Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 155 (2015) 262–299

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Dynamic pricing in the presence of individual learning ☆

Xi Weng *

Department of Applied Economics, Guanghua School of Management, Peking University, 100871 Beijing, China

Received 7 December 2013; final version received 15 November 2014; accepted 26 November 2014

Available online 3 December 2014

Abstract

This paper studies price dynamics in a setting in which a monopolist sells a new experience good over time to many buyers, and the seller can neither price discriminate among the buyers nor commit to a price rule. Buyers learn from their own experiences about the effectiveness of the product. Individual learning generates ex post heterogeneity, which affects the buyers' purchasing decisions, the monopolist's pricing strategy, and efficiency. When learning occurs through good news signals, buyers receive a rent because of the possible advantageous belief caused by short-lived deviations. If a good news signal arrives, the price can instantaneously increase or decrease depending on the arrival time of this signal. The equilibrium is inefficient because the monopolist's incentive to exploit known buyers leads to inefficient early termination of exploration. When learning occurs through bad news signals, ex post heterogeneity has no such effect, since only homogeneous unknown buyers purchase the experience good.
© 2014 Elsevier Inc. All rights reserved.

JEL classification: D83; C02; C61; C73

Keywords: Learning; Experimentation; Strategic pricing; Exponential bandit; Good news case; Bad news case

☆ I am especially grateful to George Mailath for continuous encouragement, guidance and support. I also thank Hongbin Cai, David Dillenberger, Jan Eeckhout, Hanming Fang, Qingmin Liu, Steve Matthews, Guido Menzio, Tymofiy Mylovanov, Mallesh Pai, Andrew Postlewaite, Jeroen Swinkels and seminar participants at Kellogg, Michigan, UCL, Penn, Toulouse, University of Hong Kong and UPF for helpful comments and suggestions. I also acknowledge financial support from the National Natural Science Foundation of China (Grant No. 71303014) and Guanghua Leadership Institute (Grant No. 12-02), as well as support from the Spanish Ministry of the Economy and Competitiveness (Project ECO2012-36200) and the Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education, China.

* Fax: +86 10 6275 1463.
E-mail address: wengxi125@gsm.pku.edu.cn.

http://dx.doi.org/10.1016/j.jet.2014.11.016
0022-0531/© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In many markets for new experience goods, buyers face huge uncertainty about the effectiveness or possible side effects of the product. Take as an example the market for new drugs. The effectiveness of a new drug critically depends on whether there is an appropriate match between this drug and each patient's particular problems [16]. Patients learn from their own experiences (*individual learning*) about this effectiveness.

This paper investigates how a monopolist sells a new experience good to many buyers over time in the presence of individual learning. The monopolist and the buyers initially are equally unsure about the effectiveness of the product. How will this monopolist price strategically if she observes each buyer's past actions and outcomes? Without having seen the effectiveness of the product, potential purchasers become increasingly pessimistic. In order to keep buyers purchasing the product, the price must be reduced. How will the monopolist react when the product is revealed to be effective for one buyer? Will strategic pricing achieve efficient allocation?

In this paper, dynamic monopoly pricing is modeled as an infinite-horizon, continuous-time process. The monopolist sells a perishable experience good. She can neither price-discriminate across buyers nor commit to a price rule. At each instant of time, the monopolist first posts a spot price, which is contingent on the available public information about the experiences of the buyers. Each buyer then decides to either buy one unit of the experience good or take an outside option (modeled as another good of known characteristics). The experience good generates random lump-sum payoffs according to independent Poisson processes. The arrival rate of the lump-sum payoffs depends on an unknown individual attribute, which is binary and uncorrelated across buyers. For tractability, we assume that the public arrival of lump-sum payoffs immediately resolves the idiosyncratic uncertainty of the receiver. A key feature of the model is that buyers can become *ex post heterogeneous* in two ways: heterogeneity can be induced by either different outcomes or different actions.

We consider two different cases. In the *good news case*, the experience good generates positive lump-sum payoffs; in the *bad news case*, it generates negative lump-sum damages (e.g., side effects of new drugs). This paper fully characterizes the symmetric Markov perfect equilibrium for both cases. If a monopolist sells to a single buyer, the equilibrium price is set such that the buyer is indifferent between purchasing the experience good and taking the outside option. The buyer's purchasing decision is purely myopic since her continuation value is independent of the learning outcomes. This leads to an efficient outcome since the monopolist fully internalizes the social surplus.

We first characterize the symmetric Markov perfect equilibrium for the good news case in which there are two buyers. In phase S, i.e., when no lump-sum payoff has yet arrived, the monopolist sells to both unknown buyers before quitting the market (an "unknown" buyer refers to a buyer whose valuation of the good has not been revealed); in phase I, i.e., after one buyer has received a lump-sum payoff, the critical tradeoff is whether to sell to both buyers or to sell only to the known buyer. In both phases, the equilibrium purchasing behavior is determined by a cutoff in the posterior belief about the unknown buyer's individual attribute. Each unknown buyer makes a purchase when the posterior belief is above this cutoff and takes the outside option otherwise.

In phase I, the unknown buyer's purchasing decision is purely myopic as in the single buyer case. The key reason for this is that if the monopolist sells to both buyers, the equilibrium price is set to make the more pessimistic unknown buyer indifferent. In phase S, however, the presence of ex post heterogeneity has two important implications for the equilibrium price.

First, consider the situation where two *ex ante* identical unknown buyers make different purchasing decisions. One buyer continues experimentation by purchasing the experience good, while the other buyer deviates to take the outside option for a small amount of time. If the experimenter does not receive any lump-sum payoffs during that period, she becomes more pessimistic about her individual attribute. Without price discrimination, if the monopolist sells to two different buyers, the optimal price is set to make the more pessimistic buyer indifferent between the alternatives. The deviator, who is more optimistic about the experience good, pays less than her willingness to pay. This implies that it is always profitable for the deviator to cause asymmetric beliefs between the other buyer and herself, and the monopolist must provide extra subsidy to deter such a short-lived deviation.[1] Because of this *deterrence effect*, the equilibrium price is lowered such that each unknown buyer myopically prefers purchasing the experience good than taking the outside option.

