Experimental Economics (2025), **28**, 1240–1261
doi:10.1017/eec.2026.10050

ORIGINAL PAPER

Learning by consuming in the lab

Louis Lévy-Garboua[1,2], Claire Owen[3], Laetitia Placido[4] and Benoît Rapoport[3,5]

[1]Paris School of Economics, Université Paris 1 Panthéon-Sorbonne, Centre d'économie de la Sorbonne, Paris, France
[2]CIRANO, Montreal, Quebec, Canada
[3]Université Paris1-Panthéon-Sorbonne, Centre d'économie de la Sorbonne, Paris, France
[4]City University of New York, Baruch College, New York, NY, USA
[5]Institut National d'Etudes Démographiques (INED), Paris, France
Corresponding author: Laetitia Placido; Email: laetitia.placido@baruch.cuny.edu

The authors of this paper certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

(Received 7 February 2024; revised 11 November 2024; accepted 21 April 2026)

Abstract

We build an experiment to uncover the bandit-like nature of consumer behavior usually masked by the product and time aggregation of consumption data. Subjects make repeated choices between musical styles (either all familiar or unfamiliar), and post-choice satisfaction is observed. We estimate Bayesian bandit models of learning taste by consuming with satiation. Our best model features decreasing random exploration, with openness being associated with higher exploration. Early exploration is more intensive in the unfamiliar treatment and persists throughout the experiment in both treatments. Overall, subjects make choices that deviate from their best prediction 61.5% of the time in the unfamiliar treatment versus 44.7% in the familiar treatment. Our model offers a rational interpretation of random utility in discrete consumer choices which does not rest on perception and/or decision errors.

Keywords: Bayesian bandit models; experienced utility; experimental consumer choices; learning by consuming; random exploration; random utility

JEL Codes: C91; D11; D12; D83; Z11

1. Introduction

How do consumers come to know what they like? This question is increasingly relevant in economies characterized by high innovation and differentiation. Consumers usually face several varieties of one good like different brands or different musical styles. Within a period, they may consume sequentially more than one variety and experience heterogeneous products from each variety to learn what they like. The multiplicity of varieties of each good, of unknown products within each variety, and the need to experience these products sequentially generate choice uncertainty in a dynamic setting. However, the dynamic and stochastic nature of consumer choice is usually masked by statistics of consumption data through a process of product and time aggregation. Hence, while observational data show the quantities of all varieties consumed in a discrete period as if consumers had solved a static problem of choice among a set of homogeneous varieties, we conjecture that rational consumers have been facing a stochastic multi-armed bandit problem through repeated consumption of different varieties and products. This conjecture about real-world behavior will be tested in an experiment that

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

forces subjects to make repeated consumption choices among heterogeneous varieties of an experience good thus revealing the underlying sequential decision-making process hidden by observational data.[1] Restoring the hidden dimensions of consumer choices in a controlled environment will provide a rational interpretation of random utility in discrete consumer choices which does not have to rest on unobservable perception and/or decision errors in a static setting.

We focus on the dynamic and stochastic nature of consumer choices by examining how people 'learn by consuming' in a unique lab experiment. Given our focus on consumption and preference for goods, the experiment will be conducted in the natural context of music consumption. Our experiment offers 50 repeated choices between four musical styles (i.e., varieties) and measurement of post-choice satisfaction. In addition, it contains two treatments providing unequal opportunities for learning. To focus on the dynamics of preference discovery, goods can be acquired for free. Subjects repeatedly consume the musical style they select, which provides non-pecuniary incentives to choose what they like.

We chose music because of its external validity: Consuming music in the lab is no different from streaming music in isolation. Moreover, music appears to be especially fit for uncovering the individual strategies masked by statistical aggregation of consumption data in the limited time of an experiment. Indeed, it is sufficiently universal to concern all subjects, sufficiently pleasant to avoid boredom, sufficiently unfamiliar to allow learning, and sufficiently heterogeneous to avoid rapid satiation. Finally, music allows us to contrast behavior when faced with familiar or unfamiliar styles. Thanks to the nice properties of music, this experiment is perhaps the first to investigate repeated consumption choices in the lab, with direct measure of experienced utility. Having many repetitions enables us to uncover exploration strategies, including random exploration and diversification that would otherwise be hidden in aggregate data.

This repeated consumption task can be viewed as a four-armed bandit problem in which consumers who seek to maximize their experienced utility over an extended period choose one musical style at a time, unaware of which music piece will be played and how much utility they will derive from it. Each consumption reveals an experienced utility that provides additional information on the distribution of rewards between musical styles. This problem is non-trivial, however, because even though consumers evaluate precisely the experienced utility from the products they consume after each round, their choice yields no information about non-selected varieties. Therefore, they need to explore different varieties to acquire new information, while the exploitation of the optimal decision so far would yield the highest expected utility conditional on prior information. In essence, they seek to learn which varieties are best for themselves while not spending too much time exploring. Moreover, while adoption of an optimal policy allows the consumer to converge toward stable latent preferences, preference learning is complicated in the short run by the phenomenon of satiation in consumption goods. This mechanism generates a preference for diversity.

We have two treatments, distinguishing between familiar music and unfamiliar music, to check whether consumers make more stable and concentrated choices, and less random exploration, when faced with more familiar styles. We estimate nested Bayesian bandit models of learning tastes on the data from the two treatments. They are adapted to consumer choices by allowing for satiation of utility with repeated consumption of goods in the form of diminishing experienced marginal utility. Our best performing model features decreasing random exploration of the choice options. Early exploration is more intensive in the unfamiliar treatment and persists throughout the experiment in both treatments. Overall, subjects make choices that deviate from their best prediction 61.5% of the time in the unfamiliar treatment versus 44.7% of the time in the familiar treatment. The nature of exploration however differs between treatments: subjects adopt a strategy of randomization in the

[1]Most goods are experience goods (Nelson, 1970) that need to be consumed before preferences for them can be discovered. This is true even if subjects have stable innate preferences for goods and if they are able to evaluate precisely how satisfied they are with their consumption.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

unfamiliar treatment (50% always randomize), while they also adopt a strategy of diversification in the familiar treatment (they never choose randomly all the time).

We observed substantial heterogeneity in consumer behavior. Estimated exploration rates were five times higher (resp. 3.4) for subjects with high openness in the unfamiliar (resp. familiar) treatment. We also observe differences between high explorers and low explorers. In the familiar treatment, the diversification strategy adopted by high explorers protected them from excessive satiation and resulted in higher experienced utility overall. However, exploration was not beneficial overall in the unfamiliar treatment as high explorers, choosing randomly, took the risk of consuming too much of a variety they did not like, and low explorers converged sooner on what they liked better but exposed themselves to stronger satiation.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the experimental design and data. A Bayesian bandit model of consumer choice with satiation and random exploration is then presented in Section 4. A reduced form estimation is provided in Section 5, and the structural estimation appears in Section 6. Section 7 provides evidence of the exploration-exploitation trade-off and of individual differences in exploration intensity. Section 8 concludes.

2. Related literature

Conventional economic theory often views consumption choices as static and deterministic, reflecting known and stable individual preferences. However, both consumption and experimental data suggest that choices are inherently dynamic and stochastic. Stigler and Becker (1977) reconciled the dynamic nature of consumption behavior with traditional economic theory by proposing that preferences are acquired through the accumulation of consumption-specific human capital. For example, repeated exposure to music develops 'musical capital,' thereby enhancing one's taste for music. Plott (1996) explained experimental inconsistencies of choice by suggesting that individuals have stable latent preferences, but these preferences are initially unknown and must be discovered through experience (similar ideas can be found in Smith, 1994; Binmore, 1999; Camerer & Ho, 1999). Preference reversals are then caused by some type of myopia when subjects face unfamiliar tasks. With incentives and experience, however, choices become rational and reflect stable preferences. Using data on theater consumption, Lévy-Garboua and Montmarquette (1996) proposed and tested an adaptive model of discovered preferences. Delaney et al. (2020) examine preference discovery focusing on the question of what is to be learned or not.

Another account of the stochastic nature of preferences is random utility theory, in which the utility of a discrete choice cannot be known with full certainty (e.g., Becker et al., 1963; Loomes & Sugden, 1995). Intrinsic randomness would be due to measurement errors, preference changes, taste for diversity, or choice inconsistencies. Hence, the model cannot forecast what choice is made, but it can estimate the probability that a given option in the choice set be selected as a ratio of utilities (McFadden, 2001).

Our experiment describes the process of preference learning as a multi-armed bandit problem. Theoretical literature on multi-armed bandit problems has grown exponentially during the last decades (see Gittins and Jones, 1974 for a deterministic solution in an infinite horizon with exponential discounting, Bergemann and Valimaki, 2006 for economic applications, Scott, 2010 and Burtini et al., 2015 for surveys of stochastic models of exploration-exploitation). However, few experimental tests of single-agent bandit problems have appeared so far in the literature. Banks et al. (1997) first implemented a two-armed bandit where one arm is safe and the other is a simple win-lose (Bernoulli) distribution. Implementing an exponential bandit model, Hudja and Woods (2023) observed that subjects respond to changes in the prior belief, safe action, and discount factor in the predicted direction, yet under-respond to these changes. Steyvers et al. (2009) asked 451 participants to solve

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

a number of bandit problems. Using Bayesian model selection measures that assess how well people adhere to an optimal decision process compared to simpler heuristic decision strategies, they related individual performances to observed differences in cognitive abilities and personality traits. Lévy-Garboua et al. (2024) designed an experiment where 169 high school students solved a bandit problem with a known optimal solution to draw a parallel between their measured creativity and their ability to perform optimally in a bandit task.

3. Preference learning in the lab

3.1. Experimental design

Conducting a lab experiment on preference learning requires goods that satisfy at least four conditions. First, subjects must be able to choose and consume the goods repeatedly in the lab. Second, the goods should be relatively unfamiliar to the subjects so that there is room for preference learning. Third, consuming the goods should be sufficiently pleasant to avoid boredom. Fourth, although relatively uniform, two units of the same good should be sufficiently different to limit satiation. To meet these challenges, we conceived an experiment in which subjects chose and consumed music. Besides, music is a good experimental 'model,' as music consumption has been shown to produce strong rewards, similar to those of food and money (Salimpoor et al., 2011), and consuming music in the lab is no different from streaming music in isolation.

