0.0011	0.0450
Mean probability	0.0616	0.0616	0.0616	0.0616	0.0616	0.0616
Pc75 probability	0.0000	0.0050	0.0790	0.0916	0.0089	0.0990
Max probability	1.0000	0.9702	0.2950	0.3864	0.9924	0.3060
Std Deviation	0.2404	0.1834	0.0508	0.0581	0.1967	0.0615
Number of existing categories			9 -22			

Table 7: Predictions for simulated out-of-sample decisions in periods 12,000-12,861

	Actual	Model 1	Model 2	Model 3	Model 4	Model 5
	LEGO	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
Min. probability	0.0000	0.0000	0.0001	0.0001	0.0000	0.0000
Pc25 probability	0.0000	0.0000	0.0009	0.0007	0.0000	0.0000
Median probability	0.0000	0.0000	0.0023	0.0019	0.0002	0.0000
Mean probability	0.0067	0.0067	0.0067	0.0067	0.0067	0.0067
Pc75 probability	0.0000	0.0002	0.0070	0.0057	0.0008	0.0040
Max probability	1.0000	0.6350	0.0751	0.1015	0.6236	0.1780
Std Deviation	0.0816	0.0520	0.0116	0.0133	0.0509	0.0189
Number of existing categories			148 - 150			

an Akaike weight of 13.3%. However, with a $pWAIC = 110.3$ value, it exhibits a higher number of effective parameters, which indicates a greater degree of model complexity. Thompson Sampling and the Gittins Index show similar WAIC values, although Thompson Sampling is more complex than the Gittins Index. Lastly, the Static model demonstrates the highest WAIC value but lower model complexity compared to Thompson Sampling and the Gittins Index.

To conclude, several metrics on the evaluation set shows that Greedy is the best model in terms of prediction LEGO's future launches.

5.4 Robustness checks

We perform additional analyses to evaluate the robustness of our results. For details, we include all supporting materials in Appendix D.

Longer training sample size. We replicated our analysis with a training period of 1,000 periods, instead of 350. The results are shown in Table 10. We confirm that UCB provides the best in-sample fit and the Greedy model the best out-of-sample performance, albeit with a small gap in comparison to UCB.

Other performance variables. We estimate models with alternative measures of product

Table 8: Watanabe Information Criteria (WAIC)

	Model 1	Model 2	Model 3	Model 4	Model 5
	Non-Bayesian Softmax	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
WAIC	950,441	148,491	195,465	863,209	868,745
Mean period WAIC	75.97	11.87	15.62	69	69.44
dWAIC	64.1	0	3.75	57.13	57.57
Standard Error	0.03	0.04	0.03	0.02	0.04
Akaike Weight	0.0%	86.7%	13.3%	0.0%	0.0%
pWAIC	268.7	49.77	110.3	332.83	343.75

performance: instead of $log\_Revenue$, we consider log_Volume and log_Price as the metric to maximize. We replicated our results in-sample, and find that UCB provides the best fit with log_Volume, while Greedy does so with log_Price. Estimated parameters, shown in Tables 11 and 12, remain similar to Table 5's. However, the specification with $log\_Revenue$ reported earlier provides the overall best fit.

Parameter recovery. To ensure that data generated with a specific model can also be reverse-engineered to retrieve the same model, we attempt to recover parameters of simulated model by ensuring we obtain the same model that generated the data. We find that the Greedy model exhibits a remarkable capability to accurately represent data generated by other policies. This finding can be attributed to the existence of regime changes in behavioral models across different time horizons, where the Greedy models consistently capture a behavior that persists over long periods. We discuss this further in Appendix D.4.

Other robustness checks

Alternative ground truth. We replicated our out of sample study starting with the priors identified by Greedy and Static. Figure 14 and 15 show the predictions performance, which need to be compared to Figure 6. We find that, regardless of the ground truth, the UCB model never outperform the Greedy model, suggesting that it is more adaptable than UCB.

Trends and forgetting. Since category launches exhibits a curvilinear growth rate over the years, we included trends in the different models, as well as a limited memory factor in (3b), to give more weight to recent observations. Table 17 includes the results. As obtained with our main specification, we find that the UCB model remains superior to remaining models, followed by the Greedy model.

Fixed effect and novelty only. We also tested specifications without structural parameters ($log\_MarketShare$, log_Volume, log_Price) and only fixed effect for product categories (cf. Table 19). We further consider models with novelty, i.e., the $Novel$ covariate (cf. Table 20). On the one hand, in both cases, we find that, in-sample, Thompson Sampling explains best the data, although its Log-likelihood value is worse than in our main specification. However, as in the main specifications, its predictions on the evaluation dataset degrades to become worse than remaining policies. On the other hand, the Gittins Index has the same fit value and predictions performance than remaining models, including the Static model.

6. Conclusion

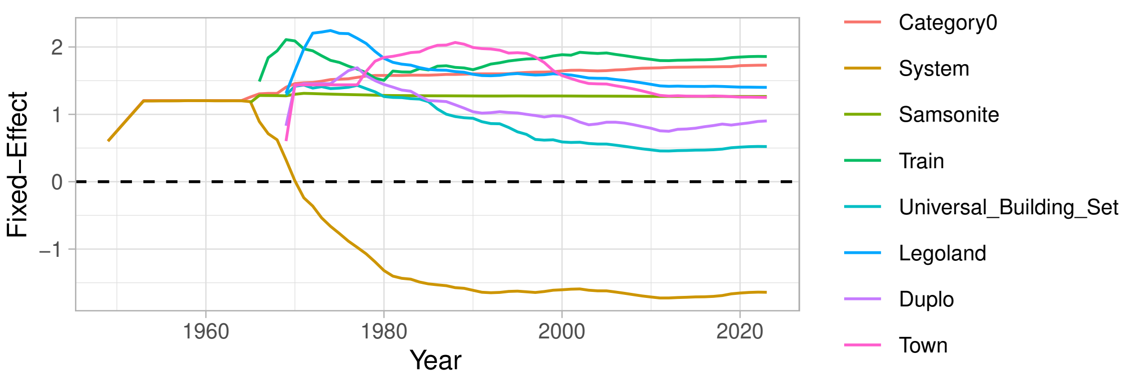
In this work, we aim to elucidate the learning process of a firm about its business environment. Our study is especially relevant when we consider a long time horizon over which consumer preferences evolve. We postulate that this learning process is revealed from the observation of a firm's new product launch decisions and the corresponding commercial performance of those. We create a model that includes both a knowledge and a choice rule, which will help the firm balance its goal of extracting higher performance from existing product categories while also exploring newer ones to renew the product offering. We represent the firm's knowledge with a sleeping linear bandit to account for newly available categories and the inherent structure in the product space. With this framework in mind, we can test alternative rule configurations and then reverse-engineer the best combination that explains the firm's data in the most plausible way.

We evaluate our framework with data from a leading toy design firm, LEGO, over a long history of product launches from the 1950s until now. On the one hand, our analysis provides evidence that LEGO is most likely to have been using either UCB, which gives low confidence to initial belief values and seeks novelty, or Greedy, which exploits existing knowledge to seek immediate performance. Both models have similar fit performance, similar parameter estimates with respect to past performance, but different uncertainty values associated with them. These models are much better than static, non-anticipatory models, i.e., those that only use information from the past, as well as less exploratory models such as Thompson Sampling and the Gittins Index, which does not reflect the increase in new categories experienced at LEGO. Interestingly, UCB and Greedy models are sufficiently adaptive so that there is no further need to discount past observations. Hence, our results indicated that LEGO has developed a feedback mechanism between the observed performance of new products, and its future launch plans. Our framework allows us to uncover this relationship, by rationalizing the launch decision as a comparison between options with different (random) rewards, such as category revenue. On the other hand, we cannot state that LEGO actively explores new product spaces via active learning –with UCB–, since it is as plausible as passive learning –with greedy. In addition to the general structure of the learning strategy, our model also reveals two aspects of adaptation.

First, category heterogeneity is of the utmost importance. Indeed, the belief associated with a given category performance varies wildly across categories, and moreover, it also dynamically evolves. Figure 7 plots the dynamics of mean belief (aggregated in each year) associated with category intrinsic features over the years for the first seven launched categories, which we compare to the intrinsic value of opening a new category.

For example, the firm first launched the category $System$. From the year 1949 to the year 1964, $System$ had a slightly lower aggregated fixed-effect than $Category0$ at a value around 1, which, after adding the rest of covariates, led LEGO to launch new $System$ products. Then, $Samsonite$ was launched and updated to a slightly better category mean value, while the belief value for category $System$ quickly decreased and hence LEGO ceased launching the latter. Subsequently, category $Train$ was launched and enjoyed a strong and increasing category worth, which later decreased in the 1980s. These bifurcations are an example of *cannibalization* across categories, due to changes in the relative values in the options. As we can see, categories with higher values are predicted to be launched more often. Another relevant observation from the figure relates to the *innovation appetite*, which is materialized with an updated value for $Category0$, which steadily increased from the 1960s until now. This led the firm to launch many new categories, including *Legoland*, *Duplo*

Figure 7: Time-Series of category fixed effects, aggregated by year. Table 24 and 25 in Appendix E.4 zoom on the periods 280 to 300 to illustrate how new categories inherit the fixed effect of Category 0.



and *Town*. *Legoland* and *Town* were very successful (the first one up until its peak in 1975, and the second up until 1988), while *Duplo* did not do as well. This demonstrates how our framework allows us to visualize the relative value of each category, in comparison with other categories, as well as the option of opening a brand new category.