Second, there is another positive *continuation value effect* on the price in phase S. This is driven by ex post heterogeneity in phase I. In phase I, the monopolist faces a trade-off between exploiting the buyer who is known to be good, and exploring the unknown buyer. As long as this trade-off is resolved in favor of exploration, the known buyer receives a rent by paying a relatively low price. This feeds back to phase S: the monopolist can charge a relatively high price due to this extra incentive to experiment. The combination of the deterrence effect and the continuation value effect implies that the price response to the first lump-sum payoff is ambiguous. When the first lump-sum payoff arrives relatively early, there is an instantaneous drop in the price; whereas, when the first lump-sum payoff arrives relatively late, there is an instantaneous jump in the price.

The equilibrium purchasing behavior in the good news case is characterized for an arbitrary number of buyers. It turns out that the equilibrium experimentation level is always lower than the socially efficient one when at least one buyer has received a lump-sum payoff. This is due to the existence of ex post heterogeneity: known buyers are willing to pay more than unknown buyers. Without price discrimination, the tradeoff between exploitation and exploration leads to inefficient early termination of experimentation.

We then characterize the symmetric Markov perfect equilibrium for the bad news case. It is shown that the equilibrium is always efficient. The key insight is that although buyers become heterogeneous, the buyers who have received lump-sum damages will never repurchase the experience good. The only potential buyers are the unknown ones, who are ex post homogeneous in a symmetric equilibrium. Another important difference between the good and bad news cases is that no extra subsidy is needed in the bad news case since deviations by an unknown buyer make the deviator more pessimistic. As a result, there is no deterrence effect and no continuation value effect.

1.1. Related literature

Refs. [4] and [19] are two early papers that analyze the impact of price competition on experimentation. They show that if there is only one buyer, dynamic duopoly competition with vertically differentiated products can achieve efficiency. However, [5] show that in the presence of social learning, dynamic duopoly competition cannot achieve efficiency. Refs. [6] and [12]

[1] This is similar to findings in the dynamic moral hazard model with experimentation (see, e.g., [3,9] and [20]). In these papers, it is shown that a misalignment of beliefs between the principal and the agent is always profitable for the agent, and must be dissuaded by providing more high-powered incentives. However, the mechanism is due to unobservable effort, and hence is different from the one discussed in this paper.

allow *ex ante heterogeneity* in the sense that buyers are different in their willingness to pay.[2] Both papers assume a continuum of buyers. At each instant of time, an individual buyer makes a myopic optimal choice and strategic interactions between buyers do not exist.

Ref. [7] also considers a dynamic monopoly pricing problem, but with a continuum of buyers. The difference in crucial modeling assumptions leads them to investigate different properties of the equilibrium price path. When the framework of a continuum of buyers is used, it implies that no individual buyer can affect the price path of the monopolist, and hence the analysis of deviation incentives is greatly simplified. However, it is also impossible to discuss the impact of a single good news signal on price.[3] Instead, [7] is more concerned with would follow a downward trajectory or whether it would eventually go up in equilibrium. Refs. [14] and [15] develop a different model for dynamic monopoly pricing under social learning. Their model is closer to that found in the herding literature: each short-lived buyer makes a purchasing decision following a pre-determined sequence. In contrast, in our model, all buyers are long-lived and make purchasing decisions repeatedly.

This paper is also closely connected to the continuous-time strategic experimentation literature. A nonexhaustive list of related papers includes [11,21,22] and [23].[4] The analysis of our model setting is greatly simplified by the use of exponential bandits, building on [23]. Most research in the strategic experimentation literature assumes a common value environment, in which the players' payoffs are perfectly correlated. This enables us to use a uni-dimensional posterior belief as the unique state variable to characterize the value functions. By considering individual learning, we introduce multi-dimensional posterior beliefs and develop methods to solve this dimensionality problem.

In addition to the theoretical body of work, several empirical studies attempt to quantify the importance of learning considerations on consumers' dynamic purchasing behavior. However, most existing works have exclusively focused on modeling individual consumer behavior and analyzing the impact of idiosyncratic uncertainty (see, e.g., [2,16,18], etc.). This paper complements this literature by considering how the monopolist seller dynamically adjusts the price in the presence of individual learning.

The remainder of this paper is organized as follows. Section 2 introduces the model and defines the solution concept. Section 3 and Section 4 solve a symmetric Markov perfect equilibrium and discuss the efficiency of the equilibrium for the good news case and the bad news case, respectively. Section 5 concludes the paper.

[2] Ref. [27] also investigates a duopoly model with *ex ante heterogeneity* along a location. This paper considers a two-period model and is mainly concerned about consumer loyalty, i.e., whether in the second period, buyers return to the seller that they originally purchased from.

[3] In the Web Appendix, we consider an extended good news model with a continuum of buyers. In equilibrium, the monopolist will initially decrease the price to encourage experimentation. It turns out that the deterrence effect and the continuation effect cancel each other out, and the price is set such that the unknown buyers are myopically indifferent between purchasing the experience good and taking the outside option. However, due to the option value of becoming a known buyer, each unknown buyer strictly prefers purchasing the experience good.

[4] The strategic experimentation framework is also used as a building block to investigate broader issues. For example, [26] investigates voting in a strategic experimentation environment; [28] considers a war-of-attrition game in which the players are learning about their private payoffs; [3,13,20] and [24] all investigate moral hazard problems when effort affects the speed of learning.

2. Model setting

We consider a continuous-time model with $t \in [0, +\infty)$ and a positive discount rate $r > 0$. The market consists of n buyers indexed by $i = 1, 2, \cdots, n$ and one monopolist, who are all risk-neutral. A monopolist with a zero cost of production sells a *risky product* with unknown value.[5] At each point in time, a buyer can either buy one unit of the risky product or take a safe outside option/product.

If a buyer purchases the safe product, she receives a known deterministic flow payoff $s > 0$.[6] The value of the risky product to a buyer i consists of two components: a deterministic flow payoff $\xi_f \geq 0$ and a random lump-sum payoff ξ_l. The presence of lump-sum payoffs depends on the quality of the match between the product and the specific buyer: it is either *relevant* ($\kappa_i = 1$) or *irrelevant* ($\kappa_i = 0$). The arrival of random lump-sum payoffs ξ_l is independent across buyers and modeled as a Poisson process with intensity $\lambda_H \kappa_i$, with $\lambda_H > 0$. The common priors are such that $\rho_0 = \Pr(\kappa_i = 1)$ for each buyer i. The product characteristic and the match qualities are initially unobservable to all players (seller and buyers), but the parameters λ_H, ξ_f, ξ_l, and ρ_0 are common knowledge.