We conducted two treatments. In the 'familiar' treatment, 68 subjects had to choose in each round between four popular musical styles: Pop/Rock, Classical, Rap/RnB, and Jazz/Blues. In the 'unfamiliar' treatment, 72 subjects had to choose in each round between traditional music from four regions of the world: Africa, America, Europe, and Asia.[2] Unlike the musical styles in the familiar treatment, our subjects (mostly university students) were unlikely to have been exposed to traditional music prior to the experiment. As a result, the treatments create asymmetric opportunities for subjects to learn their own preferences through listening experience.

Special care was taken when constructing the set of 400 pieces of music. The objective was to select sufficiently different tracks within each style, while having styles properly partitioned and an equal amount of heterogeneity within each style. To confirm that the tracks selected met these criteria, we conducted a categorization task with external judges drawn from the same subject pool (see Appendix I for the description of the task and the results).

At the start of the experiment, subjects filled out a questionnaire where they indicated their age, gender, personality (Brief Big Five, Barbot, 2011; Costa & McCrae, 1992), and prior frequency of listening to the four musical styles selected for the experiment on a 4-level scale.

The experiment consisted of 50 rounds, each partitioned in 3 stages. In stage 1, the subject selected one of four musical styles. In stage 2, a one-minute sample from the musical style chosen by the subject in stage 1 was played. The subject wore closed-back headphones and had to listen to the entire sample while facing a blank screen. The duration of the sample (one minute) was chosen to be sufficiently long to allow a precise rating of the experience, but short enough to avoid ear fatigue. In addition, an acoustic adjustment was applied to homogenize sound levels within and across music samples so that listening experiences were both comfortable and homogeneous. In stage 3, the subject was asked to rate his listening experience on a 10-point scale. This rating is used as a proxy for the subject's experienced utility (see, for instance, Kahneman et al., 1997). This rating automatically captures any change in utility resulting from increased satiation or appreciation. Furthermore, it can capture the subject's risk attitude, since the decision process takes place in a state of uncertainty.[3]

[2]America includes traditional music from both North and South America, Europe includes Russia, and Asia includes Oceania.

[3]See Glimcher (2009, p. 514).

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

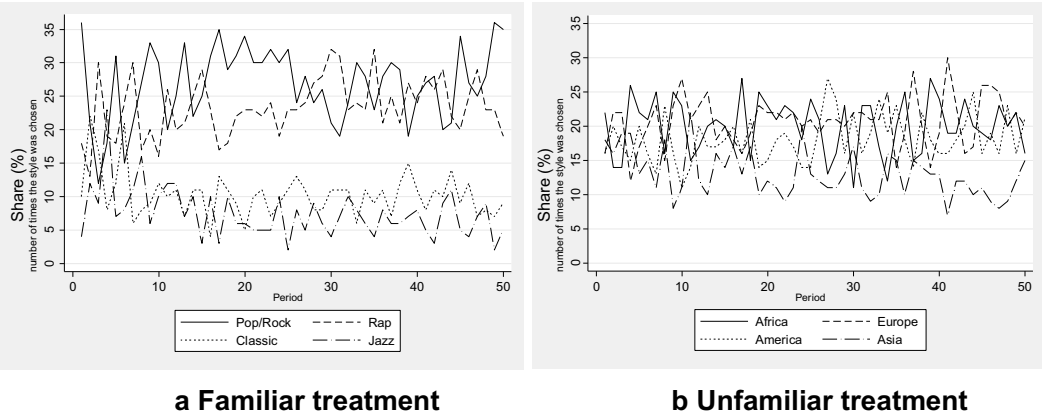


Fig. 1 Evolution of shares of the musical styles over 50 rounds. (a) Familiar treatment. (b) Unfamiliar treatment

Subjects received a show-up fee of 15 euros. In addition, they had strong non-pecuniary incentives to choose a musical style carefully in stage 1, given that they would immediately experience the consequence of their choice (i.e., listening to a one-minute sample) and that they would have to make a large number of choices. To encourage truthful reporting in stage 3, each subject received a personalized 10-track digital recording designed to mirror the experienced utility reported during the experiment. Specifically, a subject was told that 5 sets of 10 randomly-drawn rounds would be compiled at the end of the experiment to create their CD. For each set, we calculated the average experienced utility reported by the subject for the music samples he chose in those rounds. The 10 tracks from the set with the highest average experienced utility were recorded on a device which was then offered to the subject.[4] Thus, subjects had extrinsic incentives to rate their listening experience carefully.

3.2. *From a static to dynamic look at the data*

In the familiar treatment, average listening frequencies show a preference for Pop/Rock (40% of choices) and Rap/RnB (34%) versus Classical (15%) and Jazz/Blues (11%). In the unfamiliar treatment, we observe a low frequency of choices for Asian style (18% of choices) while the remaining choices are almost equally spread between Africa (28%), Europe (25%), and America (29%). From a static perspective, our data can be interpreted as if a representative consumer exhibited a preference for diversity and consumed all musical styles during the experiment.

Figure 1 describes the evolution of the listening shares for each style across the 50 rounds of the experiment. In both treatments, subjects explore a lot during the first 12 rounds or so. Convergence toward preferred styles occurs quickly in the familiar treatment while convergence is limited and delayed in the unfamiliar treatment, essentially away from the Asian style. Consumers maintain diversified preferences, probably in response to the decline in experienced marginal utility of more heavily consumed styles and to the persistent uncertainty regarding the specific sample they will have to listen to in each round.

3.3. *Exploration and convergence*

We provide a characterization of the intensive exploration phase and a synthetic proof of convergence by comparing the Herfindahl index (H) calculated over three time-intervals: 1–12, 13–25, and 26–50.

[4]The experiment was conducted in 2009, when CDs were still popular.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Table 1 Mean Herfindahl across time intervals and treatment

Time intervals	1 to 12	13 to 25	26 to 50	t-test[a] (p-values) 1 − 12 vs 13 − 25	t-test[a] (p-values) 13 − 25 vs 26 − 50
Familiar	0.42	0.54	0.52	<0.001	0.522
Unfamiliar	0.29	0.37	0.37	<0.001	0.843
t-test[b] (p-values)	<0.001	<0.001	<0.001		

Note: $H = \sum_{j=1}^{4} share_j^2$ is computed for each individual, where $share_j$ is the proportion of rounds in which style j was chosen over a given time-interval. H may vary between 0.25 and 1.
[a]Two-sided paired t-tests (equality of means).
[b]Two-sided unpaired t-tests (equality of means).

Table 2 Mean experienced utility across time intervals and treatment

Time intervals	1 to 12	13 to 25	26 to 50	t-test[a] (p-values) 1 − 12 vs 13 − 25	t-test[a] (p-values) 13 − 25 vs 26 − 50
Familiar	5.66	5.82	5.47	0.189	0.001
Unfamiliar	5.07	4.96	4.72	0.345	0.012
t-test[b] (p-values)	0.002	< 0.001	< 0.001		

[a]Two-sided paired t-tests (equality of means).
[b]Two-sided unpaired t-tests (equality of means).

The Herfindahl index is defined as $H = \sum_{j=1}^{4} share_j^2$ where $share_j$ denotes the share of consumption style j for a subject over a given time interval. It captures how subjects' selection of musical styles is concentrated during a given interval, notwithstanding which style they choose. For instance, subjects who listened three times to each of the four competing styles over 12 consecutive rounds will have a minimal H index value of 0.25. Meanwhile subjects who listened three times to two styles, but only twice to a third style and four times to the fourth style, yield the second smallest H value of 0.2639.

Table 1 reports the Herfindahl values for the two treatments. The time interval (1–12) seems to be a phase of intensive exploration in both treatments, such that the H index takes its minimal value on average. Convergence toward more concentrated choices occurs after this initial phase though, as expected, preferences are less concentrated for unfamiliar styles. After this exploration phase, the mean value of H increases significantly for familiar and unfamiliar musical styles, remaining almost constant thereafter.[5]

3.4. Repeated consumption and satiation

Table 2 reports the mean experienced utility in both treatments across time intervals 1–12, 13–25, and 26–50. Experienced utility appears uniformly lower in the unfamiliar treatment, but the two treatments follow a parallel trend. Subjects experience significantly lower utility levels on average during the second half of the experiment, although they explore less in later rounds and are thus more prone to choose their preferred styles. This confirms the erosion of marginal experienced utility through repeated consumption, that is, satiation.

In the next section, we introduce a Bayesian bandit model of taste and consumer's choice that will be estimated econometrically on the experimental data to provide further insights on the dynamic and stochastic aspects of learning.

[5]A non-parametric Pearson chi^2 test for equalities of median Herfindahl gives similar results.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

4. Bayesian bandit model of learning taste

4.1. Learning model of taste with satiation

Consumers face, for each experience *good* like wine or music, a set of heterogeneous *products* classified into *varieties* sharing common features. For example, when selecting a wine, one could choose between Pinot and Cabernet, each forming a variety. Likewise, when listening to music, one could choose between Pop/Rock and Classical music, each forming a different style. The model assumes that consumers have a taste for each variety of good, by choosing for each round one variety of this good and consuming one product randomly drawn from this variety.

Set-up. Each round $t \geqslant 1$ consists of three phases. In phase 1, agent i selects a variety j among J mutually exclusive alternatives. His discrete choice is described by a set of dichotomous variables x_{jt} s.t. $x_{jt} = 1$ and $x_{j't} = 0 \,\forall\, j' \neq j$.[6] In phase 2, the agent consumes a randomly drawn product from variety j. When he selects a variety in phase 1, the agent does not know precisely the marginal utility he will experience from consumption of the product randomly drawn from this variety in phase 2. In phase 3, the agent updates their beliefs based on the marginal utility they experienced from that consumption.