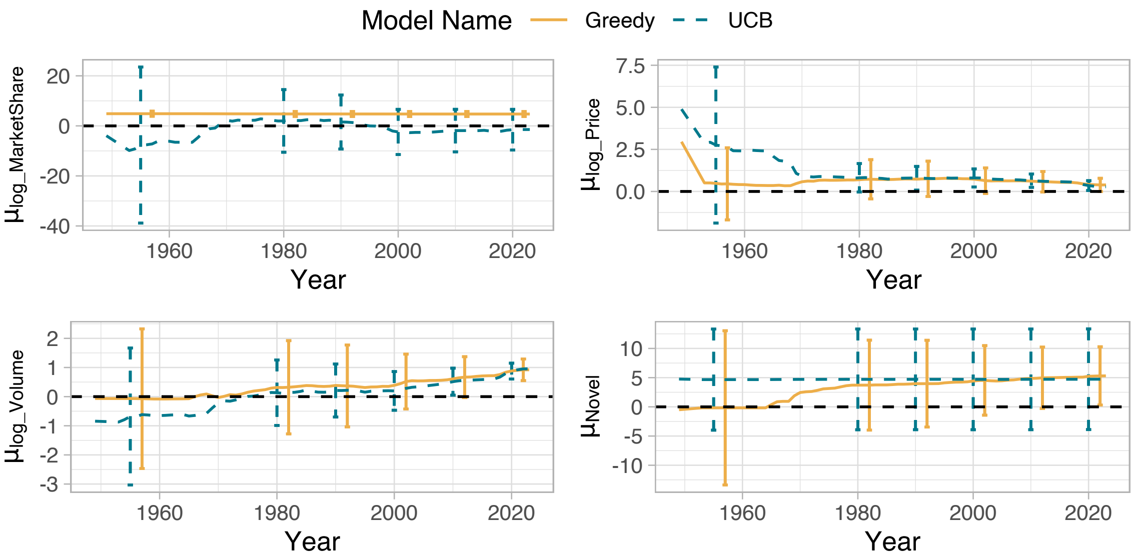
Second, our model also incorporates context-dependent covariates, that describe the observable differences between options. Besides making the model richer by controlling by category differences, this approach has a secondary effect: it allows to *transfer learning* across experiments. Namely, since each category has certain past performance covariates, it can learn through them from launches of other categories.

Specifically, contextual learning is conducted by measures of past experience, which, under passive learning, has a stable and positive effect, so that categories with higher past market share will be more likely to be chosen in the future. Interestingly, as shown in Figure 8, under UCB, the model reflects high uncertainty about this variable, which is considerably reduced over time.

Contextual learning is also taking place through the effect of past volume sales and past price. We observe that both the Greedy and UCB models initially detect very high uncertainty about the role of these past performance measures, and quickly learn about them. More expensive categories, which received a large bonus initially and led to more launches, later resulted in lower performance, which made LEGO learn that the impact of price was not as large as originally thought. This meant that the posterior of this variable was reduced from a high value (2.95 and 4.89 for Greedy and UCB respectively) in the 1950s to a value close to zero in the 2020s. More strikingly, the role of volume changed signs, from a negative to a positive value. In other words, LEGO initially gave priority to more niche categories with smaller number of units sold, but later started giving a bonus to larger volume ones.

In summary, the combination of a conceptual framework, its operationalization in a bandit model, and its application to a set of actual product launches demonstrates the potential of dynamic learning perspectives to better understand real new product introduction patterns. This is in contrast with most of the literature in innovation management, which tends to employ reduced form specifications. We hope that our work will spark interest in considering observable or latent (as in this paper) knowledge states.

Figure 8: Evolution of belief confidence bound for Greedy and UCB models.



To conclude, we study decisions as an inference of a learning process, in which the blackbox of the underlying structure of unitary experiences (consisting of a choice and its associated performance) is opened. We find that the Greedy model provides the best representation of decision data of LEGO, a leading toy product firm. Through our framework, we gain insights into the natural selection process among product categories. Specifically, we can visually assess the significance of category heterogeneity by analyzing the variation in category fixed-effects. Additionally, we observe the occurrence of cannibalization as new categories emerge. Furthermore, we quantify the level of innovation appetite, which reflects the value of untapped innovation that remains unlaunched, as well as the necessary degree of category heterogeneity within the population. Lastly, we employ modeling, estimation, and visualization techniques to examine transfer learning among categories, exemplifying category mutation. For instance, although niche categories initially perform better, over time, categories with larger sales volumes receive greater rewards.

Finally, it is worth mentioning that our work is limited by its very high computational cost. However, it can be replicated in any industry, and future research could address questions related to learning from outside events, such as competitors, or internal learning mechanisms, such as people versus processes related decisions.

Acknowledgements

Victor Martínez-de-Albéniz acknowledges the financial support provided by the Agencia Estatal de Investigación (AEI) from the Spanish Ministry of Science and Innovation, with project reference PID2020-116135GB-I00 MCIN/ AEI / 10.13039/501100011033.

References

Anand J, Mulotte L, Ren CR (2016) Does experience imply learning? *Strategic Management Journal* 37(7):1395–1412.

Anderson A, Kleinberg J, Mullainathan S (2017) Assessing human error against a benchmark of perfection. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 11(4):1–25.

Argote L, Lee S, Park J (2020) Organizational learning processes and outcomes: major findings and future research directions. *Management Science* .

Auer P (2002) Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research* 3(Nov):397–422.

Bayati M, Hamidi N, Johari R, Khosravi K (2020) The unreasonable effectiveness of greedy algorithms in multi-armed bandit with many arms. *arXiv preprint arXiv:2002.10121* .

Benjamin DJ (2019) Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1* 2:69–186.

Bennett VM, Snyder J (2017) The empirics of learning from failure. *Strategy Science* 2(1):1–12.

Besbes O, Gur Y, Zeevi A (2014) Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems* 27.

Brunswik E (1952) The conceptual framework of psychology.(int. encycl. unified sci., v. 1, no. 10.). .

Cameron AC, Trivedi PK (2010) *Microeconometrics using stata*, volume 2 (Stata press College Station, TX).

Caro F, Gallien J (2007) Dynamic assortment with demand learning for seasonal consumer goods. *Management Science* 53(2):276–292.

Cesa-Bianchi N, Lugosi G (2006) *Prediction, learning, and games* (Cambridge university press).

Chao R, Lipson M, Loutskina E (2012) Financial distress and risky innovation. *Saint Mary's College of California, Working paper* .

Chao RO, Kavadias S (2008) A theoretical framework for managing the new product development portfolio: When and how to use strategic buckets. *Management science* 54(5):907–921.

Chapelle O, Li L (2011) An empirical evaluation of thompson sampling. *Advances in neural information processing systems* 24:2249–2257.

Chick SE, Branke J, Schmidt C (2010) Sequential sampling to myopically maximize the expected value of information. *INFORMS Journal on Computing* 22(1):71–80.

Chick SE, Gans N (2009) Economic analysis of simulation selection problems. *Management Science* 55(3):421–437.

Clark JR, Kuppuswamy V, Staats BR (2018) Goal relatedness and learning: Evidence from hospitals. *Organization Science* 29(1):100–117.

Csaszar FA (2018) What makes a decision strategic? strategic representations. *Strategy Science* 3(4):606–619.

Csaszar FA, Levinthal DA (2016) Mental representation and the discovery of new strategies. *Strategic Management Journal* 37(10):2031–2049.

Csaszar FA, Steinberger T (2022) Organizations as artificial intelligences: The use of artificial intelligence analogies in organization theory. *Academy of Management Annals* 16(1):1–37.

Custódio C, Ferreira MA, Matos P (2019) Do general managerial skills spur innovation? *Management Science* 65(2):459–476.

Da Costa L, Lanillos P, Sajid N, Friston K, Khan S (2022) How active inference could help revolutionise robotics. *Entropy* 24(3):361.

Daniel K (2017) *Thinking, fast and slow.*

Denrell J, Fang C, Levinthal DA (2004) From t-mazes to labyrinths: Learning from model-based feedback. *Management Science* 50(10):1366–1378.

Ferecatu A, De Bruyn A (2022) Understanding managers' trade-offs between exploration and exploitation. *Marketing Science* 41(1):139–165.

Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ (2013) The anatomy of choice: active inference and agency. *Frontiers in human neuroscience* 7:598.

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

Gelman A, Simpson D, Betancourt M (2017) The prior can often only be understood in the context of the likelihood. *Entropy* 19(10):555.

Gilboa I, Marinacci M (2016) Ambiguity and the bayesian paradigm. *Readings in formal epistemology*, 385–439 (Springer).

Gittins J, Glazebrook K, Weber R (2011) *Multi-armed bandit allocation indices* (John Wiley & Sons).

Gittins JC (1979) Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)* 41(2):148–164.

Godart FC, Maddux WW, Shipilov AV, Galinsky AD (2015) Fashion with a foreign flair: Professional experiences abroad facilitate the creative innovations of organizations. *Academy of Management Journal* 58(1):195–220.

Gronau QF, Sarafoglou A, Matzke D, Ly A, Boehm U, Marsman M, Leslie DS, Forster JJ, Wagenmakers EJ, Steingroever H (2017) A tutorial on bridge sampling. *Journal of mathematical psychology* 81:80–97.

Han W, Powell WB (2020) Optimal online learning for nonlinear belief models using discrete priors. *Operations Research* 68(5):1538–1556.

Harsanyi JC (1967) Games with incomplete information played by "bayesian" players, i–iii part i. the basic model. *Management science* 14(3):159–182.

Huang Z, Xu Y, Hu B, Wang Q, Pan J (2020) Thompson sampling for combinatorial semi-bandits with sleeping arms and long-term fairness constraints. *arXiv preprint arXiv:2005.06725* .

Kahneman D, Sibony O, Sunstein CR (2021) *Noise: a flaw in human judgment* (Hachette UK).

Kalvit A, Zeevi A (2021) A closer look at the worst-case behavior of multi-armed bandit algorithms. *Advances in Neural Information Processing Systems* 34:8807–8819.

Kanade V, McMahan HB, Bryan B (2009) Sleeping experts and bandits with stochastic action availability and adversarial rewards. *Artificial Intelligence and Statistics*, 272–279 (PMLR).