We consider two cases for the above setting. In the *good news case*, $\xi_l > 0$ and the arrival of lump-sum payoffs makes the risky product more attractive than the safe one. We assume that the risky product is superior to the safe one only when the buyers can receive lump-sum payoffs:

Assumption 1 (*Good news case*). In the good news case, $\xi_l > 0$ and $\xi_f < s < \xi_f + \lambda_H \xi_l$.

In the *bad news case*, $\xi_l < 0$ and the arrival of lump-sum payoffs makes the risky product less attractive than the safe one. We impose the requirement that the risky product is superior to the safe one only when the buyers cannot receive lump-sum payoffs:

Assumption 2 (*Bad news case*). In the bad news case, $\xi_l < 0$ and $\xi_f > s > \xi_f + \lambda_H \xi_l$.

All players observe each buyer's past actions and outcomes. As a result, both the seller and the buyers hold common posterior beliefs about any given buyer's match quality. In both cases, if one buyer receives a lump-sum payoff from the risky product, every player immediately knows that the buyer's match is relevant. The absence of lump-sum payoffs makes it very likely that a match is irrelevant.

At each instant of time t, the monopolist first announces a spot price based on previous history and then each buyer decides which product to purchase conditional on previous history and the announced price. It is assumed that the monopolist can neither price-discriminate across buyers nor commit to a price rule.

2.1. Belief updating

Denote by N_{it} the total number of lump-sum payoffs received by buyer i before time t. Let P_t be the price charged by the monopolist at time t. Set $a_{it} = 1$ if buyer i purchases the risky

[5] The zero cost assumption is simply a normalization. The model with production cost $c > 0$ is equivalent to another one with zero production cost and the flow payoff of the safe product being $s' = s + c$.

[6] Alternatively, we can assume that the flow payoff is random but drawn from a commonly known distribution with expectation $s > 0$.

product at time t; and $a_{it} = 0$ if buyer i purchases the safe product at time t. Public history before time t is defined as:

$$h_t \triangleq \left(\{a_{i\tau}, N_{i\tau}\}_{i=1}^n, P_\tau\right)_{0\le\tau<t}.$$

Posterior belief that buyer i's match is relevant is defined as:

$$\rho_{it} \triangleq \Pr[\kappa_i = 1 \mid h_t].$$

Given prior ρ_0, the posteriors ρ_{it} evolve according to Bayes' rule. A buyer i who has not received any lump-sum payoff before time t expects to receive a lump-sum payoff from the risky product with arrival rate $\lambda_H a_{it}\rho_{it}$. If a lump-sum payoff is received, ρ_{it} immediately jumps to 1; otherwise, ρ_{it} obeys the following differential equation at those times t when a_{it} is right continuous[7]:

$$\dot\rho_{it} = -\lambda_H a_{it}\rho_{it}(1-\rho_{it}). \tag{1}$$

2.2. Strategies and payoffs

Throughout the paper, we focus on symmetric Markov perfect equilibria. The natural state variables are the posteriors $\boldsymbol{\rho}$. The state variable $\boldsymbol{\rho}_t$ is required to be feasible in the sense that

$$\boldsymbol{\rho}_t \in \Sigma = \left\{\boldsymbol{\rho} \in [0,1]^n : \text{ either } \rho_i = 1 \text{ or } \rho_i \le \rho_0 \text{ for all } i\right\}.$$

Purchasing decision Buyer i's acceptance policy is a function of states $\boldsymbol{\rho}$ and price P

$$\alpha_i : \Sigma \times \mathbb{R} \to \{0,1\}.$$

Since lump-sum payoffs arrive at rate $\rho_{it}\lambda_H$, the expected flow of utility associated with purchasing decision a_{it} is

$$a_{it}\rho_{it}\lambda_H\xi_l + a_{it}(\xi_f - P_t) + (1-a_{it})s.$$

The choice of a_{it} affects not only flow utility but also how beliefs $\boldsymbol{\rho}_t$ are updated. Given beliefs $\boldsymbol{\rho} \in \Sigma$, monopolist's strategy P and other buyers' strategies α_{-i}, buyer i's value (sum of normalized expected discounted utility) from purchasing strategy α_i is

$$U_i(\alpha_i, P, \alpha_{-i}; \boldsymbol{\rho}) = \mathbb{E}\int re^{-rt}\left\{\alpha_i(\boldsymbol{\rho}_t, P_t)(\rho_{it}\lambda_H\xi_l + \xi_f - P_t) + \left(1-\alpha_i(\boldsymbol{\rho}_t, P_t)\right)s\right\}dt$$

where the expectation is taken over $\{\boldsymbol{\rho}_t : t \in [0,\infty)\}$ with $\boldsymbol{\rho}_0 = \boldsymbol{\rho}$.

---

[7] If buyer i has not received good news within the time period $[t, t+h)$, then the posterior belief $\rho_{i,t+h}$ can be written as:

$$\rho_{i,t+h} = \frac{\rho_{it}e^{-\lambda_H\int_0^h a_{i,t+\tau}d\tau}}{\rho_{it}e^{-\lambda_H\int_0^h a_{i,t+\tau}d\tau} + 1 - \rho_{it}}.$$

Since $a_{i\tau}$ is right continuous with respect to time at time t, there exists some $\bar h > 0$ such that $a_{i,t+\tau} = a_{i,t}$ for all $\tau \le \bar h$. Hence by definition:

$$\dot\rho_{it} = \lim_{h\to 0}\frac{\rho_{i,t+h} - \rho_{i,t}}{h} = -\lambda_H a_{it}\rho_{it}(1-\rho_{it}).$$