Tastes. We assume that the agent has a stable but imperfectly known *latent taste* for each variety. The taste for variety j follows a normal distribution $v_{jt} \sim \mathcal{N}(\mu_j, h_j)$, where the mean μ_j is the unknown latent taste and h_j is the variance assumed to be known by the consumer.[7] μ_j is to be learned by the agent through repeated consumption over time.

Taste for variety j in round t, v_{jt}, is not directly experienced by the consumer but can be discovered by him through realizations of the *experienced marginal utility* u_{jt}. The latter generally decreases with repeated consumption of the variety within the duration of the experiment, due to satiation, and deviates deterministically from v_{jt}. We describe this relation by:

$$u_{jt} = v_{jt} + \frac{\alpha_j}{1 + n_{jt}} \tag{1}$$

where n_{jt}, $(0 \leqslant n_{jt} \leqslant t)$ designates the number of times variety j was consumed in the first t rounds (i.e., $n_{jt} = \sum_{\theta=1}^{t} x_{j\theta}$), and α_j is a variety-specific parameter assumed to be known by the agent. We assume that α_j is non-negative and describes the law of decreasing marginal utility.[8] While we expect satiation to occur, marginal experienced utility might increase for certain varieties if taste was acquired through repeated exposure.[9] Unlike the distribution of taste v_{jt}, the distribution of experienced marginal utility is not stable over time when $\alpha_j \neq 0$, that is $u_{jt} \sim \mathcal{N}\left(\mu_j + \frac{\alpha_j}{1+n_{jt}}, h_j\right)$.

After consuming a product from variety j, the agent discovers his taste for this variety, v_{jt}.[10] He needs to sample many products from the same variety to get a precise estimate of his mean taste μ_j for this variety.

If the agent knew every latent taste μ_j, then his latent preferences would produce a stable ranking for the J varieties. However, because he does not know μ_j in any round t, consequently latent ranking is also unknown. Instead, the agent's choices entail an unstable ranking of the J varieties. Unlike latent preferences, choices can be path-dependent and endogenous.

[6] To ease notations, we omit the index i in this section. The index is reintroduced in Section 5.

[7] We could also assume that the agent holds beliefs and learns about h_j (see DeGroot, 1970, p. 168). However, this complication does not seem to be necessary on our data.

[8] Satiation is a temporary phenomenon, which is expected to occur within the duration of an experiment like ours'. Thus, n_{jt} refers to a limited interval of time, and needs to be reinitialized beyond this horizon.

[9] In this case, parameter α_j would be negative, indicating that marginal experienced utility rises on average from an initial level of $\mu_j + \alpha_j$ to a higher asymptotic level μ_j. The estimated value of α_j will capture the net effect of satiation and taste acquisition.

[10] Since products are tasted sequentially, the product index and the time index coincide.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Taste Learning. In our model of taste learning, the agent's choices converge toward his latent preferences. After consuming a random product from variety j in round t (i.e., $x_{jt} = 1$), the agent updates his prior beliefs (i.e., estimate of own taste) by applying Bayes' rule, based on the information contained in the signal u_{jt}. For varieties which are not consumed in round t, beliefs do not change. The prior distribution of beliefs is assumed to be normal, with known mean and variance (μ_{j0}, k_{j0}). We will also assume in the following that k_{j0} is homogeneous across agents. The agent predicts he will have taste μ_{j0} for variety j at round 1. After consuming a random product from this variety, he transforms the mean of his normally distributed prior taste in a Bayesian fashion (see DeGroot, 1970, chapter 9). The law of motion for the *predicted taste* μ_{jt} and variance k_{jt} in subsequent rounds is as follows (see Appendix III)[11]:

$$\mu_{jt} = (1-\rho_{jt})\mu_{j,t-1} + \rho_{jt}\upsilon_{jt};\ k_{jt} = (1-\rho_{jt})k_{j,t-1};\ \text{with}\ \rho_{jt} = x_{jt}\frac{k_{jt}}{h_j} \tag{2}$$

ρ_{jt} can be interpreted as a 'reinforcement rate' $(0 \leqslant \rho_{jt} < 1)$, that is, the weight given to the informative signal υ_{jt} relative to that of the prior μ_{jt-1}. The three equations (2) determine the three unknowns $\mu_{jt}, k_{jt}, \rho_{jt}$ in each round t, given the fixed parameters μ_{j0}, h_j, α_j, and the observables x_{jt} and u_{jt}.

Alternatively, the first equation in (2) can be written as $\mu_{jt} - \mu_{jt-1} = \rho_{jt}(\upsilon_{jt} - \mu_{jt-1})$ and interpreted as an adjustment of predicted taste proportionally to the 'reward prediction error' $(\upsilon_{jt} - \mu_{jt-1})$.[12] The predicted taste μ_{jt} of an experienced variety converges through the learning process and the reinforcement rate tends toward zero when a given variety is experienced repeatedly. Note that convergence of the reinforcement rate toward zero *predicts* the tendency for learning curves to be steep initially and then flatter (see Blackburn, 1936). Equation (2) also demonstrates that first impressions matter a lot for less familiar varieties, characterized by a high variance of prior beliefs k_{j0} and, plausibly, by a low prior taste μ_{j0}.

Since taste is not directly experienced by the consumer in the presence of satiation, we estimate the law of motion of mean experienced marginal utility $\mu'_{jt} \equiv \mu_{jt} + \frac{\alpha_j}{1+n_{jt}}$, for all $t \geqslant 0$, provided parameter α_j is known by the econometrician. It can be derived from the law of motion of taste in equation (2), with the help of definition (1) to express taste as a function of experienced utility. An explicit recurrence equation for updating the mean experienced marginal utility μ'_{jt} is provided by:

$$\mu'_{jt} = (1-\rho_{jt})\left(\mu'_{j,t-1} - \frac{\alpha_j x_{jt}}{(1+n_{jt})(1+n_{jt-1})}\right) + \rho_{jt}u_{jt},\ \forall t \geqslant 1 \tag{3}$$

which is the law of motion of taste with an additional term $-(1-\rho_{jt})\frac{\alpha_j x_{jt}}{(1+n_{jt})(1+n_{jt-1})}$. The latter converges to 0 when n_{jt} rises indefinitely. After only four experiences, it is divided by 20. Thus, even if $\alpha_j \neq 0$, the predicted mean taste and mean experienced marginal utility should eventually converge with experience to a common value $(\mu'_j = \mu_j)$.

4.2 Decision model with random exploration

We assume that the agent's decision utility U_{jt} in phase 1 of round t incorporates a random component ξ_{jt} observed by the agent during this phase.[13] If the agent correctly anticipates a decrease in

[11] Distinguishing between predicted utility (decision phase) and experienced utility (consumption phase) is generally supported by fMRI studies which show their correspondence with two separate neural mechanisms (also found in the case of music by Salimpoor et al., 2011).

[12] See Schultz et al. (1997). Dopamine, which has been proven to play a role in preference formation, is known to encode the reward prediction error (see Caplin and Dean, 2008 for an axiomatic approach).

[13] For the distinction between experienced utility and decision utility, see for instance Kahneman and Thaler (2006).

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

experienced marginal utility if variety j were to be consumed in the next round, then his stochastic decision utility is given by equation (4):

$$U_{jt} = \mu_{j,t-1} + \frac{\alpha_j}{2 + n_{jt-1}} + \xi_{jt}, \tag{4}$$

where $\mu_{j,t-1} + \frac{\alpha_j}{2+n_{jt-1}}$ is the predicted marginal experienced utility and ξ_{jt} is a realization of the random decision component $\xi_{jt} \sim Logistic\,(0, \tau)$ of zero mean and standard deviation τ.

Several reasons for introducing a stochastic element in consumer choices can be found in the 'random utility' literature, which rests on a static decision model. In contrast, our consumer faces a stochastic bandit problem, and we interpret the random component of decision utility as originating from the nature of experience goods and from the random exploration of different varieties. We do not estimate the standard version of the bandit model in which the reward distribution is not assumed because it is difficult to apply and restricted to special cases. Instead, we adopt the 'randomized probability matching' heuristic put forward by the multi-armed bandit literature, which is compatible with Bayesian updates of the posterior distribution. This Bayesian modern approach can be more easily implemented and may even outperform the standard 'optimal' solution, as explained by Scott (2010, p.640).[14] In equation (5) below, the agent deliberately matches the computable probability of selecting variety j in round t with its probability of being optimal conditional on his updated beliefs. Thus, randomized probability matching endogenizes the random utility model. Ex ante (i.e., before ξ_{jt} is realized), the probability that variety j be selected in round t is:

$$P_{jt} = P\left(x_{jt} = 1\right) = \int_{\Theta_{jt}} \phi\left(\xi_t\right) d\xi_t \tag{5}$$

where $\xi_t = \left(\xi_{1t}, \ldots, \xi_{Jt}\right)'$, $\Theta_{jt} = \left\{\xi_t \left| U_{jt} > U_{j't}\ \forall j' \neq j\right.\right\}$ and $\phi\left(.\right)$ is the pdf of the logistic distribution with zero mean and covariance matrix $\tau^2 I_j$ (Train, 2009).

Observe that this decision model is general since it captures the full range of behavior from pure exploitation to pure exploration. When $\tau = 0$, decision and predicted utilities are equal, according to equation (4), and decision behavior is deterministic. In every round t, the agent maximizes his predicted marginal utility conditional on the information acquired from prior experiences and selects $j^* = Argmax_{j \in J}\left(\mu_{j,t-1} + \frac{\alpha_j}{2+n_{jt-1}}\right)$, so that $P_{j^*t} = 1$. This is a pure exploitation scenario. In contrast, an agent does not systematically select the variety he currently believes is the best for himself when $\tau > 0$. Instead, he learns his preferences through exploration, by randomly sampling varieties. When τ is large, every variety is selected with almost equal probability, leading to a scenario of pure exploration.