Kc DS, Staats BR (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management* 14(4):618–633.

Keil T, Posen HE, Workiewicz M (2022) Aspirations, beliefs and a new idea: Building on march's other model of performance feedback. *Academy of Management Review* (ja).

Keskin NB, Zeevi A (2017) Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research* 42(2):277–307.

Kim WC, Mauborgne R (2014) *Blue ocean strategy, expanded edition: How to create uncontested market space and make the competition irrelevant* (Harvard business review Press).

Kruschke JK, Liddell TM (2018) Bayesian data analysis for newcomers. *Psychonomic bulletin & review* 25(1):155–177.

Lattimore T, Szepesvári C (2018) Bandit algorithms. *preprint* 28.

LEGO Group (2022) Lego annual report 2022. URL `https://www.lego.com/cdn/cs/aboutus/assets/blt70ef2efdd8d21dc7/LEGO_Annual_Report2022_Final_WEB.pdf`, accessed on April 9th, 2023.

Levine S, Kumar A, Tucker G, Fu J (2020) Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643* .

Li D, Raymond LR, Bergman P (2020) Hiring as exploration. Technical report.

Li L, Chu W, Langford J, Schapire RE (2010) A contextual-bandit approach to personalized news article recommendation. *Proceedings of the 19th international conference on World wide web*, 661–670.

Loch CH, Kavadias S (2002) Dynamic portfolio selection of npd programs using marginal returns. *Management Science* 48(10):1227–1241.

Mullen K, Ardia D, Gil DL, Windover D, Cline J (2011) Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software* 40(6):1–26.

Musaji S, Schulze WS, De Castro JO (2020) How long does it take to get to the learning curve? *Academy of Management Journal* 63(1):205–223.

Ortner R, Ryabko D, Auer P, Munos R (2014) Regret bounds for restless markov bandits. *Theoretical Computer Science* 558:62–76.

Posen HE, Keil T, Kim S, Meissner FD (2018) Renewing research on problemistic search—a review and research agenda. *Academy of Management Annals* 12(1):208–251.

Powell WB, Ryzhov IO (2012) *Optimal learning*, volume 841 (John Wiley & Sons).

Qin C, Russo D (2022) Adaptivity and confounding in multi-armed bandit experiments. *arXiv preprint arXiv:2202.09036* .

Rusmevichientong P, Shen ZJM, Shmoys DB (2010) Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* 58(6):1666–1680.

Russell SJ (2010) *Artificial intelligence a modern approach* (Pearson Education, Inc.).

Russo D (2021) A note on the equivalence of upper confidence bounds and gittins indices for patient agents. *Operations Research* 69(1):273–278.

Sauré D, Zeevi A (2013) Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* 15(3):387–404.

Slivkins A (2019) Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272* .

Smith R, Friston KJ, Whyte CJ (2022) A step-by-step tutorial on active inference and its application to empirical data. *Journal of mathematical psychology* 107:102632.

Sommer SC, Loch CH (2004) Selectionism and learning in projects with complexity and unforeseeable uncertainty. *Management science* 50(10):1334–1347.

Sommer SC, Loch CH, Dong J (2009) Managing complexity and unforeseeable uncertainty in startup companies: An empirical study. *Organization Science* 20(1):118–133.

Speekenbrink M, Konstantinidis E (2015) Uncertainty and exploration in a restless bandit problem. *Topics in cognitive science* 7(2):351–367.

Staats BR, KC DS, Gino F (2015) Blinded by experience: Prior experience, negative news and belief updating. *Harvard Business School NOM Unit Working Paper* (16-015).

Staats BR, Kc DS, Gino F (2018) Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Science* 64(2):804–824.

Su Y, Wang L, Santacatterina M, Joachims T (2019) CAB: Continuous adaptive blending for policy evaluation and learning. Chaudhuri K, Salakhutdinov R, eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 6005–6014 (PMLR), URL `https://proceedings.mlr.press/v97/su19a.html`.

Teodoridis F, Bikard M, Vakili K (2019) Creativity at the knowledge frontier: The impact of specialization in fast-and slow-paced domains. *Administrative Science Quarterly* 64(4):894–927.

Terwiesch C (2008) Product development as a problem-solving process. *Handbook of Technology and Innovation Management* 143.

Tirole J (1988) *The theory of industrial organization* (MIT press).

Van Parys BP, Golrezaei N (2020) Optimal learning for structured bandits. *arXiv preprint arXiv:2007.07302* .

Vanpaemel W, Lee MD (2012) Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review* 19(6):1047–1056.

WSJ (2016a) Barbie gains help soften sales decline at mattel. https://www.wsj.com/articles/mattels-loss-widens-on-sales-slump-though-barbie-shows-some-life-1469045759, accessed: 2011-11-01.

WSJ (2016b) Mattel to add curvy, petite, tall barbies. https://www.wsj.com/articles/mattel-to-add-curvy-petite-tall-barbies-1453991134, accessed: 2011-11-01.

WSJ (2018a) Coca-cola launched 500 drinks last year; most taste nothing like coke. *The Wall Street Journal* URL https://www.wsj.com/articles/coca-cola-launched-500-drinks-last-year-most-taste-nothing-like-coke-1535025601.

WSJ (2018b) Low-sugar innovations sweeten coke's growth. *The Wall Street Journal* URL https://www.wsj.com/articles/low-sugar-innovations-sweeten-cokes-growth-1532532426.

Appendices

A. Pseudo-algorithms

Algorithm 1 Static Model/Softmax Non-Bayesian

Tune γ, Infer μ_0
for $t \in \{1, \ldots, T\}$ **do**
  for $i \in \{1, \ldots, I_t\}$ **do**
    Compute $\hat{R}_{t,i} = \mu_t^T \cdot X_i$
    $\tilde{Y}_{t,i} \sim \tilde{p}_{t,i} = Softmax_\gamma(\hat{R}_{t,i})$ cf. Equation (2a)
  end for
end for

Algorithm 2 Greedy Model/ Softmax Bayesian

Set $\gamma = 1$, Infer μ_0
for $t \in \{1, \ldots, T\}$ **do**
  for $i \in \{1, \ldots, I_t)\}$ **do**
    Compute $\hat{R}_{t,i} = \mu_t^T \cdot X_i$
    $\tilde{Y}_{t,i} \sim \tilde{p}_{t,i} = Softmax_\gamma(\hat{R}_{t,i})$ cf. Equation (2a)
  end for
  Observe $(y_t, r_{y_t,t})$, Update((μ_t, Σ_t) cf. Equation (2b)
end for

Algorithm 3 Softmax UCB (Auer 2002)

Set $\gamma = 1$, Infer $\mu_0, \Sigma_0^\theta, \sigma_\epsilon, \beta_{UCB}$
for $t \in \{1, \ldots, T\}$ **do**
  for $i \in \{1, \ldots, I_t\}$ **do**
    Compute $\hat{R}_{t,i} = \mu_t^T \cdot X_i + \beta_{UCB}(X_i^T \cdot \Sigma_t \cdot X_i) + \sigma_\epsilon^2$
    $\tilde{Y}_{t,i} \sim \tilde{p}_{t,i} = Softmax_\gamma(\hat{R}_{t,i}))$ cf. Equation (2a)
  end for
  Observe $(y_t, r_{y_t,t})$, Update((μ_t, Σ_t) cf. Equation (2b)
end for

Algorithm 4 Softmax Gittins Index, GI (Chick and Gans 2009)

Set $\gamma = 1, \gamma_{GI} = 0.99$, Infer $\mu_0, \Sigma_0^\theta, \sigma_\epsilon$
for $t \in \{1, \dots, T\}$ **do**
  for $i \in \{1, \dots, I_t)\}$ **do**
    Compute $\hat{R}_{t,i} = \mu_t^T \cdot X_i + \sigma_\epsilon \cdot \tilde{b}\left(\frac{-X_i^T \cdot \Sigma_t \cdot X_i}{\sigma_\epsilon^2 * log\gamma_{GI}}\right)\sqrt{-log\gamma_{GI}}$
    $\tilde{Y}_{t,i} \sim \tilde{p}_{t,i} = Softmax_\gamma(\hat{R}_{t,i})$ cf. Equation (2a)
  end for
  Observe $(y_t, r_{y_t,t})$, Update((μ_t, Σ_t) cf. Equation (2b)
end for

Algorithm 5 Thompson Sampling

Infer $\mu_0, \Sigma_0^\theta, \sigma_\epsilon$
for $t \in \{1, \dots, T\}$ **do**
  for $i \in \{1, \dots, I_t)\}$ **do**
    for $k \in \{1, \dots, n\}$ **do**
      $\tilde{\theta}_t \sim \mathcal{N}(\mu_t, \Sigma_t)$
      Compute $\bar{R}_{t,i,k} = \tilde{\theta}_t^T \cdot X_i + \sigma_\epsilon^2$
      $\tilde{y}_t = \text{argmax}_{i \in \{1,\dots,I_t\}} \tilde{R}_{t,i,k}$
      $\tilde{Y}_{t,i,k} = \mathbf{1}_{\tilde{y}_t = i}$
    end for
    $\tilde{Y}_{t,i} \sim \tilde{p}_{t,i} = \frac{1}{n}\sum_{k=1}^n \tilde{Y}_{t,i,k}$ cf. Equation (2a)
  end for
  Observe $(y_t, r_{y_t,t})$, Update((μ_t, Σ_t) cf. Equation (2b)
end for

B. Skewness of Category Market Share

We present summary statistics for the variable $log_MarketShare$ based on periods of 1000 launches and compare these statistics to the descriptive statistics for $log\_MarketShare$ across the entire dataset. Each row in the table represents a specific period, where "ms" represents market share and e.g., "ms2000" refers to the market share within the period of 1,001 to 2,000 launches. Additionally, we include the kurtosis2 value, which is calculated as follows: $kurtosis2 = 3 \times (mean - median)/stdev$. A kurtosis value between 0.5 and 1 indicates slightly positive skewness, while a value greater than 1 suggests extreme skewness. From the provided statistics, we observe that the $log\_MarketShare$ variable exhibits a slightly positively skewed distribution.