Pricing decision Given prior ρ_0,

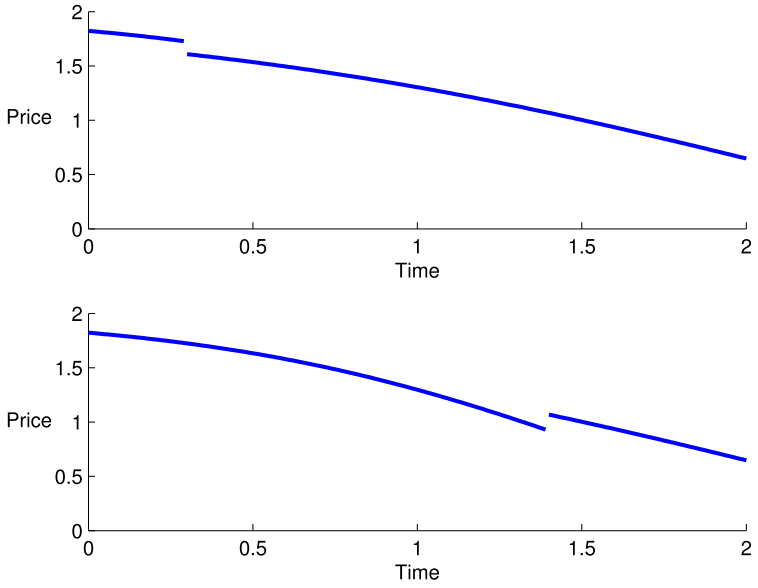

Fig. 4. Instantaneous price response to the first arrival of good news.

$$\tilde{P}_S(\rho) - P_I(\rho) = \frac{\lambda_H}{r}\rho\big(V_I(\rho) - s\big) > 0.$$

After the arrival of the first lump-sum payoff, the known buyer receives a rent by paying a relatively low price. Thus, in phase S, the monopolist can charge a relatively high price due to the extra incentive to experiment as a result of being the first to receive a lump-sum payoff gives access to an additional rent in phase I.

The combination of the deterrence effect and the continuation value effect implies that the instantaneous price reaction to the arrival of the first lump-sum payoff is ambiguous. In particular, when the first lump-sum payoff arrives early, there could be an instantaneous drop in price in order to encourage the unknown buyer to experiment as shown by Fig. 2.

Corollary 1. *There exists $\bar{\rho} > \rho_I^\star$, such that $P_S(\rho) > P_I(\rho)$ for $\rho > \bar{\rho}$ and $P_S(\rho) < P_I(\rho)$ for $\rho < \bar{\rho}$.*

The above corollary implies: in the early days of the market, ρ is higher and it is more likely to have $P_S(\rho) > P_I(\rho)$; in the late days of the market, ρ is lower and it is more likely to have $P_S(\rho) < P_I(\rho)$. Fig. 4 describes a situation in which with the same prior, the price might either drop or jump depending on when the first lump-sum payoff arrives.

3.3. Efficiency

This section discusses the efficiency property of the symmetric Markov perfect equilibrium for an arbitrary number of buyers.

Theorem 1. *Consider a market with any $n \geq 2$ buyers. The symmetric Markov perfect equilibrium is inefficient. Moreover, the equilibrium experimentation terminates too early except when no buyer has received a lump-sum payoff.*

The result of Theorem 1 is intuitive. As in Proposition 1, when more than one buyer has received a lump-sum payoff, the monopolist faces a trade-off between exploiting those buyers who are known to be good, and exploring those buyers who have not yet determined their valuation. This trade-off always leads to inefficient early termination of exploration. Moreover, as shown in the proof of Theorem 1, the equilibrium cutoff when k buyers are known to be good satisfies

$$\rho_k^\star = \frac{nrs + kr(g-s)}{nrg + (n-k)\lambda_H(g-s)},$$

which is obviously increasing in k: the monopolist is more likely to stop experimentation as the temptation to exploit current known buyers increases, and is deceasing in n: the incentive to exploration becomes higher as the number of buyers grows.

We are now in a position to summarize the roles played by ex post heterogeneity. First, in phase S, ex post heterogeneity means that there is a future benefit for the deviator by becoming more optimistic than the non-deviators. The monopolist must provide an extra subsidy to deter deviations. Second, in the individual learning phase, ex post heterogeneity implies that the receivers of lump-sum payoffs are more optimistic than the unknown buyers. If the monopolist wishes to serve all buyers, the known buyers extract rents. This continuation value effect has a positive impact on the price in phase S: the monopolist can charge a relatively high price due to this extra incentive to experiment. The interaction of deterrence and continuation value effects leads to an ambiguous instantaneous price reaction to the arrival of the first lump-sum payoff. Finally, ex post heterogeneity generates a tradeoff between exploitation and exploration for the monopolist. The equilibrium experimentation level is lower than the socially efficient level as we have seen in the two-buyer case.[11]

One important implicit assumption is that the monopolist cannot commit to a certain price rule. It is common in the dynamic monopoly pricing literature to assume lack of commitment (see, e.g., the literature on the Coase conjecture). With commitment, efficiency can be easily restored by using the following "trial" contract. The monopolist offers a contract (p, T) where T is the length of the trial period and p is the lump-sum price charged at the beginning of the game. The monopolist commits that by paying p, the consumers can purchase the product for T length of time for free. After the trial period, the monopolist charges price $g - s$ and the known buyers will purchase the product.

By setting T such that the posterior belief at T, ρ_T, is the same as $\rho^e = \frac{rs}{rg+\lambda_H(g-s)}$, the monopolist achieves expected profits from each buyer:

$$\pi = \left(1 - e^{-rT}\right)(\rho_0 g - s) + \rho_0\left(1 - e^{-\lambda_H T}\right)e^{-rT}(g-s),$$

where $(1 - e^{-rT})(\rho_0 g - s)$ is the price charged at the beginning of the game to incentivize buyers to purchase. After time T, with probability $\rho_0(1 - e^{-\lambda_H T})$, the buyer receives at least one lump-sum payoff, and will pay $g - s$ to purchase the product.

Obviously, $\rho_T = \rho^e$ implies that experimentation stops efficiently. Meanwhile, the monopolist is willing to offer this trial contract because it extracts the full surplus received by the buyers and, hence, achieves the highest possible expected profits. One attractive property of this contract is that no buyer is incentivized to deviate given that the deviation has no impact on future price.

[11] In the Web Appendix, it is shown that if the buyers' valuations are perfectly correlated instead of independent, there is no deterrence or continuation value effect. Furthermore, the equilibrium is always efficient since the monopolist can fully internalize the social surplus.

4. Equilibrium in

homogeneous. Second, off the equilibrium path, a buyer cannot benefit from deviations because the deviator becomes more pessimistic.