Random exploration is necessary for an agent to learn their latent preferences and draw the most from consumption experience in the long run. Indeed, as in the standard Bayesian learning model (see DeGroot, 1970), an agent is only guaranteed to fully learn his latent preferences (i.e., $\left(\mu_{jt}, k_{jt}\right) \rightarrow \left(\mu_j, 0\right)\ \forall j$) when every variety has a positive probability of being sampled almost all the time. This is not the case when $\tau = 0$ as some varieties may only be experienced a finite number of times.[15] In contrast, $P_{jt} > 0$ when $\tau > 0$, in which case the agent fully learns his preferences.

[14]With minor adaptations (in brackets): 'The multi-armed bandit problem is notoriously resistant to analysis, although optimal [nonparametric] solutions are available in certain special cases (Gittins & Jones, 1974). The optimal solutions are typically hard to compute, rely on artificial discount factors, and fail to generalize to realistic reward distributions [like the normal or logistic distributions]. They can also exhibit *incomplete learning*, meaning that there is a positive probability of playing the wrong arm forever (Brezzi & Lai, 2000). […] [Consumers facing heterogeneous products of the same variety can assume] that a model, such as a probit or logistic regression, determines the reward distributions of the different arms. Assuming a parametric model allows the bandit to focus on a lower-dimensional parameter space and thus potentially achieve greater rewards than "optimal" solutions that make no parametric assumptions.'

[15]For instance, consider an agent initially familiar with a low variance variety j (i.e., $\mu_{j0} \cong \mu_j$, k_{j0} and h_j small) who has low initial beliefs about the alternatives (i.e., $\mu_{j'0} \ll \mu_{j0}\ \forall j' \neq j$) and doesn't feel satiation. When $\tau = 0$, the agent consumes j in period 1 as it has the highest predicted utility μ_{j0}. Because the agent is already familiar with j and because h_j is small,

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

4.3. Decreasing random exploration

An agent relies on random exploration to learn about varieties of goods. However, once the agent has acquired better knowledge of his preferences, random exploration can be expected to fade. This is especially true in finite horizon. If decline in random exploration is slow enough, the agent will select his preferred style $j^* = Argmax_{j\in J}\left(\mu_j\right)$ with a probability that converges to 1. To model decreasing random exploration, we specify the exploration parameter as $\tau^i = \tau_0 \exp\left(X\beta - \lambda t + \delta^i\right)$, where t is the round, $\lambda \geqslant 0$ captures decreasing exploration, τ_0 is a positive constant, X a vector of individual characteristics and $\delta^i \sim \mathcal{N}\left(0, \sigma_\tau^2\right)$ is the individual-specific random effect in equation (4). $\lambda = 0$ yields the Bayesian model with constant random exploration.

4.4. Summary

In the Bayesian bandit model of consumer choice, an agent selects variety j with highest stochastic decision utility in phase 1 of round t; then consumes j in phase 2, which generates an experienced marginal utility u_{jt} drawn from $\mathcal{N}\left(\mu'_j, h_j\right)$; and in phase 3, the agent applies Bayes' rule to update his beliefs about μ'_j from $\mathcal{N}\left(\mu'_{j,t-1}, k_{j,t-1}\right)$ to $\mathcal{N}\left(\mu'_{jt}, k_{j,t}\right)$ based on the information contained in u_{jt}. Over time, μ'_{jt} converges toward μ_j for all j, and the agent fully learns his latent taste for this variety.

We estimate this Bayesian bandit model of consumer choice on our experimental data in two steps. We first implement a reduced form analysis in Section 5 and conduct a structural estimation of the model in Section 6.

5. Reduced form analysis

In this section, we report panel data estimation of experienced utilities. For each musical style, we first convert the reported listening frequency into a prior marginal utility μ'_{ij0} using linear interpolation at the individual level.[16] This makes priors comparable to the reported marginal experienced utilities. Then, we conduct a panel data regression analysis to estimate experienced utilities:

$$u_{ijt} = \mu'_{ij} + \varepsilon_{ijt} = \mu_{ij} + \frac{\alpha_j}{1 + n_{ijt}} + \varepsilon_{ijt} = \bar{\mu}_j + \frac{\alpha_j}{1 + n_{ijt}} + \eta_{ij} + \varepsilon_{ijt} \tag{6}$$

where $\bar{\mu}_j$ is average latent taste for variety j across subjects, $\eta_{ij} \sim \mathcal{N}\left(0, \sigma^2_{\eta_j}\right)$ is a variety and individual-specific random effect, and $\varepsilon_{ijt} \sim \mathcal{N}\left(0, h_j\right)$ is the residual of the random experienced utility. This allows us to estimate the α_js and the h_js which are assumed to be known by subjects but not observed by the econometrician. We implemented this analysis at two levels of aggregation, considering varieties, first as treatments of familiar and unfamiliar music (Table 3), then as musical styles within each treatment (Table 4). In both Tables, the average prior experienced utilities $\overline{\mu'}_{j0}$ (the average value of μ'_{ij0} across subjects) and results of the panel regressions of equation (6) are reported.

Table 3 presents the results of a variant of equation (6), in which all styles within a treatment are aggregated, and all model parameters (σ^2_η, h, $\overline{\mu}$, α) are allowed to vary across treatments. This allows us to test the statistical significance of treatment effects to validate our assumption that traditional world music would be less known a priori than the so-called 'familiar' genres. The results

consuming j has little impact on the agent's beliefs, so $\mu_{j1} \cong \mu_{j0}$. As a result, j is also selected in period 2 and may be so in a finite number of subsequent periods. Note that j may be selected in every period in this example even if there is another variety j' the agent would ultimately prefer (i.e., $\mu_{j'} > \mu_j$).

[16]The listening frequency f_{ij0} takes discrete values from 1 to 4 and is converted into $\mu'_{ij0} = \frac{1}{3}\left(\left(max\left(u_{ijt}\right) - min\left(u_{ijt}\right)\right) * f_{ij0} + 4min\left(u_{ijt}\right) - max\left(u_{ijt}\right)\right)$, where the *max* and *min* refer to the highest and lowest satisfactions reported by individual i. μ'_{ij0} is the affine function of f_{ij0} which takes value $min\left(u_{ijt}\right)$ if $f_{ij0} = 1$ and value $max\left(u_{ijt}\right)$ if $f_{ij0} = 4$.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Table 3 Aggregate varieties: Familiar versus unfamiliar treatment

	First Step: Panel data estimation for experienced utility		
	Familiar treatment	Unfamiliar treatment	Tests of differences
$\overline{\mu'}_0$	5.63	4.28	$t = 5.18^{***}$
	(1.34)	(1.70)	$(p < 0.001)^{a}$
DMU term $(\hat{\alpha})$	1.84***	0.82**	$\chi_2 = 2.94^{*}$
	(0.43)	(0.41)	$(p = 0.09)^{b}$
Mean latent taste $\left(\overline{\hat{\mu}}\right)$	5.48***	4.81***	$\chi_2 = 13.38^{***}$
	(0.12)	(0.14)	$(p < 0.001)^{c}$
$\hat{h}$	4.86***	4.50***	$\chi_2 = 4.94^{**}$
	(0.12)	(0.11)	$(p = 0.03)^{d}$
$Ln\,(L)$		−15,792.79	

*significant at the 10%-level.
**significant at the 5%-level.
***significant at the 1%-level.
[a]Mean μ'_{ij0}, across individuals i and across styles j; see footnote 16. Two-sided unpaired t-test (equality of means).
[b]Wald test of comparison of the two alpha parameters.
[c]Wald test of comparison of the two intercepts of the model (one for the Familiar treatment, one for the Unfamiliar treatment).
[d]Likelihood ratio test of comparison of a model in which all parameters are allowed to differ (including the h) and a model in which all parameters but the h are allowed to differ.

Table 4 Disaggregated varieties: Musical styles within treatments

	Two-step estimation of Bayesian bandit model						
	First Step: Panel data estimation for experienced utility						
	Familiar treatment			Unfamiliar treatment			
Parameters	Pop/Rock	Rap/RnB	Classic/Jazz	Africa	Europe	America	Asia
$\overline{\mu'}_{j0}$ [a]	7.67	5.89	3.32	4.09	4.56	4.30	4.19
	(2.60)	(3.13)	(1.80)	(2.99)	(2.94)	(2.91)	(2.82)
DMU term $(\hat{\alpha}_j)$	1.33**	2.68***	0.41	3.98***	2.54***	−0.83	0.01
	(0.55)	(0.57)	(0.54)	(0.55)	(0.56)	(0.58)	(0.54)
Mean latent taste $(\bar{\mu}_j = \overline{\mu'}_j)$	5.43***	4.62***	5.24***	4.30***	4.53***	5.14***	3.70***
	(0.16)	(0.19)	(0.24)	(0.18)	(0.19)	(0.19)	(0.22)
$\hat{h}_j$	4.93***	4.50***	3.43***	4.10***	4.37***	4.38***	2.81***
	(0.20)	(0.19)	(0.17)	(0.19)	(0.20)	(0.21)	(0.16)
$Ln\,(L)$	−3024.00	−2592.37	−1877.88	−2188.32	−2299.02	−2014.64	−1327.97

Estimation of equation (6): Linear random-effect model, estimated by maximum likelihood.
**significant at the 5%-level.
***significant at the 1%-level.
[a]Mean μ'_{ij0}, see footnote 16.
In parentheses, we report the standard deviations of the variables μ'_{ij0}, not the standard deviation of the means, which are much lower.

confirm that subjects, on average, had more prior experience of familiar genres than unfamiliar ones $(t = 5.18,\ p < 0.001)$ and a higher latent taste as well (Wald test $\chi^2 = 13.38,\ p < 0.001$).