Table 9: Kurtosis of Market Share variable

	mean	stdev	min	pc25	median	pc75	max	kurtosis2
Overall Market Share	0.0120	0.0302	0.0000	0.0009	0.0028	0.0097	1.0000	0.9108
ms1000	0.0843	0.1910	0.0000	0.0043	0.0168	0.0595	1.0000	1.0613
ms2000	0.0426	0.0507	0.0000	0.0059	0.0200	0.0672	0.2980	1.3382
ms3000	0.0365	0.0410	0.0000	0.0037	0.0140	0.0618	0.1490	1.6462
ms4000	0.0257	0.0354	0.0000	0.0033	0.0074	0.0419	0.1471	1.5504
ms5000	0.0178	0.0283	0.0000	0.0020	0.0061	0.0145	0.1438	1.2464
ms6000	0.0144	0.0232	0.0000	0.0015	0.0048	0.0141	0.1208	1.2414
ms7000	0.0115	0.0195	0.0000	0.0010	0.0029	0.0109	0.1076	1.3198
ms8000	0.0095	0.0169	0.0000	0.0009	0.0026	0.0093	0.1030	1.2359
ms9000	0.0085	0.0153	0.0000	0.0009	0.0024	0.0085	0.1013	1.2086
ms10000	0.0078	0.0142	0.0000	0.0008	0.0022	0.0072	0.0973	1.1975
ms11000	0.0074	0.0134	0.0000	0.0007	0.0021	0.0071	0.0928	1.1888
ms12000	0.0068	0.0127	0.0000	0.0006	0.0019	0.0063	0.0884	1.1688
ms13000	0.0067	0.0123	0.0000	0.0006	0.0020	0.0059	0.0836	1.1378

C. Data Distribution Visualization and Correlation Matrix

Figure 13 provides a visualization of the variables' distribution in the diagonal, their correlation matrix in the upper diagonal, boxplots, and the histogram for the dummy variables in the last two columns, respectively. The x-axis on the upper diagonal provides the relative frequency for every density plot on the diagonal. The distribution for $log\_MarketShare$ has a long tail on its right, highlighting that a large number of categories have a low market share value. Specifically, about $600,000$ observables have less than 10% market share.

The variable log_Price is skewed to the right and has several modes. Thus, many categories seem to belong to specific bins associating a category with a price level. The distribution for log_Volume has a fatter tail on the right, meaning that many categories have a very low level of sales. Regarding variables' correlation, unsurprisingly, log_Price and log_Volume are highly correlated with *Revenue*. The relationship between all variables is positive, but for the relationship between $log\_MarketShare$ and log_Volume. Hence, there is a positive association between past experience and past price as well as past revenue. However, there is a tiny negative linear relationship between past experience and past sale.

In this work, the relationship investigated is between past experience and future performance for existing categories, and to what extent past experience can be transferred to new category variants. The barplot for the dummy $Novel$ shows $12,861$ observations for the value $Novel$. However, only 150 of them are selected, as observed on the right-hand side of the cell ($Novel$, *Selected*), representing the number of categories in our dataset. Finally, boxplots show that selected categories either have a very high or very low market share, and there is no difference in mean past performance between selected and not selected categories.

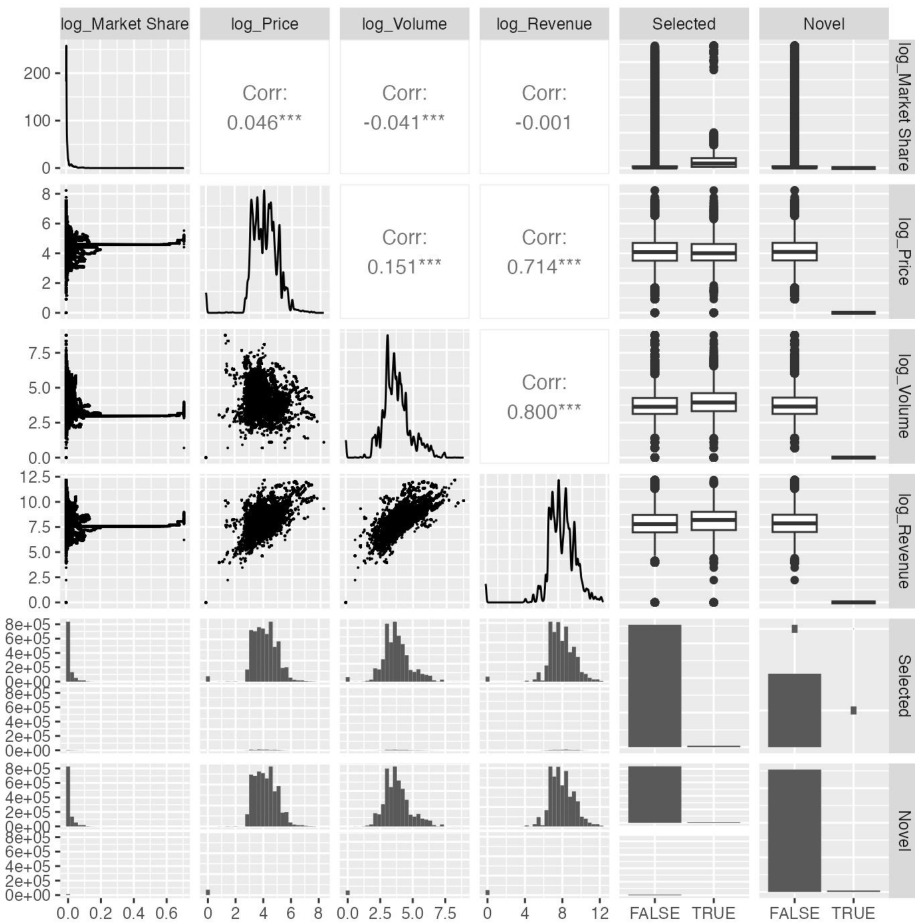


Figure 9: Visualization of data distribution and correlation matrix

D. Robustness Test Results

D.1 Longer training Samples, 1000 periods

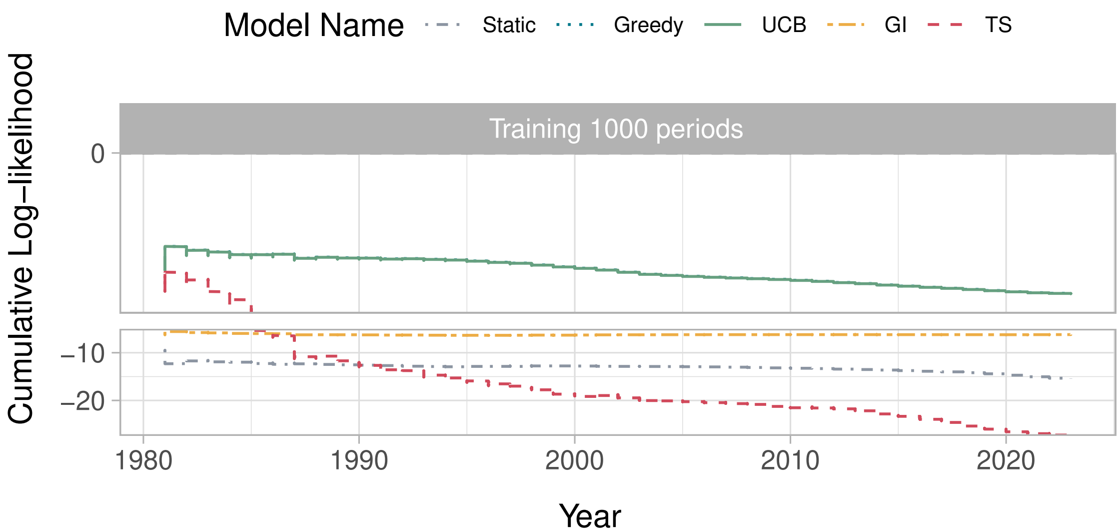
In-Sample Results. We redid the estimation of our main model displayed in Table 5 over 1000 samples. The results of this estimation are presented in Table 10. Similar to the findings from the inference with 350 samples, the UCB model exhibits a better fit on the training dataset compared to the other models. It is followed by the Greedy model. In all models, the parameters for $log\_MarketShare$ are consistently strong and positive. However, the confidence interval for UCB is quite large, indicating that the prior attributed to experience is uncertain. Interestingly, none of the models show a premium associated with category novelty, which is surprising, especially for an active learning policy like UCB. Additionally, the exploration bonus is very low, with a value of $\beta_{UCB} = 0.02$. This suggests that experimentation occurs over short horizons, and capturing this behavior becomes challenging over longer timeframes. This finding is supported by the close fit values of the UCB and Greedy models.

Table 10: Inference results. Training sample includes product launches in periods 1-1000.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	2.34	2.33	5	4.84	5	4.91
	0.18	[2.33, 2.33]	[-10.31, 19.62]	[-29.29, 37.77]	[3.40, 6.61]	[2.66, 7.31]
log_Volume	-0.04	-0.04	-2.23	-3.8	-0.74	2.35
	0.06	[-0.04, -0.04]	[-5.36, 0.96]	[-7.60, 0.21]	[-2.87, 1.54]	[-32.03, 37.84]
log_Price	-0.08	-0.08	-0.76	-1.17	-3.92	-0.11
	0.05	[-0.08, -0.08]	[-1.91, 0.29]	[-1.74, -0.55]	[-31.09, 21.65]	[-1.04, 0.90]
Novel	-2.01	-2.01	-4.37	-4.86	-4.08	-3.9
	0.36	[-2.01, -2.01]	[-17.15, 7.90]	[-5.85, -3.86]	[-6.66, -1.50]	[-7.93, 0.19]
Noise (σ_ϵ)		0	18.86	14.46	6.74	17.36
Bonus (β_{UCB})		0	0	0.02	0	0
Logtest	418.14					
McFadden pseudo-R2	0.03					
Wald Test	283.65					
Log-likelihood	-1990.57	-1990.57	-1779.89	-1770.18	-1955.66	-1928.01
AIC		3989.13	3567.78	3548.36	3919.32	3864.02
Observations	11844	11844	11844	11844	11844	11844
Simulations per period						1,000

Out-of-Sample Results. Figure 10 displays the predictions on the evaluation set spanning from period 1001 to 12,861. It is evident that the UCB and Greedy models exhibit complete overlap in their predictions. In contrast, the Gittins Index model demonstrates a significant performance advantage over the Static model, indicating that it requires a larger sample size to effectively learn from the available data. Additionally, the predictions from the Thompson Sampling model consistently underperform compared to both the Greedy and UCB models, and its performance continues to deteriorate to the extent that it performs worse than the Gittins Index and Static models.