In the Web Appendix, we consider the correlated case by introducing another dimension of uncertainty on product quality. A lump-sum payoff can occur only when both the product quality is high and the match quality is relevant. The analysis of the correlated case is more complicated since there is one more state variable: the belief about the product characteristic. As the belief about the product characteristic can be expressed as a function of the belief about individual match quality, we can apply similar techniques to analyze the problem. It is shown that adding this common uncertainty does not change the main results as long as idiosyncratic uncertainty still exists.

There are several extensions to consider in the future. For tractability, we have assumed that the arrival of lump-sum payoffs immediately resolves the idiosyncratic uncertainty of the receiver. We could consider a model where the arrival of lump-sum payoffs cannot immediately resolve the idiosyncratic uncertainty of the receiver as in [22]. As long as ex post heterogeneity exists, the resulting equilibrium would be inefficient as well.

Another natural extension of the current model is to consider a dynamic duopoly pricing environment. This issue is partially investigated by [6], who consider a model with a continuum of buyers such that buyers choose according to their myopic preferences at each instant. It would be interesting to consider a model with a finite number of buyers such that each buyer's choice has non-trivial effects on learning and future prices.

Appendix A. Admissible strategies

Before formally defining admissible Markovian strategies, we define admissibility for general strategies. First denote an outcome h to be

$$h \triangleq \left(\{a_{it}, N_{it}\}_{i=1}^{n}, P_t\right)_{0\le t<\infty};$$

and H is the set of all possible outcomes. A sub-outcome $h^- \subset h$ only includes information about purchasing decisions and lump-sum payoffs:

$$h^- \triangleq \left(\{a_{it}, N_{it}\}_{i=1}^{n}\right)_{0\le t<\infty};$$

and H^- is the set of all possible sub-outcomes.

In general, a strategy can be viewed as a map from the set of outcomes to actions. We focus on strategies that are independent of previous prices; this is because allowing pricing as a function of previous prices may generate more complicated problems.[14] The monopolist's pricing decision is given by the mapping:

$$P : H^- \times [0,\infty) \to \mathbb{R};$$

and the buyers' acceptance decision is given by the mapping:

$$\alpha_i : H \times [0,\infty) \to \{0,1\}.$$

$P(h^-, t)$ is the price charged by the monopolist at time t; and $\alpha_i(h,t)$ is the purchasing decision made by buyer i at time t. Assumptions A1 and A2 stated below guarantee that the strategies are well defined.

[14] For example, any decreasing price path is consistent with the pricing function $P(h,t) = \inf_{\tau<t} P_\tau$.

Denote vector $a = (a_1, \cdots, a_n)$ and vector $N = (N_1, \cdots, N_n)$. A metric on the sets of outcomes is defined as:

$$D^-\left(\hat{h}_t^-, \tilde{h}_t^-\right) = \int_0^t \left[d(\hat{a}_\tau, \tilde{a}_\tau) + d(\hat{N}_\tau, \tilde{N}_\tau)\right] d\tau$$

and

$$D(\hat{h}_t, \tilde{h}_t) = \int_0^t \left[d(\hat{a}_\tau, \tilde{a}_\tau) + d(\hat{N}_\tau, \tilde{N}_\tau)\right] d\tau + |\hat{P}_t - \tilde{P}_t|$$

where d is the Euclidean norm. In particular, the previous prices do not enter in the definition of $D(\hat{h}_t, \tilde{h}_t)$; only the current price matters. The metric D (D^-) determines a Borel σ-algebra $\mathcal{B}_H$ ($\mathcal{B}_{H^-}$). The first restriction on strategies is that:

A1. P is a $\mathcal{B}_{H^-} \times \mathcal{B}_{[0,\infty)}$ measurable function and α_i is a $\mathcal{B}_H \times \mathcal{B}_{[0,\infty)}$ measurable function.

The second restriction requires that the strategies take the same actions if two histories are almost the same:

A2. For all t, and $\hat{h}, \tilde{h} \in H$ such that $D(\hat{h}_t, \tilde{h}_t) = 0$, then $P(\hat{h}^-, t) = P(\tilde{h}^-, t)$ and $\alpha_i(\hat{h}, t) = \alpha_i(\tilde{h}, t)$.

A1 and A2 are two natural restrictions on strategies. Additional conditions must be imposed to guarantee that the induced outcome is unique. Before doing that, we define an outcome h to be *compatible* with a given strategy profile $\{P, \alpha\}$ if h satisfies: $P(h^-, t) = P_t$ and $\alpha_i(h, t) = a_{it}$. A straightforward modification of the argument in [8] shows the following:

Proposition 4. *A strategy profile (P, α) generates a unique distribution over compatible outcomes if it satisfies:*

1. *for any outcomes $\hat{h}$ and $\tilde{h}$ and any time t such that $D(\hat{h}_t, \tilde{h}_t) = 0$ and $\hat{N}_t = \tilde{N}_t$,*

$$\lim_{\epsilon \searrow 0} P(\hat{h}, t + \epsilon) = \lim_{\epsilon \searrow 0} P(\tilde{h}, t + \epsilon);$$

and

2. *for any $\hat{h}$ and $\tilde{h}$ and any t such that $D(\hat{h}_t, \tilde{h}_t) = 0$, $\hat{N}_t = \tilde{N}_t$ and $\lim_{\epsilon \searrow 0} \hat{P}_{t+\epsilon} = \lim_{\epsilon \searrow 0} \tilde{P}_{t+\epsilon}$, then there exists $\epsilon > 0$ and $a \in \{0, 1\}$ such that $\alpha_i(\hat{h}, \tilde{t}) = \alpha_i(\tilde{h}, \tilde{t}) = a$ for any $\tilde{t} \in (t, t + \epsilon)$.*

We say a strategy profile (P, α) is *weakly admissible* if it satisfies conditions 1 and 2 in Proposition 4. In Proposition 4, condition 2 is the key condition. This condition is slightly different from the inertia condition proposed in [8]. The modification is needed to handle the possible situation in which the arrival of a lump-sum payoff at time t results in the purchasing decisions a_t not being right continuous in time.

Any Markovian strategy profile (P, α) which induces a weakly admissible strategy profile generates a unique distribution over compatible outcomes. However, the notion of weak admissibility does not guarantee that the induced outcome allows us to use Eq. (1) to update beliefs.