In the within-treatment analysis (Table 4), Classical music and Jazz/Blues are aggregated into one style to increase their share and facilitate the econometric analysis.[17] As expected, subjects have clear priors about their least and most liked music in the familiar treatment ($\overline{\mu'}_{j0}$ is highest for Pop/Rock

[17]Classical music and Jazz form a homogenous style. Prior listening frequency indexes are almost equal and low for Classical and Jazz (resp. 1.75(0.90) and 1.66 (0.82)). Individual listening frequencies of familiar styles for the whole experiment are all negatively correlated with the exception of Classical and Jazz (corr. = 0.116). All average listening frequencies per round are also strongly negatively correlated with the exception of Classical and Jazz (corr. = 0.037). Moreover, satisfaction indexes for Classical and Jazz are highly and significantly correlated (0.56***), and they exhibit similar correlations with the other styles (resp. 0.55*** and 0.48*** for Pop, and − 0.05 and 0.10 for Rap).

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

and lowest for Classical/Jazz), while subjects assign similar, intermediate values to all four musical styles on average in the unfamiliar treatment.

The DMU term (α_j) is positive and significantly different from 0 only for the most popular styles, Pop/Rock, Rap/RnB, African, and European style (two-sided test), suggesting that music listening is not subject to strong satiation as it takes lots of repetition to identify a significant declining trend in marginal experienced utility. However, Table 3 shows that DMU terms are positive on average at an aggregate level in both treatments at the 5% level of significance at least.

Table 4 shows that the $\bar{\mu}_j$s and the $\overline{\mu'}_{j0}$s behave differently in both treatments both in terms of magnitude as well as ranking. In the familiar treatment, initially low-ranked styles (Classical and Jazz) are liked more than anticipated, whereas the highest-ranked style (Pop/Rock) is liked less. At the end of the learning process, Classical/Jazz comes close to Pop/Rock and even surpasses Rap/RnB.[18] By contrast, in the unfamiliar treatment, the $\bar{\mu}_j$s never differ significantly from the $\overline{\mu'}_{j0}$s.

For a rational subject with stable known preferences, marginal utility would be equalized over the whole experimental period for all consumed musical styles. A comparison of the distribution of mean latent tastes $(\bar{\mu}_j = \bar{\mu}'_j)$ with prior experienced utility $(\overline{\mu'}_{j0})$ in Table 4 reveals that this is approximately verified in the familiar treatment but not in the unfamiliar treatment. We observe a significant difference between prior experienced marginal utilities and a partial convergence in the familiar treatment, yet a diverging trend in the unfamiliar treatment.[19] While learning has taken place in both treatments, the process is converging in the familiar treatment but remains incomplete in the unfamiliar treatment. In this treatment, there is a striking difference between a quasi-indifference in the prior experienced utilities and the wide appreciation gap between Asian music (least appreciated in our sample) and other styles. Clearly, subjects consciously avoided the musical style that they liked least, as already apparent in Figure 1b.

Finally, the variances $\hat{h}_j$ are precisely estimated in Tables 3 and 4, which legitimates our assumption of a known variance of taste h_j.[20] Moreover, variances are relatively similar across musical styles and treatments, which provides some evidence that tracks used in the experiment portray a similar degree of heterogeneity within each style. There are two exceptions, however, as variances of taste are lower for Classical/Jazz and traditional Asian music. Subjects had strong prior beliefs about what they would like least. As a matter of fact, they were wrong for Classical/Jazz and right for Asian musical style.

6. Structural estimation

6.1. Specification and method

The remaining structural parameters to estimate are variances of initial beliefs $\boldsymbol{k}_0 = (k_{1,0}, k_{2,0}, k_{3,0}, k_{4,0})$ for the non-familiar treatment and $\boldsymbol{k}_0 = (k_{1,0}, k_{2,0}, k_{3,0})$ for the familiar treatment; and τ^i, the (individual) exploration parameter. The estimation procedure combines the choice probabilities $P_{ijt}(\boldsymbol{k}_0, \tau^i) = P(x_{ijt} = 1)$ in equation (5) with the Bayesian law of motion in equation (3), where h_j is replaced by $\hat{h}_j$ which has been estimated in the previous section.

To account for possible heterogeneity in learning, we specify τ^i as:

$$\tau^i = \tau_0 \exp(-\lambda t + \omega Open_i + \delta^i)$$

[18] In the Familiar treatment: the equality of $\bar{\mu}_j$ with $\overline{\mu'}_{j0}$ is rejected at the 1% threshold for Pop/Rock and Classical/Jazz and at the 10% threshold (6%) for Rap/RnB.

[19] In the Familiar treatment, we systematically reject equality of the mean $\overline{\mu'}_{j0}$ for the 3 pairs at all the usual thresholds, while in the Unfamiliar treatment, we never reject equality for the 6 pairs at all the usual thresholds (the lowest p-limit is 35.7% for Europe versus Asia). This means that the $\overline{\mu'}_{j0}$ are very similar across styles in the Unfamiliar treatment, but very different in the Familiar treatment.

[20] We could also assume that the agent holds beliefs and learns about h_j (see DeGroot, 1970, p. 168). However, this complication does not seem to be necessary on our data.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

In practice, the model is estimated as follows, to ensure that τ_0 stays positive.

$$\tau^i = \exp\left(\beta_0 - \lambda t + \omega Open_i + \delta^i\right) \tag{7}$$

As mentioned above, $\lambda \geqslant 0$ captures decreasing random exploration. Observed heterogeneity is captured by one of the Big Five personality traits, namely openness.[21] Unobserved heterogeneity is captured by the individual-specific random effect $\delta^i \sim \mathcal{N}(0, \sigma_\tau^2)$.

When the model does not include random effects (i.e., $\sigma_\tau^2 = 0$), $\boldsymbol{k}_0$, β_0, λ and ω are estimated by Maximum Likelihood based on the likelihood function:

$$L\left(\boldsymbol{k}_0, \beta_0, \lambda, \omega\right) = \prod_{i,j,t} \left[P_{ijt}\left(\boldsymbol{k}_0, \beta_0, \lambda, \omega\right)\right]^{x_{ijt}}$$

where $x_{ijt} = 1$ if subject i has chosen the style j in round t, and 0 otherwise.

In this case, the model is a multinomial logit model with a non-linear index. It is the non-linearity of the index that allows identification of the β_0, λ and ω parameters.

When the model includes random effects (i.e., $\sigma_\tau^2 \neq 0$), it is a mixed multinomial logit model, once more with a non-linear index. In this case, we use a Maximum Simulated Likelihood approach (see Train (2009) for example). The unconditional choice probability may be written:

$$\bar{P}_{ijt}\left(\mathbf{k}_0, \beta_0, \lambda, \omega, \sigma_\tau\right) = E\left[P_{ijt}\left(\mathbf{k}_0, \tau^i\left(\delta^i\right)\right)\right] \tag{8}$$

where the expectation operator is taken as a function of individual-specific random effect δ^i. We use Monte Carlo simulations to approximate numerically $\bar{P}_{ijt}\left(\boldsymbol{k}_0, \beta_0, \lambda, \omega, \sigma_\tau\right)$. Specifically, we generate sequences of R pseudo-random numbers $\{\delta_i^r\}_{r=1,\ldots,R}$ from $\mathcal{N}(0, \sigma_\tau^2)$ (more precisely, we use Halton sequences with $R = 1000$), and we replace the expectation in (8) by its empirical counterpart. The likelihood is then approximated by:

$$\tilde{L}\left(\boldsymbol{k}_0, \beta_0, \lambda, \omega, \sigma_\tau\right) = \prod_{i,j} \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t} \left[P_{ijt}\left(\boldsymbol{k}_0, \tau^i\left(\delta_i^r\right)\right)\right]^{x_{ijt}} \right\}$$

This second step allows us to estimate the variances of initial beliefs $\boldsymbol{k}_0$ along with components of the exploration parameter. As mentioned above, we do not use true values of h_j, but their first-step estimates $\hat{h}_j$ instead. Because of this, the second-step matrix of variances-covariances is biased. We correct this bias by using the Murphy and Topel (1985) correction.

6.2. Estimation results

Six versions of structural estimation of the Bayesian bandit model of consumer choice are reported below, for the familiar treatment in Table 5 and the unfamiliar treatment in Table 6. M1 (col. 1) is the simple model of pure exploitation (with k_0 only), M2 (col. 2) includes exploration (with $\tau_0 > 0$) without unobserved heterogeneity, M3 (col. 3) is M2 + unobserved heterogeneity σ_τ, M4 (col. 4) is M3 + Openness which accounts for observed heterogeneity, M5 (col. 5) is M3 + decreasing exploration λ, and M6 (col. 6) is M5 + Openness.

In both treatments, a considerable increase of the log-likelihood can be observed when random exploration and decreasing exploration are added respectively to the model, which is consistent with predictions of the multi-armed bandit literature. Unobserved heterogeneity also improves the model's fit with high significance. Openness is the only observable factor of heterogeneity that was found significant, but it has a strong impact on exploration intensity, which conforms with intuition.

[21] We tried different specifications, by including the other Big Five (conscientiousness, extraversion, agreeableness, and neuroticism), and/or the age and gender of the subject. Those variables were never statistically significant.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Table 5 Structural estimation of Bayesian bandit model: Familiar treatment

Two-step estimation of Bayesian bandit Model						
Second step						
Familiar treatment	Pure exploitation M1	Exploration-exploitation M2	M2 + Unobs.het. M3	M3 + Open. M4	M3 + Dec.explor M5	M5 + Open. M6
			k_0			
$k_{1,0}$	8.88***	3.91**	6.89***	6.52***	1.13***	1.00***
Pop/Rock	(1.86)	(1.54)	(2.64)	(2.50)	(0.42)	(0.36)
$k_{2,0}$	32.53**	6.53***	4.23***	4.20***	6.46**	6.09**
Rap/RnB	(15.68)	(2.18)	(1.54)	(1.48)	(3.3)	(2.91)
$k_{3,0}$	2.74***	0.78***	0.42***	0.43***	0.64***	0.68***
Classical/Jazz	(0.35)	(0.16)	(0.10)	(0.11)	(0.15)	(0.16)
			$\tau\ (\tau = \exp(\beta_0 - \lambda t + \omega Open_i + \delta^i))$			
$\beta_0\ (\cdot = log(\tau_0))$		0.93***	1.10***	−0.48	1.95***	−0.07
		(0.04)	(0.11)	(0.64)	(0.14)	(0.73)
$\omega \cdot (= \beta_{Openness})$				0.43**		0.55***
				(0.17)		(0.20)
τ_0[a] $(= exp(\beta_0))$		2.53***	3.00***	3.01***	7.03***	7.01***
		(0.10)	(0.33)	(0.32)	(0.98)	(0.93)
$\tau \cdot (\text{at } min(Openness))$				1.67**		3.32***
				(0.42)		(0.97)
$\tau \cdot (\text{at } max(Openness))$				4.34***		11.17***
				(0.82)		(2.48)
σ_τ (variance of δ^i, random effect in τ)			0.77***	0.74***	0.79***	0.72***
			(0.10)	(0.09)	(0.10)	(0.09)
λ					0.036***	0.037***
(associated to t)					(0.003)	(0.003)
$Ln(L)$	−3734.26	−3194.24	−3086.06	−3084.13	−2992.65	−2989.50

[a] τ (at $mean(Openness)$) for models M4 and M6.
**significant at the 5%-level.
***significant at the 1%-level.
Standard errors corrected by using the Murphy and Topel (1985) correction, to consider the use of $\hat{h}_j$s instead of the true h_js.