Figure 10: Out of Sample Performance with Revenue as product performance variable. 1000 training samples.



D.2 Other Performance Variable: Price and Volume

Inference with Current Volume. We conducted model estimation using category volume (unit sold) as the performance variable. The results of this estimation are presented in Table 11. The parameter estimates and goodness-of-fit values obtained in this analysis are comparable to those reported in Table 5.

Figure 11 further illustrates the superiority of the Greedy model over the UCB model on the evaluation set. However, in this specific specification, the Static model exhibits a performance that is closely comparable to the UCB model.

Inference with Current Price. We also attempt to confirm the validity of our findings by using log_Price as the performance variable. In this case, the Greedy model demonstrates a superior fit compared to the other models. Furthermore, the parameters associated with $log\_MarketShare$ are consistently positive, and the confidence intervals for these parameter estimates are significant. These results indicate that when aiming to introduce higher-priced products, there is limited scope for experimentation, and the Greedy algorithm is likely to be the best fit for the observed data.

Compared to the in-sample results with price as performance measure, the out-of-sample analysis yields even more interesting findings. The Static model produces predictions that are as accurate as those of the Greedy model and even outperforms the UCB model, as shown in Table 13 with comparable WAIC, pWAIC, and Akaike Weight values. This result is further supported by Figure 12. It suggests that when the performance metric revolves around increasing the category price, a Greedy model is just as effective as a model that follows previous launch patterns by focusing on categories with higher prices. However, our in-sample analysis, as shown in Table **??**, indicates that the fit of the UCB model is superior to the fit of this specification with log_Price as the performance variable. Nonetheless, the Greedy model demonstrates superior performance across all alternative specifications and does not exhibit overfitting.

Figure 11: Out of Sample Performance with Volume as product performance variable.

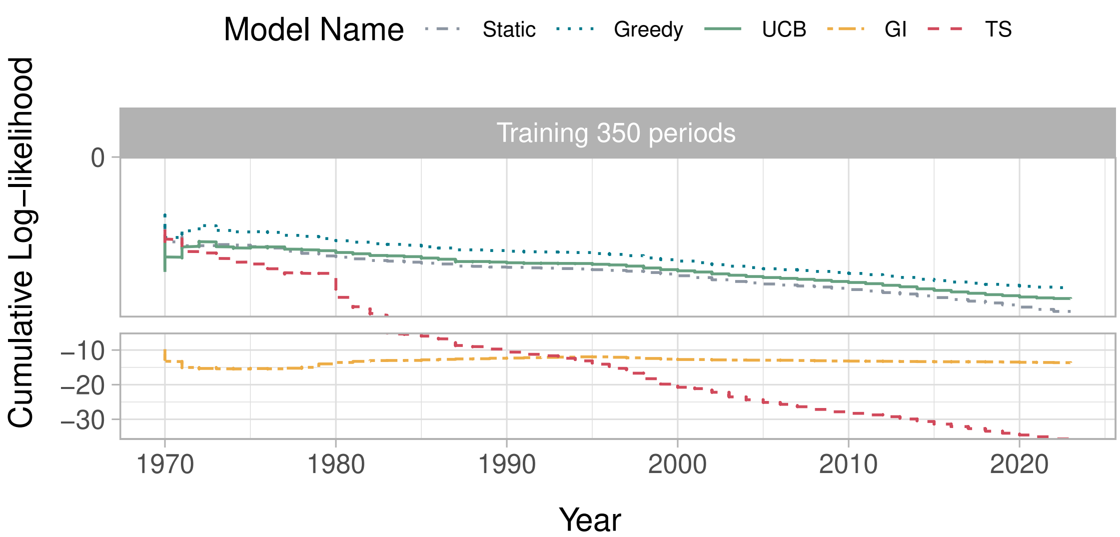


Figure 12: Out of Sample Performance with Price as product performance variable.

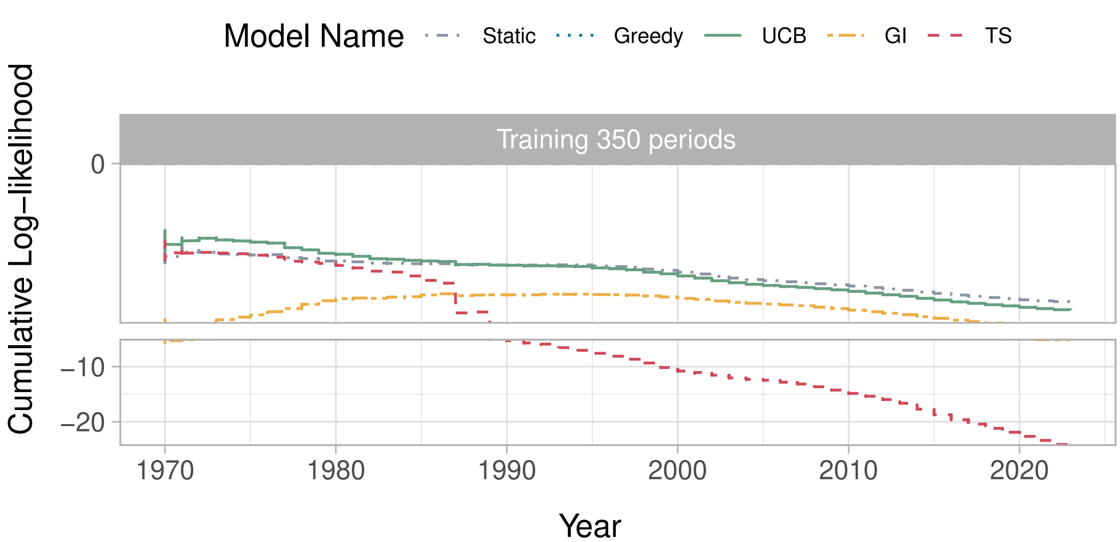


Table 11: Inference results with Volume as performance variable. Training sample decision period 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	3.17	3.17	4.82	0.31	4.5	4.55
	(0.29)	[3.17, 3.17]	[4.08, 5.53]	[-32.05, 31.54]	[3.07, 5.95]	[1.86, 7.43]
log_Volume	0.29	0.29	0.01	-0.54	0.25	0.54
	(0.15)	[0.29, 0.29]	[-1.72, 1.78]	[-2.25, 1.27]	[-0.38, 0.92]	[0.48, 0.61]
log_Price	0.7	0.7	2.17	0.47	0.07	3.66
	(0.17)	[0.70, 0.70]	[-34.24, 35.44]	[-5.99, 7.37]	[-15.69, 14.90]	[-3.48, 11.35]
Novel	2.74	2.74	-4.15	2.34	-2.77	-3.96
	(1.15)	[2.74, 2.74]	[-37.60, 27.95]	[-16.86, 21.65]	[-10.31, 4.77]	[-14.01, 6.27]
Noise (σ_ϵ)			18.5	15.55	18.59	18.09
Bonus (β_{UCB})				0.19		
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-182.4	-174.97	-186.81	-187.68
Observations	1290	1290	1290	1290	1290	1290
Simulations per period						1,000

D.3 Breakpoints

We plot the correlation between the yearly average predictions for product categories and LEGO's average choices in Figure 13. We add horizontal lines for correlation values at 0.5 and 0.25. This allows us to see that the Greedy model has a higher correlation than the competitive models for several years until 1978, where it falls below the 0.5 line. We use this year to undertake the breakpoint analysis. Even after 1978, the correlation between the Greedy model and LEGO's choices remains above the value of 0.25, unlike the remaining competing policies. Interestingly, 1977 and 1979 are the years that Kjeld Kirk Kristiansen, the grandson of LEGO's founder, joined the company and took over as CEO respectively.