Definition 2. A Markovian strategy profile (P, α) is *strongly admissible* in the good news case if it satisfies[15]:

1. $P(\rho)$ is left continuous and non-decreasing when it is continuous: for each $\rho \in \Sigma$ and $\delta > 0$, there exists some $\epsilon > 0$ s.t. $P(\rho') \leq P(\rho)$ and $|P(\rho') - P(\rho)| \leq \delta$ for all feasible $\rho' \leq \rho$ such that $\|\rho' - \rho\| \leq \epsilon$[16];
2. $\alpha_i(\rho, P)$ is left continuous: for each $\rho \in \Sigma$ and $\delta > 0$, there exists some $\epsilon' > 0$ s.t. $\alpha_i(\rho', P') = \alpha_i(\rho, P)$ for all feasible $(\rho', P') \leq (\rho, P)$ such that $\|(\rho', P') - (\rho, P)\| \leq \epsilon'$; and
3. if h is a history compatible with (P, α), $\mathcal{C}(t; h) < \infty$ for $t < \infty$, where $\mathcal{C}(t; h)$ denotes the number of times τ before t such that purchasing behavior a_τ is discontinuous.

It is straightforward to check that conditions 1 and 2 in Definition 2 are sufficient to guarantee that (P, α) induces a weakly admissible strategy profile. In addition, these two conditions imply that any outcome induced by the Markovian strategy profile (P, α) is well behaved in the sense that the purchasing decisions a_{it} and pricing decisions P_t are right continuous functions when there is no arrival of lump-sum payoffs. This enables us to use Eq. (1) to update beliefs. In the good news case, condition 1 implies P_t is decreasing when it is continuous but we also allow jumps in the price path. Condition 3 requires that each buyer can change actions no more than a finite number of times in a finite time interval, since condition 2 does not preclude the possibility of an infinite number of changes on any time interval. This additional condition is needed to simplify the analysis of the equilibrium.

Definition 2 is too strong in the sense that even cutoff strategies may not be strongly admissible.[17] We use the completion argument in [8] to overcome this issue. First define a metric on the space of strongly admissible strategies. A Markovian strategy profile (P, α) is *admissible* if there exists strongly admissible Markovian strategy profiles $\{(P_k, \alpha_k)\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty}(P_k, \alpha_k) = (P, \alpha)$. An outcome h is *consistent* with an admissible strategy profile (P, α) if there exists strongly admissible Markovian strategy profiles $\{(P_k, \alpha_k)\}_{k=1}^{\infty}$ and outcomes $\{h_k\}_{k=1}^{\infty}$ satisfying the following three conditions: i) for each k, h_k is compatible with (P_k, α_k), ii) $\lim_{k\to\infty}(P_k, \alpha_k) = (P, \alpha)$ and iii) $\lim_{k\to\infty} h_k = h$. An admissible Markovian strategy profile (P, α) may not generate a unique distribution over compatible outcomes. However, the proof of Theorem 2 in [8] applies here, as well, to show that each admissible Markovian strategy profile (P, α) is identified with a unique distribution over consistent outcomes. When referring to outcomes generated by an admissible Markovian strategy profile (P, α), we restrict attention to the consistent outcomes.

In our definition of Markov perfect equilibrium, we allow the deviating strategies to be non-Markovian. Additional conditions on the non-Markovian strategies are also needed to make sure that the induced outcome will be well behaved even off the equilibrium path. The conditions imposed are counterparts of conditions 1–3 in Definition 2.

[15] For the bad news case, condition 1 should be changed to require that P is piecewise non-increasing.

[16] We write $(x_1, \cdots, x_n) \leq (y_1, \cdots, y_n)$ if $x_i \leq y_i$ for $i = 1, \cdots, n$, and $\|\cdot\|$ is the Euclidean norm.

[17] For example, consider a cutoff strategy such that the cutoff price for buyer i is strictly increasing in beliefs and buyer i takes the risky product at the cutoff price. This strategy violates the condition that α_i is left continuous in beliefs.

Definition 3. Define time t as a regular time for outcome h if there is no arrival of lump-sum payoffs at time t. A weakly admissible strategy profile (P, α) is *strongly admissible* in the good news case if it satisfies:

1. P is right continuous and non-increasing when continuous at any regular time: for any outcomes h and any regular time t,
$$\lim_{\epsilon \searrow 0} P(h, t+\epsilon) = P(h,t);$$
and there exists $\bar{\epsilon}_1 > 0$ such that $P(h, t+\epsilon) \leq P(h,t)$ for all $\epsilon \leq \bar{\epsilon}_1$;
2. for any h and any regular t such that P_t is right continuous and non-increasing at time t, there exists $\bar{\epsilon}_2 > 0$ and $a \in \{0, 1\}$ such that $\alpha_i(h, \tilde{t}) = \alpha_i(h,t)$ for any $\tilde{t} \in (t, t+\bar{\epsilon}_2)$; and
3. if h is a history compatible with (P, α), $\mathcal{C}(t; h) < \infty$ for $t < \infty$.

A non-Markovian strategy profile (P, α) is *admissible* if there exists strongly admissible non-Markovian strategy profiles $\{(P_k, \alpha_k)\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty}(P_k, \alpha_k) = (P, \alpha)$. For an admissible non-Markovian strategy profile (P, α), we also restrict attention to those consistent outcomes that can be similarly defined.

Appendix B. Proofs of results

B.1. Proof of Proposition 1

Proof. In phase I, denote ρ to be the common posterior belief about the unknown buyer's idiosyncratic uncertainty. Denote $P_I(\rho)$ as the price set by the monopolist for $\rho > \rho_I^\star$, where $\rho_I^\star$ is the equilibrium cutoff. Then, the value function for the unknown buyer satisfies

$$rU_I(\rho) = r\big(g\rho - P_I(\rho)\big) + \rho\lambda_H\big(s - U_I(\rho)\big) - \lambda_H\rho(1-\rho)U_I'(\rho).$$

Certainly, a profit-maximizing monopolist always sets prices $P_I(\rho) = g\rho - s$ such that $U_I(\rho) = s$. The monopolist's problem is to choose between charging a low price $g\rho - s$ in order to keep experimenting and charging a high price $g - s$ to extract the full surplus from the known buyer. Obviously, this is an optimal stopping problem with HJB equation

$$rJ_I(\rho) = \max\big\{r(g-s), 2r(g\rho - s) + \rho\lambda_H\big(2(g-s) - J_I(\rho)\big) - \lambda_H\rho(1-\rho)J_I'(\rho)\big\}. \tag{12}$$