The estimated variances of the initial priors, k_{j0}, are quite sensitive to the specification and appear to be biased upward when exploration is not accounted for. These variances are smallest for Classical/Jazz and Pop/Rock, the musical styles that subjects were most familiar with before the experiment. Even with styles that were more familiar to them, subjects' prior knowledge of preferences was imperfect, since the estimated variances of the initial priors are always highly significant. Thus, there was room for learning in both treatments. However, variances of the prior distribution of tastes are significantly larger in the unfamiliar treatment, which confirms our assumption that subjects knew little about their taste for traditional world music before the experiment.

The exploration parameter τ_0 is between 1.8 and 3.8 times higher in the unfamiliar treatment, which indicates that subjects relied more on random exploration to learn about the musical styles that they were less familiar with. Decreasing exploration (λ) is higher in absolute value for the unfamiliar treatment, however. The combined result on τ_0 and λ indicates that early exploration was more intensive for less familiar musical styles, yet exploration decreased more rapidly. The half-life duration for the exploration parameter is of 12.84 rounds (s.d. = 0.95) in the unfamiliar treatment and of 18.73 rounds (s.d. = 1.52) in the familiar treatment.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Table 6 Structural estimation of Bayesian bandit model: Unfamiliar treatment

Structural estimation of Bayesian bandit model						
Second step						
Unfamiliar treatment	Pure exploitation M1	Exploration-exploitation M2	M2 + Unobs.het. M3	M3 + Open. M4	M3 + Dec.explor M5	M5 + Open. M6
			k_0			
$k_{1,0}$Africa	14.96***	14.57*	15.88**	15.98***	11.13***	11.17***
	(2.07)	(7.63)	(7.18)	(4.05)	(1.33)	(1.36)
$k_{2,0}$Europa	11.31***	6.98***	5.16***	5.10***	6.99***	6.99***
	(1.51)	(2.33)	(1.44)	(0.74)	(0.97)	(0.99)
$k_{3,0}$America	12.28***	14.44	3.91***	3.91***	5.20***	5.22***
	(1.76)	(9.04)	(0.56)	(0.19)	(0.17)	(0.19)
$k_{4,0}$Asia	5.22***	67.36	14.86	14.73*	15.10	14.97
	(0.70)	(214.71)	(8.97)	(7.99)	(17.74)	(18.12)
		$\tau\ (\tau = \exp(\beta_0 - \lambda t + \omega Open_i + \delta^i))$				
$\beta_0\ (= log\ (\tau_0))$		1.51***	2.06***	0.39	3.29***	1.16
		(0.07)	(0.22)	(1.66)	(0.25)	(1.56)
$\omega\ (= \beta_{Openness})$				0.47		0.60
				(0.46)		(0.43)
τ_0[a] $(= exp\ (\beta_0))$		4.53***	7.85***	7.86***	26.84***	27.21***
		(0.32)	(1.73)	(1.48)	(6.71)	(6.80)
τ(at $min(Openness)$)				3.68*		10.26***
				(2.87)		(7.63)
τ(at $max(Openness)$)				12.93***		51.57***
				6.59)		(26.69)
σ_τ (variance of δ^i, random effect in τ)			1.24***	1.22***	1.41***	1.38***
			(0.19)	(0.01)	(0.01)	(0.25)
λ					0.052***	0.053***
(associated to t)					(0.003)	(0.003)
$Ln\ (L)$	−5828.77	−4871.42	−4786.03	−4785.28	−4673.61	−4672.51

[a] τ (at $mean(Openness)$) for models M4 and M6.
*Significant at the 10%-level.
**significant at the 5%-level.
***significant at the 1%-level.
Standard errors corrected by using the Murphy and Topel (1985) correction, to consider the use of $\hat{h}_j$s instead of the true h_js.

To better appreciate learning differences across subjects, we introduce unobserved heterogeneity in model M3 of Tables 5 and 6. The term σ_τ is always highly significant and hardly diminishes when a personality trait like openness is introduced in model M4. Therefore, subjects appear to have learned their preferences in different ways during the experiment: some were more likely to choose the style they most preferred at the time, whereas others tended to sample styles to better identify their preferences.

The impact of openness on the exploration rate τ is visible in models M4 and M6. In our preferred model M6, exploration rates of subjects with the maximum score of openness are 3.4 times those of subjects with the minimum score in the familiar treatment and 5.0 times those in the unfamiliar treatment. Thus, while everybody tends to explore more when musical styles are unfamiliar, more open people explore far more than others.

In the next section, we further examine the convergence process and the heterogeneity of learning strategies used by subjects.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

7. Discussion

7.1. Initial and final beliefs

Table 7 reports summary statistics for the average distance between predicted utility and latent utility at the beginning (columns 1 and 4) and end (columns 2 and 5) of the experiment. The initial distance $|\mu'_{ij0} - \bar{\mu}'_{ij}|$ reflects the average 'projection bias' of subjects in predicting their latent utility (Loewenstein & Adler, 1995; Loewenstein et al., 2003) whereas the final distance $|\mu'_{ij50} - \bar{\mu}'_{ij}|$ is a measure of the average convergence of consumers toward their latent taste. Every mean $|\mu'_{ij0} - \bar{\mu}'_{ij}|$ significantly differs from the corresponding mean $|\mu'_{ij50} - \bar{\mu}'_{ij}|$ and all those mean values differ significantly from 0 ($p < 0.001$). The projection bias, which is large relative to the prior and relatively larger in the unfamiliar treatment, has been significantly reduced by learning during the experiment as shown in columns (3) and (6). However, convergence remains partial since the gap remains significant by the end of the experiment.

7.2. Evidence about the exploration-exploitation tradeoff

7.2.1. Learning curves

Learning curves measure the impact of the exploitation strategy for each musical style. Figure 2 shows the *theoretical* and *empirical* learning curves, which depict the reinforcement rates ρ_{jt} after t rounds of theoretical or actual experience respectively. These curves are plotted using equations (2) and (3), for the estimated values of the parameters (k_{j0}, etc.).[22] The reinforcement rates vary widely across musical genres in the first rounds of the experiment and always converge toward zero as subjects accumulate experience.

The estimated rates of the theoretical learning curves, computed midway in the first exploration time interval (1–12), were equal to 8% for Pop/Rock and Classic/Jazz, and 14% for Asia and Africa, with an intermediate value of 13% for Rap/RnB, Europe and America. The disparity between these reinforcement rates essentially reflects a disparity between the estimated prior knowledge $k_{j,0}$s in

Table 7 Average distance between predicted utility and latent utility of musical styles at the beginning and end of the experiment

	$\lvert\mu'_{ij0} - \bar{\mu}'_{ij}\rvert$ (1)	$\lvert\mu'_{ij50} - \bar{\mu}'_{ij}\rvert$ (2)	t-test of equality of means (3)		$\lvert\mu'_{ij0} - \bar{\mu}'_{ij}\rvert$ (4)	$\lvert\mu'_{ij50} - \bar{\mu}'_{ij}\rvert$ (5)	t-test of equality of means (6)
Pop/Rock	1.88 (1.21)[a]	0.50 (0.40)	$t = 11.09$ ($p < 0.001$)	Africa	2.28 (1.25)	0.43 (0.39)	$t = 11.80$ ($p < 0.001$)
Rap/RnB	1.85 (1.55)	0.43 (0.37)	$t = 6.67$ ($p < 0.001$)	Europe	1.82 (1.28)	0.46 (0.34)	$t = 9.80$ ($p < 0.001$)
Classical/Jazz	0.70 (0.61)	0.31 (0.28)	$t = 8.14$ ($p < 0.001$)	America	1.81 (1.16)	0.42 (0.33)	$t = 9.01$ ($p < 0.001$)
				Asia	2.19 (1.91)	0.33 (0.22)	$t = 8.24$ ($p < 0.001$)

[a]The standard deviation of the *individual* experienced utility difference is displayed in parentheses below the *average* distance (i.e., the absolute value of this difference). The (non-reported) standard deviation of the average distance is much lower than the standard deviation of the variable.