D.4 Parameter recovery

To ensure that data generated with a specific model can also be reverse-engineered to retrieve the same model, we attempt to recover parameters of simulated model by undertaking the following optimization exercise:

$$\max_{\substack{\mu_1\\ \Sigma_1\\ \sigma_\epsilon^2}} \sum_{t=1}^{T} \ln\left(p(\tilde{y}_t^{M'}|M, X_{\cdot t}, \mu_t, \Sigma_t, \sigma_\epsilon^2)\right) \tag{8}$$

where $\tilde{y}_t^{M'}$ is simulated with model M', but is conditioned on another model M from which the inference is undertaken. We attempt to confirm that the maximum likelihood is obtained when $M' = M$. Table 14 presents the results of recovering parameters from simulated data obtained with the Static model. We find that the Greedy model is most likely to have generated this data,

Table 12: Inference results with Price as Performance variable. Training sample decision period 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	3.17	3.17	4.37	4.8	5	4.87
	(0.29)	[3.17, 3.17]	[3.46, 5.25]	[4.52, 5.07]	[4.90, 5.09]	[0.01, 10.06]
log_Volume	0.29	0.29	0.18	-0.19	-2.32	0.44
	(0.15)	[0.29, 0.29]	[-3.19, 3.62]	[-8.66, 8.77]	[-10.15, 6.05]	[-0.16, 1.06]
log_Price	0.7	0.7	0.52	-0.15	-4.17	4.41
	(0.17)	[0.70, 0.70]	[-17.81, 17.26]	[-19.80, 20.84]	[-27.43, 17.72]	[-19.18, 29.83]
Novel	2.74	2.74	-1.73	-4.39	-3.76	-3.04
	(1.15)	[2.74, 2.74]	[-28.13, 23.62]	[-38.24, 29.66]	[-6.05, -1.46]	[-11.31, 5.37]
Noise (σ_ϵ)			18.88	18.77	2.26	18.18
Bonus (β_{UCB})				0.05		
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-182.44	-183.22	-186.09	-189.44
Observations	1290	1290	1290	1290	1290	1290
Simulations per period						1,000

Table 13: Watanabe Information Criteria (WAIC) with Price as Performance.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Non-Bayesian Softmax	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
WAIC	154,764	146,397	204,923	1,129,073	834,463
Mean period WAIC	12.37	11.70	16.38	90.25	66.70
dWAIC	0.67	0.00	4.68	78.55	55.00
Standard Error	0.03	0.04	0.03	0.02	0.04
Akaike Weight	39.5%	55.2%	5.3%	0%	0%
pWAIC	51.59	45.62	120.64	85.10	382.24

followed by the Thompson Sampling and Gittins Index models, respectively. The Static model performs worse compared to these models.

Firstly, the Greedy model exhibits a remarkable capability to accurately represent data generated by other policies. This finding can be attributed to the existence of regime changes in behavioral models across different time horizons, where the Greedy models consistently capture a behavior that persists over long periods. Consequently, relying solely on inference results is insufficient to draw conclusions about whether data generated by an agent originates from this specific policy. However, it is evident from Figure 13 that the Greedy model's predictions exhibit the highest correlation with LEGO's choices over a significantly long period. This serves as further evidence that our results are valid and robust.

Secondly, the high likelihood values for Thompson Sampling and the Gittins Index demonstrate that these policies are more adept at matching data generated by another bandit policy rather than data generated by a non-bandit policy.

Figure 13: Correlation between LEGO average choices and predictions from Model 0 to 5.

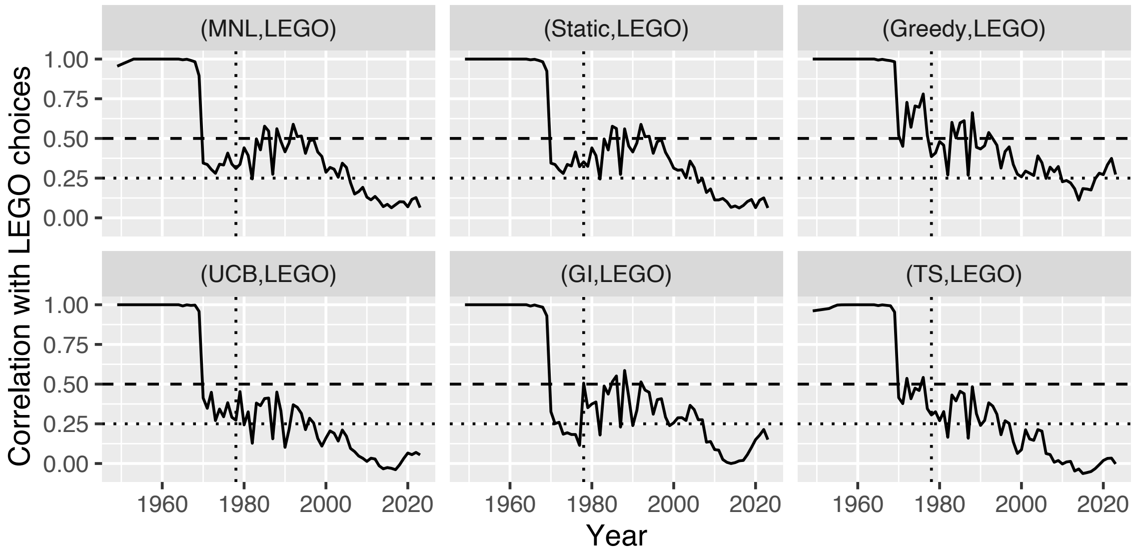


Table 15 and Table 16 display the outcomes of parameter recovery using launch patterns simulated from the Greedy and UCB models, respectively. These results corroborate the findings from Table 14, indicating that the Greedy model consistently outperforms the other models. However, it is noteworthy that the fit values and parameter estimates remain identical for all models incorporating a Softmax layer (Model 1 to Model 4).

Table 14: Parameter recovery from Static Model. Training sample decision period 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	-5	-3.29	-4.9	-4.06	-4.98
	[-5.00, -5.00]	[-37.40, 29.28]	[-7.69, -2.21]	[-26.06, 18.19]	[-5.76, -4.14]
log_Volume	0.62	4.36	4.93	2.64	0.2
	[0.62, 0.62]	[3.15, 5.60]	[-1.38, 11.60]	[0.56, 4.88]	[-0.31, 0.73]
log_Price	2.55	4.91	2.34	4.42	-2.96
	[2.55, 2.55]	[4.35, 5.43]	[2.07, 2.62]	[3.16, 5.61]	[-23.28, 18.93]
Novel	-5	-3.74	-4.53	-4.16	-1.48
	[-5.00, -5.00]	[-20.60, 12.45]	[-35.20, 26.31]	[-5.67, -2.64]	[-4.87, 1.97]
Noise (σ_ϵ)	0	18.97	18.93	18.99	18.3
Bonus (β_{UCB})	0	0	0.01	0	0
Log-likelihood	-61.02	-11.55	-60.46	-39.57	-20.46
Observations	1290	1290	1290	1290	1290

D.5 Other robustness checks

Alternative Ground Truth. We analyze the predictions on the evaluation set data by simulating policies with a prior other than the UCB model, which had the maximum likelihood in our main

Table 15: Parameter recovery from Greedy Model. Training sample decision period 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	-5	-3.29	-4.9	-4.06	-4.85
	[-5.00, -5.00]	[-37.40, 29.28]	[-7.69, -2.21]	[-26.06, 18.19]	[-5.53, -4.13]
log_Volume	0.62	4.36	4.93	2.64	0.47
	[0.62, 0.62]	[3.15, 5.60]	[-1.38, 11.60]	[0.56, 4.88]	[0.02, 0.93]
log_Price	2.55	4.91	2.34	4.42	0.75
	[2.55, 2.55]	[4.35, 5.43]	[2.07, 2.62]	[3.16, 5.61]	[-29.72, 33.59]
Novel	-5	-3.74	-4.53	-4.16	-4.55
	[-5.00, -5.00]	[-20.60, 12.45]	[-35.20, 26.31]	[-5.67, -2.64]	[-7.09, -1.97]
Noise (σ_ϵ)	0	18.97	18.93	18.99	18.55
Bonus (β_{UCB})			0.01		
Log-likelihood	-61.02	-11.55	-60.46	-39.57	-20.97
Observations	1290	1290	1290	1290	1290

Table 16: Parameter recovery from UCB Model. Training sample decision period 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	-5	-3.29	-4.9	-4.06	-4.99
	[-5.00, -5.00]	[-37.40, 29.28]	[-7.69, -2.21]	[-26.06, 18.19]	[-6.03, -3.87]
log_Volume	0.62	4.36	4.93	2.64	4.21
	[0.62, 0.62]	[3.15, 5.60]	[-1.38, 11.60]	[0.56, 4.88]	[-1.36, 9.97]
log_Price	2.55	4.91	2.34	4.42	1.98
	[2.55, 2.55]	[4.35, 5.43]	[2.07, 2.62]	[3.16, 5.61]	[0.96, 3.07]
Novel	-5	-3.74	-4.53	-4.16	-3.21
	[-5.00, -5.00]	[-20.60, 12.45]	[-35.20, 26.31]	[-5.67, -2.64]	[-6.57, 0.22]
Noise (σ_ϵ)	0	18.97	18.93	18.99	18.72
Bonus (β_{UCB})			0.01		
Log-likelihood	-61.02	-11.55	-60.46	-39.57	-34.51
Observations	1290	1290	1290	1290	1290

estimation. Specifically, we use the Greedy and Static models sequentially. Figure 14 illustrates that when using the prior from the Greedy model, the predictions from both the Greedy and UCB models completely overlap. This finding indicates that both the UCB and Greedy models outperform other competing models when it comes to predicting unseen data.

Additionally, Figure 15 illustrates that when considering the Static model as the ground truth, it is evident that the Static model outperforms the predictions of the competing models. On the one hand, the predictions of the Greedy, UCB, and Gittins Index models show complete overlap. This suggests that the prior information from the Static model does not provide strong guidance for these models to learn unique characteristics from the data. On the other hand, the predictions of the Thompson Sampling model deteriorate from the beginning and remain consistently poor. This highlights the sensitivity of these policies to the underlying ground truth, particularly in the case of Thompson Sampling.