On the RHS of Eq. (12), $g - s$ is the value if the monopolist only sells to the good buyer by charging $g - s$; if the monopolist decides to continue experimentation, she not only receives instantaneous revenue $2(g\rho - s)$ by selling to both buyers, but may also receive a future value of $2(g-s)$ if the unknown buyer receives a lump-sum payoff. From the value matching and smooth pasting conditions, it is straightforward to characterize the equilibrium cutoff as

$$\rho_I^\star = \frac{r(g+s)}{2rg + \lambda_H(g-s)}.$$

The equilibrium value function $J_I(\rho)$ could be solved as:

$$J_I(\rho) = \begin{cases} 2(g\rho - s) + (g + s - 2g\rho_I^\star)\frac{1-\rho}{1-\rho_I^\star}\left[\frac{(1-\rho)\rho_I^\star}{(1-\rho_I^\star)\rho}\right]^{r/\lambda_H} & \text{if } \rho > \rho_I^\star \\ g - s & \text{otherwise.} \end{cases}$$

The known buyer only needs to pay $P_I(\rho) = g\rho - s < g - s$ before ρ reaches $\rho_I^\star$, but has to pay $g - s$ afterwards. The value function for this buyer is given by differential equation

$$rV_I(\rho) = r\big(g(1-\rho)+s\big) + \rho\lambda_H\big(s - V_I(\rho)\big) - \lambda_H\rho(1-\rho)V_I'(\rho) \tag{13}$$

for $\rho > \rho_I^\star = \frac{r(g+s)}{2rg+\lambda_H(g-s)}$ and $V_I(\rho) = s$ for $\rho \le \rho_I^\star = \frac{r(g+s)}{2rg+\lambda_H(g-s)}$. Eq. (13) is an ordinary differential equation with boundary condition: $V_I(\rho_I^\star) = s$. This gives us the expression of $V_I(\rho)$ in the proposition. □

B.2. Characterize $\lim_{h\to 0}\frac{U_S(\rho)-\hat{U}(\rho;h)}{h}$

Lemma 1. *Fix prior ρ_0 and let $\rho_S^\star$ be the equilibrium cutoff in phase S. For $\rho_S^\star < \rho \le \rho_I^\star$,*

$$\begin{aligned}\lim_{h\to 0}\frac{U_S(\rho)-\hat{U}(\rho;h)}{h} &= 2(r+\lambda_H\rho)\big(U_S(\rho)-s\big) + \lambda_H\rho(1-\rho)U_S'(\rho)\\ &\quad - \frac{rg}{r+\lambda_H}\lambda_H\rho(1-\rho)\\ &\quad + \frac{r\lambda_H g}{r+\lambda_H}\frac{\rho_S^\star(1-\rho)^2}{1-\rho_S^\star}\left(\frac{(1-\rho)\rho_S^\star}{\rho(1-\rho_S^\star)}\right)^{r/\lambda_H};\end{aligned} \tag{14}$$

and for $\rho > \rho_I^\star$,

$$\begin{aligned}\lim_{h\to 0}\frac{U_S(\rho)-\hat{U}(\rho;h)}{h} &= 2(r+\lambda_H\rho)\big(U_S(\rho)-s\big) + \lambda_H\rho(1-\rho)U_S'(\rho)\\ &\quad + (r+\lambda_H\rho)g(1-\rho)\left(\frac{(1-\rho)\rho_I^\star}{\rho(1-\rho_I^\star)}\right)^{r/\lambda_H} - \lambda_H g\rho(1-\rho)\\ &\quad - r\left[\frac{r+\lambda_H+\lambda_H\rho_I^\star}{(r+\lambda_H)(1-\rho_I^\star)}\left(\frac{\rho_I^\star}{1-\rho_I^\star}\right)^{r/\lambda_H}\right.\\ &\quad \left. - \frac{\lambda_H}{r+\lambda_H}\left(\frac{\rho_S^\star}{1-\rho_S^\star}\right)^{1+r/\lambda_H}\right]g(1-\rho)^2\left(\frac{1-\rho}{\rho}\right)^{r/\lambda_H}.\end{aligned} \tag{15}$$

Proof. First notice that if $\lim_{h\to 0}\frac{U_S(\rho)-U^D(\rho,\rho_h)}{h}$ exists, $\lim_{h\to 0}\frac{U_S(\rho)-\hat{U}(\rho;h)}{h}$ can be written as:

$$\lim_{h\to 0}\frac{U_S(\rho)-\hat{U}(\rho;h)}{h} = (r+\lambda_H\rho)\big(U_S(\rho)-s\big) + \lim_{h\to 0}\frac{U_S(\rho)-U^D(\rho,\rho_h)}{h}. \tag{16}$$

The main issue is to evaluate $U^D(\rho,\rho_h)$ for $\rho > \rho_h$. We proceed in the following three steps: first, we decompose U^D as the sum of the non-deviator's value U^{ND} and the difference $Z = U^D - U^{ND}$; second, we derive an ordinary differential equation about U^{ND}, and explicitly solve it; finally, we derive an ordinary differential equation about Z with respect to time t, and explicitly solve it.

1. **Decompose off-equilibrium value function.**
Fix $h > 0$ to be sufficiently small and the monopolist will still sell to both buyers after an h-deviation.[18] Therefore, there exists $\bar{h}'$ such that for all $h' \leq \bar{h}'$, we have:

$$\begin{aligned} U^D(\rho, \rho_h) = \mathbb{E} \int_{t=0}^{h'} re^{-rt}(\rho_t g - \tilde{P}_t)dt \\ + \rho\big(1 - e^{-\lambda_H h'}\big)e^{-rh'} V_I(\rho_{h+h'}) + \rho_h\big(1 - e^{-\lambda_H h'}\big)e^{-rh'} s \\ + \big[1 - \rho\big(1 - e^{-\lambda_H h'}\big) - \rho_h\big(1 - e^{-\lambda_H h'}\big)\big]e^{-rh'} U(\rho_{h'}, \rho_{h+h'}). \end{aligned} \tag{17}$$