[22]The theoretical curves hypothesize that $x_{jt} = 1$ at each round while the empirical ones hypothesize that $x_{jt} = 1$ if style j was chosen at round t.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

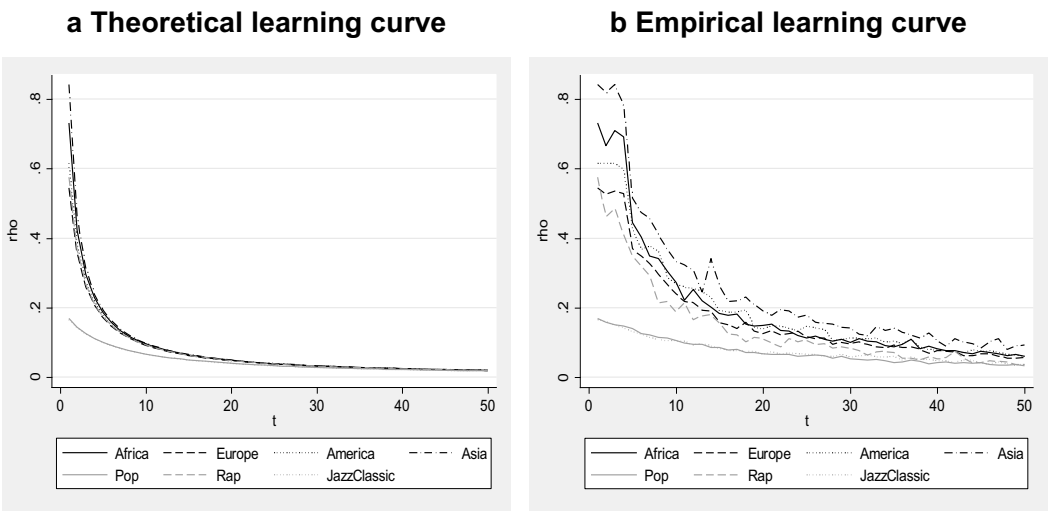


Fig. 2 Theoretical and empirical learning curves $\left(\boldsymbol{\rho}_{jt}\right)$. (a) Theoretical learning curve. (b) Empirical learning curve

Tables 5 and 6 which capture the prior knowledge of musical styles. These results confirm that subjects learned in both treatments and that the learning curve was steeper for less familiar musical styles.[23]

For empirical curves, posterior variances k_{jt} and observed reinforcement rates ρ_{jt} do not depend directly on t but on the frequency of listening to a style j during the first t rounds, since they can only be revised when musical style j is selected. For instance, on average, Rock was listened to 19.8 times during the 50 rounds of the experiment, which resulted in an empirical reinforcement rate of 7.2% per round, whereas Classical/Jazz was listened to 13.1 times only, leading to a higher empirical reinforcement rate of 8.1% per round.

7.2.2. How much do consumers randomize?

Random exploration contributes to long run optimization in consumers' choices. To interpret estimated exploration parameters τ and λ, Figure 3 plots the average probability of not choosing the best predicted style during the experiment. This probability for individual i at time t is equal to $1 - \bar{P}\left(x_{ijt} = 1\right)$, where j is the predicted utility-maximizing musical style at round t for i and $\bar{P}\left(x_{ijt} = 1\right) = \bar{P}_{ijt}\left(\hat{\boldsymbol{k}}_0, \hat{\beta}_0, \hat{\lambda}, \hat{\omega}, \hat{\sigma}_\tau\right)$ is the estimated probability of choosing this style, given by equation (8) as a function of the estimated parameters $\hat{k}_{0j}$ and $\tau^i\left(\delta^i\right)$. Figure 3 illustrates to what extent the average consumer chose a style that did not maximize his predicted utility during our experiment.

Random choices are most frequent throughout the early exploration time interval 1–12, while the frequency of non-maximizing choices stabilizes at a lower level during the rest of the experiment. The latter frequency reaches a high level of 61.5% for unfamiliar styles and 44.7% for familiar styles, and the difference is significant after the first time interval. The evolutions of the two curves are parallel, as shown by a linear fit on time intervals 1–12 and 13–50. Figure 3 is a vivid illustration of the randomness of experienced utility in discrete consumer choices.

[23]The convergence of preferences may also be observed in Table 7 by comparing the average distance between subjects' latent experienced utility and beliefs μ'_{ijt} at the beginning and at the end of the experiment (i.e., $\left|\mu'_{ij0} - \mu'_{ij}\right|$ *vs.* $\left|\mu'_{ij50} - \mu'_{ij}\right|$). In particular, the average distance is divided by 2.3 for Classical/Jazz and by 5.3 for traditional music from Africa.

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

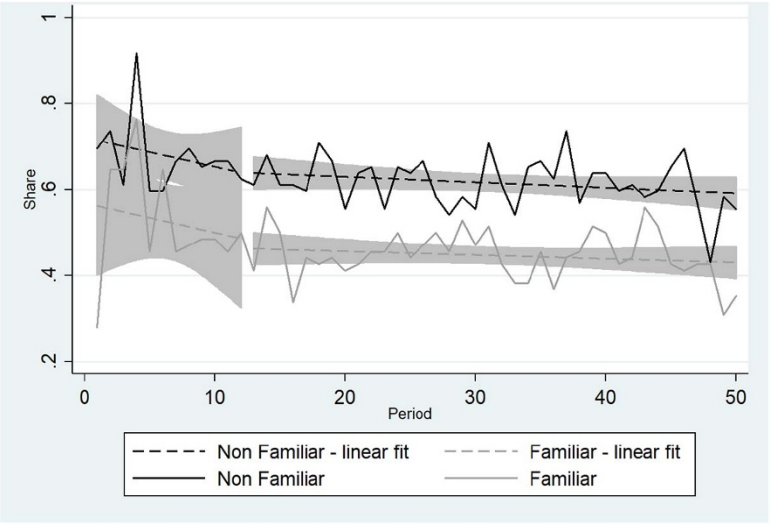


Fig. 3 Share of subjects not choosing the expected utility maximizing style
Note: Grey-shaded areas represent confidence intervals at the 95% level.

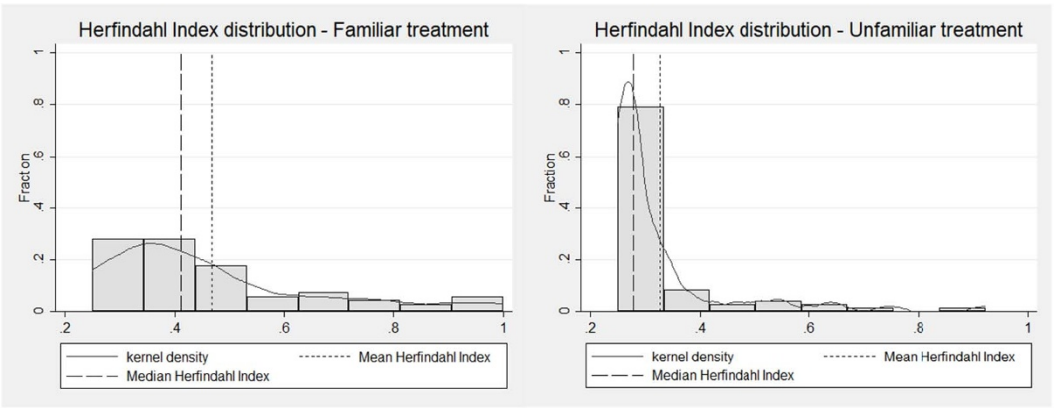


Fig. 4 Histograms of the Herfindahl concentration index for both treatments

7.2.3. Diversification or randomization?

The different degree of knowledge imperfection between familiar and unfamiliar musical styles changes the nature and impact of exploration behavior across treatments in our experimental context. Figure 4 displays the histograms of the Herfindahl indexes for both treatments. Recall that the Herfindahl is an inverse index of exploration intensity.

The Herfindahl index distribution is skewed to the right in both treatments, with mean higher than median, yet major differences exist between treatments. The distribution is far more concentrated in the high exploration zone for the unfamiliar treatment. A Kolmogorov-Smirnov test of comparison of distributions rejects the hypothesis that the Unfamiliar group is more concentrated than the Familiar group $(D = 0.527\ (p - value < 0.001))$ and does not reject the hypothesis that the Familiar group is more concentrated than the Unfamiliar group $(D = -0.003\ \ (p - value = 0.999))$.

Almost 50% of participants randomize all the time to choose a style of traditional world music (median Herfindahl is 0.279), and the cutoff lies between those who randomize throughout the

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

experiment and those who do so only part time, with time left to exploit satisfying experiences. In comparison, amidst the familiar musical styles, participants never choose randomly all the time (first percentile Herfindahl is 0.334) and there is a clear divide between subjects who diversify their music choices more and those who do not.

To summarize, high explorers look like *diversifiers* who took advantage of exploration in the familiar treatment and *randomizers* who failed to do so in the unfamiliar treatment, as will be apparent in the next section.

7.3. Did more intensive early exploration benefit consumers?

If rational consumers have an imperfect knowledge of their preferences, they should learn what they like most and least through early exploration and make better decisions in the long run by exploiting received information. Early exploration is thus expected to benefit consumers more in the unfamiliar treatment than in the familiar treatment.

We test the prediction of a beneficial effect of exploration in Table 8, which reports the mean experienced utility depending on the early exploration intensity, that is, during the first 1–12 time-interval. A proxy is used to capture an individual's early exploration intensity, which is their position relative to the mean Herfindahl index for the initial time interval (1–12). A higher concentration means here a lower early exploration intensity. We defined low (resp. high) explorers as subjects whose Herfindahl index is above (resp. below) the mean value in each treatment.

No matter how much effort they spend on exploration, consumers experience lower utility on average in the unfamiliar treatment, as means differ at the 0.1% level. High explorers in the familiar treatment reported on average a higher mean experienced utility compared to low explorers and benefited from higher choice diversification by listening more frequently to a musical style which they ultimately appreciated (Classical/Jazz). In contrast, high explorers in the unfamiliar treatment were no better-off than low explorers overall. Indeed, while high explorers in this treatment devoted more time to musical styles that they liked less (Asian style), low explorers on the other hand suffered in the long run from listening too much to musical styles they liked.

Table 9 focuses on the evolution of mean experienced utility over the three time-intervals. Comparison of the two treatments shows that the Unfamiliar group experiences on average significantly lower utility than the Familiar group in all time-intervals and for all levels of early exploration[24] except for low explorers in the 1–12 time-interval ($p = 0.916$).