Inference with trends covariates. Table 17 displays the results of the estimation with a curvilinear effect of years. It is observed that only the Greedy model demonstrates improvement,

Figure 14: Out of Sample Performance with the Greedy model as the ground truth

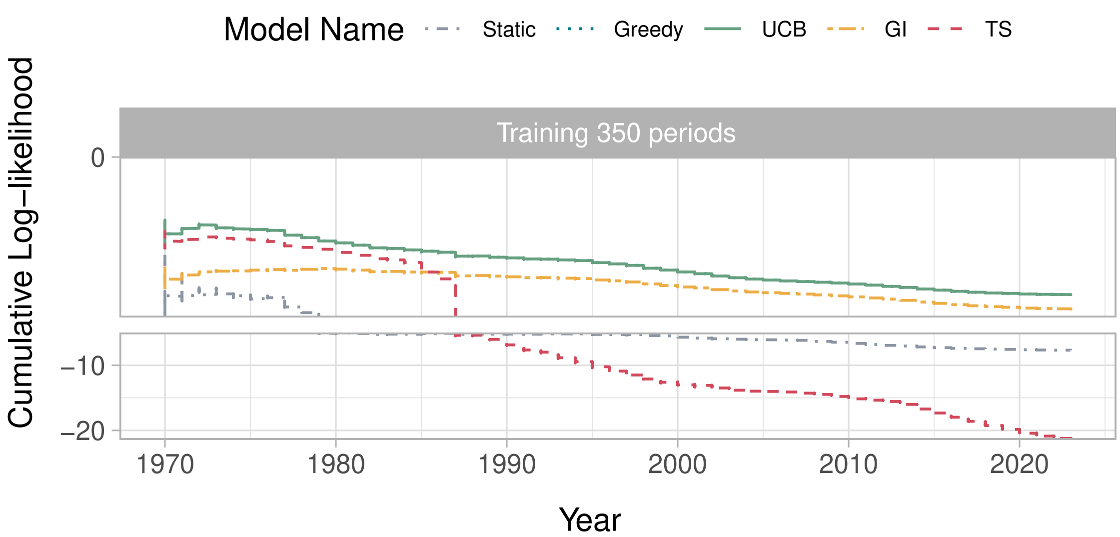
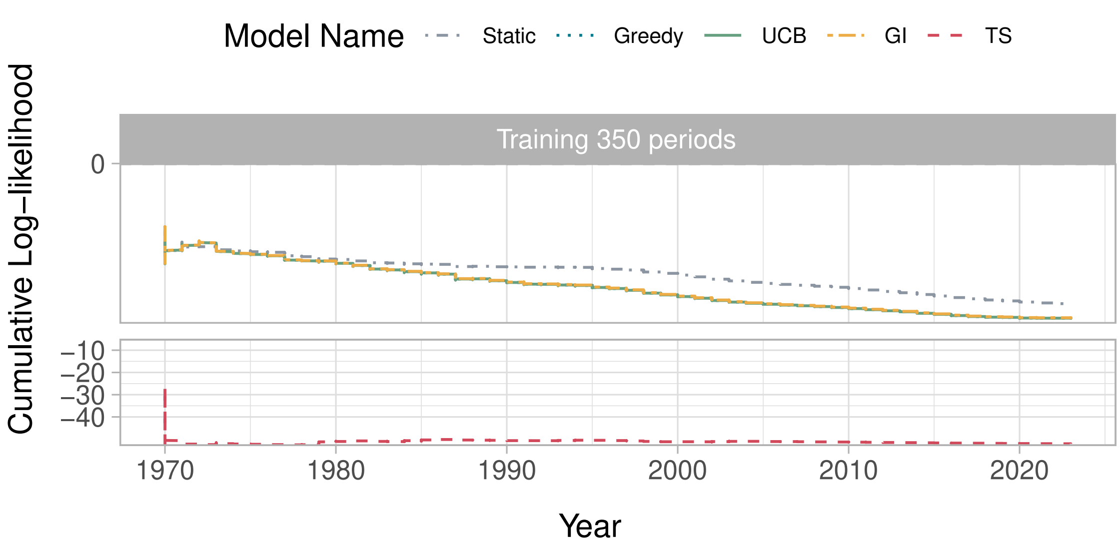


Figure 15: Out of Sample Performance with the Static model as the ground truth



indicating a higher likelihood of launching an increasing number of products in categories as the years progress. Conversely, the fit for the remaining policies deteriorates.

Table 17: Curvilinear trends, revenue as performance. Training sample decision periods 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	3.17	3.18	4.99	-2.94	4.94	4.87
	(0.29)	[3.18, 3.18]	[3.18, 6.72]	[-27.70, 21.67]	[4.29, 5.61]	[0.53, 9.34]
log_Volume	0.29	0.30	-0.04	-2.03	0.27	0.32
	(0.15)	[0.30, 0.30]	[-2.57, 2.46]	[-4.08, 0.03]	[-0.27, 0.78]	[-0.04, 0.70]
log_Price	0.7	0.7	0.53	1.4	4.07	4.8
	(0.17)	[0.70, 0.70]	[-18.36, 19.94]	[-1.56, 4.33]	[-7.81, 15.86]	[-29.16, 34.97]
Novel	2.74	2.71	-1.25	2.67	-1.73	-2.88
	(1.15)	[2.71, 2.71]	[-14.92, 13.25]	[-4.61, 9.94]	[-7.00, 3.81]	[-9.14, 3.22]
log_Year		0.55	2.63	2.16	-3.81	-0.52
	0	[0.55, 0.55]	[-7.62, 12.96]	[-3.03, 7.54]	[-22.61, 15.60]	[-9.27, 8.11]
$(log\_Year)^2$		-4.17	4.95	-2.41	0.94	-2.58
	0	[-4.17, -4.17]	[-1.77, 11.42]	[-4.89, 0.04]	[-14.79, 17.23]	[-20.27, 16.42]
Noise (σ_ϵ)			18.83	4.63	14.42	17.34
Bonus (β_{UCB})				0.51		
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-181.66	-178.32	-187.37	-188.95
Observations	1290	1290	1290	1290	1290	1290

Inference with forgetting rate. Table 18 demonstrates that incorporating a forgetting rate does not enhance predictions. Additionally, the UCB and Gittins Index models fail to converge. These models already inflate parameter variance through their exploration bonus, and introducing a forgetting rate for older observations further amplifies the variance, leading to an explosion in variance.

Inference with only category fixed-effect. Table 19 provides insights into the explanatory power of only category fixed-effects. It reveals that Thompson Sampling exhibits a better fit than the other models. The Greedy model's fit improves compared to our main model, indicating that category heterogeneity alone can account for category launches. However, when shared covariates are introduced, the fit of the Greedy model deteriorates, likely due to the path dependence of these covariates. Additionally, the Greedy, UCB, and Gittins Index models demonstrate comparable model fit. This suggests that without shared features, category identity alone does not provide sufficient information to construct a category launch strategy. Moreover, the Gittins Index exhibits a very low noise value, second only to Thompson Sampling. This suggests that in the absence of problem structure, such as that introduced by shared covariates, these policies offer better explanations for the launch pattern, though their sampling behavior cannot be differentiated as indicated earlier.

Inference with only category fixed-effect and novelty dummy. Table 20 presents the results of including category novelty in addition to category fixed effects. It can be observed that classic bandit policies such as Thompson Sampling, Gittins Index, and UCB exhibit better explanatory power compared to the other models. These policies provide a stronger fit to the data

Table 18: Inference results with a forgetting rate factor. Training set: periods 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax	Softmax	Softmax	Softmax	Thompson
	Non-Bayesian	Greedy	UCB	GI	Sampling
log_MarketShare	3.18	4.75	0.62	0.55	4.77
	[3.18, 3.18]	[4.59, 4.92]	[-26.59, 26.89]	[-9.83, 11.05]	[1.52, 8.24]
log_Volume	0.3	-0.02	-3.12	1.02	1.21
	[0.30, 0.30]	[-14.63, 14.88]	[-11.86, 6.11]	[-23.58, 27.31]	[0.37, 2.07]
log_Price	0.7	0.11	-0.62	-0.7	3.06
	[0.70, 0.70]	[-0.73, 0.87]	[-35.77, 36.93]	[-13.66, 11.49]	[-12.08, 19.37]
Novel	2.75	-1.6	3.12	-0.39	-1.72
	[2.75, 2.75]	[-6.64, 3.23]	[-18.06, 24.44]	[-29.36, 28.65]	[-4.54, 1.16]
Noise (σ_ϵ)	0	18.94	14.25	11.01	18.71
Bonus (β_{UCB})			0.13		
Log-likelihood	-186.43	-187.74	-578.22	-9027.88	-214.53
Observations	1290	1290	1290	1290	1290

Table 19: Inference results with only Category fixed effect. Training sample decision period 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax	Softmax	Softmax	Softmax	Thompson
	Non-Bayesian	Greedy	UCB	GI	Sampling
Noise (σ_ϵ)		18.08	17.50	0.93	8.93
Bonus (β_{UCB})			0		
Log-likelihood	-415.67	-181.22	-181.22	-181.28	-178.50
Observations	1290	1290	1290	1290	1290
Simulations per period					1,000

when considering both category fixed effects and category novelty.

All policies have slightly better performance compared to the model without Novelty Dummy.

Figure 16: Out of Sample Performance with Category fixed effect only as covariate.

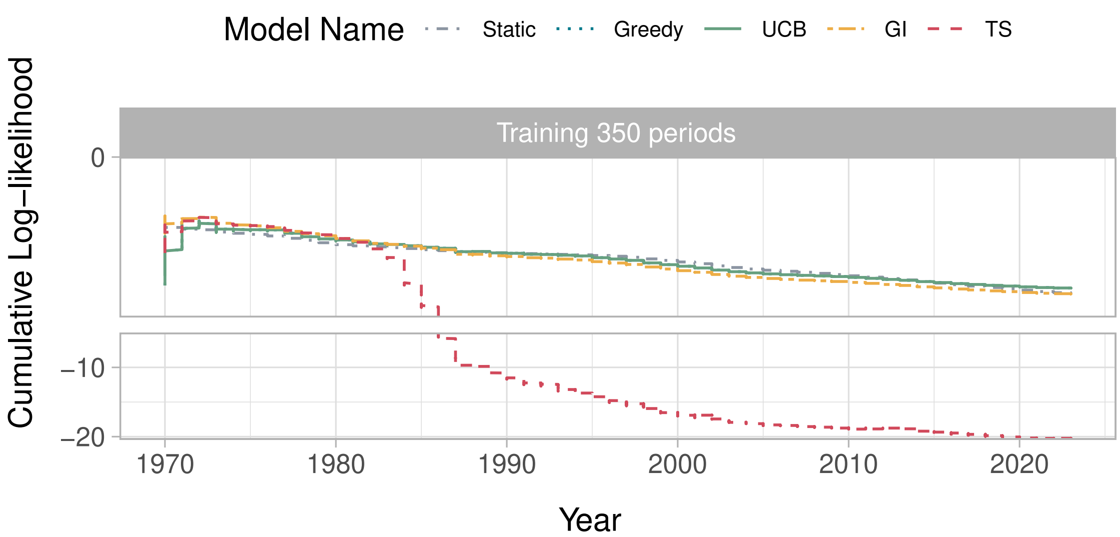


Figure 17: Out of Sample Performance with Category fixed effect and Novelty only as covariate.