In the above expression, ρ_t is the posterior about the deviator's match quality, and starts from ρ; and $\tilde{P}_t$ is the off-equilibrium price set by the monopolist after an h-deviation.
By purchasing the risky product, the non-deviator gets value

$$\begin{aligned} U^{ND}(\rho, \rho_h) = \mathbb{E} \int_{t=0}^{h'} re^{-rt}\big(\rho'_t g - \tilde{P}_t\big)dt \\ + \rho\big(1 - e^{-\lambda_H h'}\big)e^{-rh'} s + \rho_h\big(1 - e^{-\lambda_H h'}\big)e^{-rh'} V_I(\rho_{h'}) \\ + \big[1 - \rho\big(1 - e^{-\lambda_H h'}\big) - \rho_h\big(1 - e^{-\lambda_H h'}\big)\big]e^{-rh'} U(\rho_{h+h'}, \rho_{h'}), \end{aligned} \tag{18}$$

where ρ'_t is the posterior about the non-deviator's match quality, and starts from ρ_h.
Obviously, the off-equilibrium value function $U^D(\rho, \rho_h)$ can be decomposed as

$$U^D(\rho, \rho_h) = U^{ND}(\rho, \rho_h) + Z(\rho, \rho_h)$$

where $Z(\rho, \rho_h) = U^D(\rho, \rho_h) - U^{ND}(\rho, \rho_h)$.
The fact that the ρ_h buyer purchases the risky product means that it is not profitable for her to have "one-shot" deviations:

$$\begin{aligned} U^{ND}(\rho, \rho_h) \geq \tilde{U}\big(h'\big) = \int_{t=0}^{h'} re^{-rt} s dt + \rho\big(1 - e^{-\lambda_H h'}\big)e^{-rh'} s \\ + \big[1 - \rho\big(1 - e^{-\lambda_H h'}\big)\big]e^{-rh'} U(\rho_h, \rho_{h'}). \end{aligned} \tag{19}$$

Since the ρ_h buyer is more pessimistic about the probability of receiving lump-sum payoffs, the optimal off-equilibrium price $\tilde{P}$ is set such that the ρ_h buyer has incentives to experiment. Denote $\tilde{U}(\rho; \rho_h)$ as $U^{ND}(\rho, \rho_h)$ for a fixed ρ_h since ρ_h does not change in the expression of $\tilde{U}(h')$. The fact that

$$\begin{aligned} \lim_{h' \to 0} \frac{U^{ND}(\rho, \rho_h) - \tilde{U}(h')}{h'} = (r + \lambda_H \rho)\tilde{U}(\rho; \rho_h) - (r + \lambda_H \rho)s \\ + \lambda_H \rho(1 - \rho)\tilde{U}'(\rho; \rho_h) \end{aligned}$$

[18] If the monopolist only sells to the deviator, the loss from not selling to the non-deviator is proportional to the equilibrium value $J_S(\rho_h) > 0$, but the gain is proportional to $\rho - \rho_h$. As h goes to zero, the loss always dominates the gain.

is left-continuous in ρ and ρ_h implies that in equilibrium, the following equation is satisfied[19]:

$$\lim_{h' \to 0} \frac{U^{ND}(\rho, \rho_h) - \tilde{U}(h')}{h'} = 0.$$

Thus we derive an ordinary differential equation for $\tilde{U}(\rho; \rho_h)$

$$(r + \lambda$$

[6] D. Bergemann, J. Välimäki, Entry and vertical differentiation, J. Econ. Theory 106 (1) (2002) 91–125.
[7] D. Bergemann, J. Välimäki, Dynamic pricing of new experience goods, J. Polit. Economy 114 (4) (2006) 713–743.
[8] J. Bergin, W. MacLeod, Continuous time repeated games, Int. Econ. Rev. 34 (1) (1993) 21–37.
[9] V. Bhaskar, Dynamic moral hazard, learning and belief manipulation, mimeo, 2012.
[10] B. Biais, T. Mariotti, J.-C. Rochet, S. Villeneuve, Large risks, limited liability, and dynamic moral hazard, Econometrica 78 (1) (2010) 73–118.
[11] P. Bolton, C. Harris, Strategic experimentation, Econometrica 67 (2) (1999) 349–374.
[12] A. Bonatti, Menu pricing and learning, AmeAm. Econ. J. Microecon. 3 (3) (2011) 124–163.
[13] A. Bonatti, J. Hörner, Collaborating, Am. Econ. Rev. 101 (2) (2011) 632–663.
[14] S. Bose, G. Orosel, M. Ottaviani, L. Versterlund, Dynamic monopoly pricing and herding, RAND J. Econ. 37 (4) (2006) 910–928.
[15] S. Bose, G. Orosel, M. Ottaviani, L. Versterlund, Monopoly pricing in the binary herding model, Econ. Theory 37 (2) (2008) 203–241.
[16] G. Crawford, M. Shum, Uncertainty and learning in pharmaceutical demand, Econometrica 73 (4) (2005) 1137–1173.
[17] A. Dixit, The Art of Smooth Pasting, Routledge, New York, 1993.
[18] T. Erdem, M. Keane, Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets, Marketing Sci. 15 (1) (1996) 1–20.
[19] L. Felli, C. Harris, Learning, wage dynamics and firm-specific human capital, J. Polit. Economy 104 (4) (1996) 838–868.
[20] J. Hörner, L. Samuelson, Incentives for experimenting agents, Yale mimeo, 2013.
[21] G. Keller, S. Rady, Optimal experimentation in a changing environment, Rev. Econ. Stud. 66 (3) (1999) 475–507.
[22] G. Keller, S. Rady, Strategic experimentation with Poisson bandits, Theoretical Econ. 5 (2) (2010) 275–311.
[23] G. Keller, S. Rady, M. Cripps, Strategic experimentation with exponential bandits, Econometrica 73 (1) (2005) 39–68.
[24] J. Li, N. Matouschek, M. Powell, X. Weng, Implementing change in organizations, mimeo, 2013.
[25] L. Simon, M. Stinchcombe, Extensive form games in continuous time: pure strategies, Econometrica 57 (5) (1989) 1171–1214.
[26] B. Strulovici, Learning while voting: determinants of collective experimentation, Econometrica 78 (3) (2010) 933–971.
[27] J. Villas-Boas, Consumer learning, brand loyalty, and competition, Marketing Sci. 23 (1) (2004) 134–145.
[28] X. Weng, Can learning cause shorter delays in reaching agreements? mimeo, 2013.