In the familiar treatment, experienced utility is not significantly different between the first and second intervals no matter how intensive was early exploration. This observation corroborates the conventional economic assumption that people have little to learn with familiar music since they have good, albeit imperfect, prior knowledge of their preferences. However, high early exploration slows down the decline in experienced marginal utility for the second half of the experiment as consumers

Table 8 Mean experienced utility and early exploration intensity in both treatments

t = 1 to 50	All	Low early exploration H > H mean	High early exploration H < H mean	t-test[a] (p-values) Low vs High
Familiar	5.61	5.43	5.68	0.004
Unfamiliar	4.87	4.87	4.87	0.983
t-test[b] (p-values)	< 0.001	< 0.001	< 0.001	

[a]Two-sided unpaired t-tests (equality of means).
[b]Two-sided unpaired t-tests (equality of means).

[24]Two-sided unpaired t-tests (equality of means) yield a p-value almost everywhere < 0.001 ($p = 0.003$ for low explorers in the second half of the experiment).

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Table 9 Mean experienced utility and early exploration intensity per time interval in both treatments

	1 to 12	13 to 25	26 to 50	t-test[a] (p-values) 1 − 12 vs 13 − 25	t-test[a] (p-values) 13 − 25 vs 26 − 50
			Familiar treatment		
Low early exploration H > H mean	5.60	5.81	5.14	0.379	0.001
High early exploration H < H mean	5.69	5.82	5.61	0.336	0.085
t-test[b] (p-values) High vs Low	0.768	0.985	0.087		
			Unfamiliar treatment		
Low early exploration H > H mean	5.58	4.64	4.65	0.009	0.970
High early exploration H < H mean	4.93	5.05	4.74	0.226	0.003
t-test[b] (p-values) High vs Low	0.072	0.264	0.794		

[a]Two-sided paired t-tests (equality of means).
[b]Two-sided unpaired t-tests (equality of means).

counterbalance the decreasing marginal utility of the more heavily consumed styles through a greater diversification of consumption. Setting the cutoff at the median or mean does not make a difference and does not alter the conclusion.

The situation is very different for the unfamiliar treatment. Low early exploration is now associated with a very significant decline in experienced marginal utility as soon as the second time interval (13–25). In contrast, high early exploration appears like a costly investment. While mean experienced utility is significantly reduced during the exploration phase (one-tailed test, $p = 0.036$), the benefits drawn from this investment are small and short-lived. Indeed, the experienced marginal utility of high explorers rises insignificantly in the second interval and suffers from a significant fall in the second half of the experiment to the level attained by low explorers. Our results mitigate the benefits from early exploration in the unfamiliar treatment. Indeed, early exploration does not guarantee higher average reward over a finite horizon, since it may also reveal bad surprises that cannot be fully compensated.

8. Conclusion

The traditional economic assumption that consumers have innate, stable preferences and make utility-maximizing choices has been increasingly challenged by findings from econometric models and psychological experiments. This view seems too rigid, especially for experience goods in a world with many varieties and constant product innovation.

We believe that the main virtue of our experimental data about the repeated choices of musical styles has been to uncover the sequential nature of the decision-making process that is usually masked by the product and time aggregation of consumption data. Consumers need time to explore and discover which variety they prefer. They do not take a series of independent static choices but typically face a stochastic multi-armed bandit problem over their lifetime. Rational consumers are uncertain about the mean utility different varieties will provide, and they must choose to explore lesser-known options. Their optimal policy thus contains two types of randomness, one due to their prior ignorance of products and varieties and the other due to their conscious exploration behavior. The two-stage

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

mechanism of decision-making (evaluation and decision) described by neuroscientists (Glimcher, 2009) is consistent with this dual randomness of utility.

Our experimental approach enabled us to observe satisfaction and choice separately, thus disentangling experienced utility from decision utility as advocated by Kahneman et al. (1997). The notions of experienced utility and decision utility coincide in conventional economics due to the neglect of preference uncertainty. We however never assume that experienced utility deviates from decision utility because consumers make errors of perception and/or decision. Instead, their difference is the consequence of rational, or quasi-rational, exploration behavior.

Using disaggregated individual data, we estimated a Bayesian bandit model of consumer choice that accounts for satiation and declining random exploration. Our model explains seemingly unstable choices as optimal behavior under incomplete preferences. Over time, choices stabilize, but short-term fluctuations in marginal utility lead consumers to switch between varieties – which static models interpret as a preference for diversity.

People may know precisely what they like after consuming a single product, but the multiplicity of varieties and products in real life makes exploration essential to discover which one they like better. This may be the main reason for random utility.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2026.10050.

Acknowledgements. We especially thank Peter Klibanoff, Olivier L'Haridon, and Chantal Marlats for helpful comments on previous drafts. We are grateful to Aurélien Baillon, Ido Erev, Charles Noussair, and Véronique Simonnet for early discussions on the paper, and to Todd Lubart for his help in the conceptualization of the categorization task. We acknowledge the contribution of Olivier Armantier in a previous version of this paper. We thank seminar and conference participants at the Paris School of Economics, Erasmus University in Rotterdam, Asfee in Lyon, ESA in Zurich, AEW in Rome, BEELab in Florence, Afse in Lyon, ACEI in Montreal, SABE in Lake Tahoe, University of Bilbao, JMA at Montpellier, UNU-MERIT in Maastricht, SABE-IAREP meeting in Dundee. Louis Lévy-Garboua acknowledges Agence Nationale pour la Recherche for providing experiment funding (ANR-10-CREA-008 MACCAN). The experiments were conducted at LEEP (Paris School of Economics and Centre d'Economie de la Sorbonne). The replication material for the study is available at https://doi.org/10.5281/zenodo.18866032.

References

Banks, J., Olson, M., & Porter, D. (1997). An experimental analysis of the bandit problem. *Economic Theory*, *10*, 55–77. https://doi.org/10.1007/s001990050146

Barbot, B. (2011). *Brief big five (BB5)*. Editions Hogrefe France.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Systems Research and Behavioral Science*, *8*(1), 41–55. https://doi.org/10.1002/bs.3830080106

Bergemann, D., & Valimaki, J. (2006). Bandit problems. Technical report, Cowles Foundation for Research in Economics, Yale University.

Binmore, K. (1999). Why experiment in economics? *The Economic Journal*, *109*(453), 16–24. https://doi.org/10.1111/1468-0297.00399

Blackburn, J. (1936). *Acquisition of skill: An analysis of learning curves*. Medical Research Council, Industrial Health Research Board, Report No. 73. London: His Majesty's Stationery Office.

Brezzi, M., & Lai, T. L. (2000). Incomplete learning from endogenous data in dynamic allocation. *Econometrica*, *68*(6), 1511–1516. https://doi.org/10.1111/1468-0262.00170

Burtini, G., Loeppky, J., & Lawrence, R. (2015). A survey of online experiment design with the stochastic multi-armed bandit. *arXiv, 1510*, 00757. https://doi.org/10.48550/arXiv.1510.00757

Camerer, C., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Caplin, A., & Dean, M. (2008). Dopamine, reward prediction error, and economics. *The Quarterly Journal of Economics*, *123*, 663–701. https://doi.org/10.1162/qjec.2008.123.2.663

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

DeGroot, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill, Inc.

Delaney, J., Jacobson, S., & Moenig, T. (2020). Preference discovery. *Experimental Economics*, *23*, 694–715. https://doi.org/10.1007/s10683-019-09628-9

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press

Gittins, J., & Jones, D. (1974). A dynamic allocation index for the sequential allocation of experiments. In J. Gani (Ed.), *Progress in Statistics* (pp. 241–266). North Holland.
Glimcher, P. W. (2009). Choice: Towards a standard back-pocket model. In P. W. Glimcher, C. F. Camerer, E. Fehr, & Poldrack R.A. (Eds.), *Neuroeconomics* (pp. 503–521). Academic Press.
Hudja, S., & Woods, D. (2023). Exploration versus exploitation: A laboratory test of the single-agent exponential bandit model. *Economic Inquiry*, *62*(1), 267–286. https://doi.org/10.1111/ecin.13164
Kahneman, D., & Thaler, R. H. (2006). Anomalies: Utility maximization and experienced utility. *The Journal of Economic Perspectives*, *20*(1), 221–234. https://doi.org/10.1257/089533006776526076
Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235
Lévy-Garboua, L., Gazel, M., Berlin, N., Dul, J., & Lubart, T. (2024). Creative cognition as a bandit problem. *Learning and Individual Differences*, *111*, 102348. https://doi.org/10.1016/j.lindif.2024.102438
Lévy-Garboua, L., & Montmarquette, C. (1996). A microeconometric study of theatre demand. *Journal of Cultural Economics*, *20*(1), 25–50. https://doi.org/10.1007/s10824-005-5050-y
Loewenstein, G., & Adler, D. (1995). A bias in the prediction of tastes. *The Economic Journal*, *105*, 929–937. https://doi.org/10.2307/2235159
Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, *118*, 1209–1248. https://doi.org/10.1162/003355303322552784
Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, *39*(3-4), 641–648. https://doi.org/10.1016/0014-2921(94)00071-7
McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351–378. https://doi.org/10.1257/aer.91.3.351
Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economics Statistics*, *3*, 370–379. https://doi.org/10.1080/07350015.1985.10509471
Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311–329. https://doi.org/10.1086/259630
Plott, C. R. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perlman, & Schmidt C., (Eds.), *Rational Foundations of Economic Behaviour* (pp. 225–250). Macmillan Press Ltd.
Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593
Scott, S. L. (2010). A modern Bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, *26*, 639–658. https://doi.org/10.1002/asmb.874
Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making in bandit problems. *Journal of Mathematical Psychology*, *53*(3), 168–179. https://doi.org/10.1016/j.jmp.2008.11.002
Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *American Economic Review*, *67*, 76–90. https://doi.org/10.3917/idee.159.0054
Train, K. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Cite this article: Lévy-Garboua, L., Owen, C., Placido, L., & Rapoport, B. (2025). Learning by consuming in the lab. *Experimental Economics*, 28(6), 1240–1261. https://doi.org/10.1017/eec.2026.10050

https://doi.org/10.1017/eec.2026.10050 Published online by Cambridge University Press