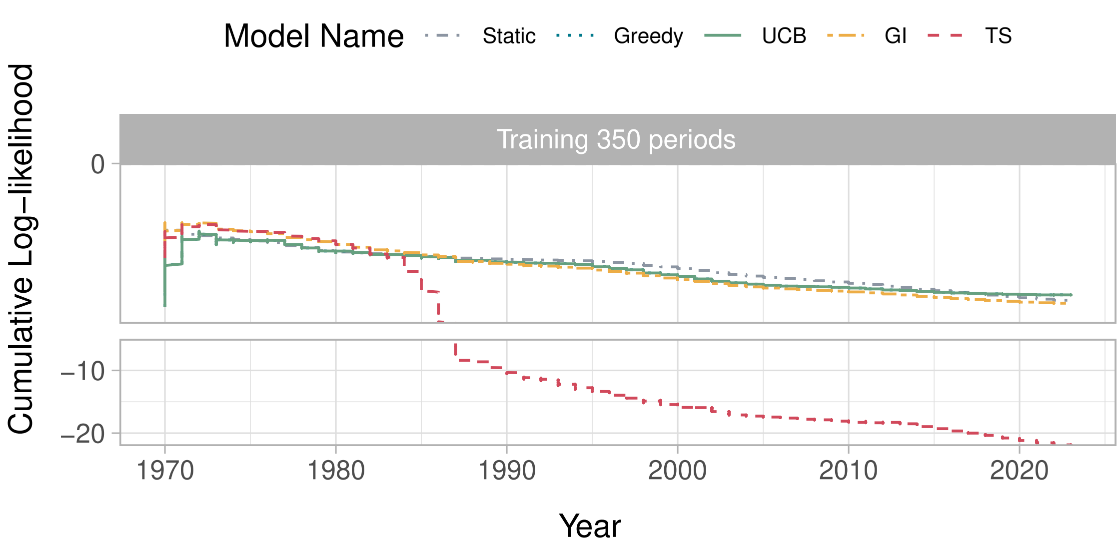


Table 20: Inference results with only Category fixed effect and Novelty. Training sample decision period 1-350.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
Novel	-3.34	-2.37	-3.89	-4.86	-4.84
	[-3.34, -3.34]	[-41.84, 32.94]	[-32.27, 25.40]	[-16.14, 7.20]	[-35.58, 22.44]
Noise (σ_ϵ)		18.87	17.67	10.37	10.32
Bonus (β_{UCB})			0.31		
Log-likelihood	-294.60	-179.11	- 178.87	-178.76	-176.64
Observations	1290	1290	1290	1290	1290
Simulations per period					1,000

E. Additional Analysis

E.1 Inference with random launch within year

Table 21 showcases the results of an analysis where category launches are shuffled within the same year due to the lack of precise launch dates. The fit values have improved compared to our main specifications. However, our results and conclusions remain consistent. The UCB and Greedy models exhibit superior in-sample performance. For the UCB model, the prior information based on past experience and novelty is diffuse. Similarly, for the Greedy model, the prior value for novelty is also diffuse.

Table 21: Inference results with Product launched schuffled within year. Training set: periods 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax Non-Bayesian	Softmax Greedy	Softmax UCB	Softmax GI	Thompson Sampling
log_MarketShare	2.79	2.79	3.84	1.96	4.99	4.95
	0.3	[2.79, 2.79]	[3.14, 4.50]	[-12.93, 16.33]	[4.53, 5.45]	[4.27, 5.67]
log_Volume	0.4	0.4	-2.61	-3.09	-1.55	-0.88
	0.15	[0.40, 0.40]	[-22.20, 17.37]	[-4.70, -1.38]	[-6.41, 3.66]	[-5.80, 4.21]
log_Price	0.7	0.7	4	-2.1	2.12	1.11
	0.2	[0.70, 0.70]	[-30.53, 35.54]	[-34.63, 32.65]	[-0.61, 4.69]	[0.01, 2.30]
Novel	2.62	2.64	0.77	1.76	1.88	-4.81
	1.14	[2.64, 2.64]	[-9.27, 10.40]	[-25.71, 29.38]	[0.72, 3.04]	[-13.95, 4.48]
Noise (σ_ϵ)		0	18.49	3.05	3	16.29
Bonus (β_{UCB})		0	0	0.44	0	0
Logtest	463.93					
McFadden pseudo-R2	0.3					
Wald Test	210.77					
Log-likelihood	-181.92	-181.92	-175.13	-168.1	-176.84	-181.58
Observations	1282	1290	1290	1290	1290	1290
Simulations per period						1,000

E.2 Inference with retail sales revenue

Table 22 presents the results of an estimation using revenue generated from actual product sales in the market. Our results and conclusions remain consistent and unchanged.

Table 22: Retail revenue as performance variable. Training sample decision periods 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax	Softmax	Softmax	Softmax	Thompson
		Non-Bayesian	Greedy	UCB	GI	Sampling
log_MarketShare	3.17	3.18	4.92	0.01	4.97	4.66
	(0.29)	[3.18, 3.18]	[2.05, 7.66]	[-32.59, 31.49]	[2.85, 7.11]	[-2.26, 12.05]
log_Volume	0.29	0.29	0.05	-0.76	0.17	-2.24
	(0.15)	[0.29, 0.29]	[-1.09, 1.21]	[-3.39, 2.02]	[-0.88, 1.28]	[-28.57, 24.94]
log_Price	0.7	0.7	4.54	1.32	-0.21	0.39
	(0.17)	[0.70, 0.70]	[-10.44, 18.22]	[-3.35, 6.32]	[-6.57, 5.78]	[0.10, 0.70]
Novel	2.74	2.73	-0.03	4.8	-3.19	-4.22
	(1.15)	[2.73, 2.73]	[-19.99, 19.13]	[-20.43, 30.19]	[-10.08, 3.71]	[-10.88, 2.56]
Noise (σ_ϵ)		0	18.91	12.82	18.98	18.92
Bonus (β_{UCB})		0	0	0.23	0	0
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-184.27	-179.54	-188.32	-185.52
Observations	1290	1290	1290	1290	1290	1290

E.3 Inference results with wholesale revenue

Table 22 displays the results of an estimation utilizing revenue derived from the inventory sitting at the reseller for the last 6 months. It also provides results supporting our main findings.

Table 23: Wholesale revenue as performance variable. Training sample decision periods 1-350.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	MNL	Softmax	Softmax	Softmax	Softmax	Thompson
		Non-Bayesian	Greedy	UCB	GI	Sampling
log_MarketShare	3.17	3.18	4.99	0.56	4.66	4.6
	(0.29)	[3.18, 3.18]	[4.57, 5.38]	[-34.35, 34.26]	[2.22, 7.12]	[0.94, 8.52]
log_Volume	0.29	0.29	0	-1.53	0.29	0.24
	(0.15)	[0.29, 0.29]	[-2.05, 2.09]	[-5.10, 2.26]	[0.13, 0.47]	[-0.30, 0.81]
log_Price	0.7	0.7	2.08	0.71	0.47	0.34
	(0.17)	[0.70, 0.70]	[-14.92, 17.62]	[-5.86, 7.72]	[-24.61, 24.07]	[-17.17, 19.21]
Novel	2.74	2.73	-0.93	2.3	-0.66	-4.31
	(1.15)	[2.73, 2.73]	[-23.25, 20.49]	[-12.76, 17.45]	[-6.81, 5.51]	[-12.35, 3.87]
Noise (σ_ϵ)		0	17.95	10.54	16.18	14.98
Bonus (β_{UCB})		0	0	0.19	0	0
Logtest	458.5					
McFadden pseudo-R2	0.3					
Wald Test	215.28					
Log-likelihood	-186.43	-186.43	-181.56	-176.61	-186.41	-186.71
Observations	1290	1290	1290	1290	1290	1290

E.4 Category Fixed Effect Inheritance

Table 24 and 25 demonstrate how new categories inherit their mean and variance priors from Category0. In period 290, the Universal Building Set (UBS) category is introduced, resulting in an update to the values of both the Novel variable and Category 0 priors. The mean value increases from 1.76 to 2.16, while the variance decreases from 6 to 5.66. Furthermore, the new UBS category inherits these updated priors. In period 291, the Legoland category is launched and similarly acquires the updated prior from Category0.

Table 24: Mean fixed effect inheritance from Category0 in period 290-292.

Period	Novel	Category0	System	Samsonite	Train	UBS	Legoland
289	4.6860	1.7693	0.5789	1.0437	1.8543		
290	4.7088	2.1679	0.5789	1.0437	1.8543	2.1679	
291	4.7018	2.0448	0.5789	1.0437	1.8543	2.1679	2.0448
292	4.7018	2.0448	0.8898	1.4578	2.9275	1.0384	2.0448

Table 25: Variance fixed effect inheritance from Category0 in period 290-292.

Period	Novel	Category0	System	Samsonite	Train	UBS	Legoland
289	4.4071	6.0073	6.4721	5.2885	4.5344		
290	4.4056	5.6626	6.4721	5.2885	4.5344	5.6626	
291	4.4047	5.4408	6.4721	5.2885	4.5344	5.6626	5.4408
292	4.4047	5.4408	6.4352	5.2079	3.8572	5.0763	5.4408