UNIVERSITY OF HOHENHEIM

Hohenheim Discussion Papers in Business, Economics and Social Sciences

PRICE DISCRIMINATION WITH INEQUITY-AVERSE CONSUMERS: A REINFORCEMENT LEARNING APPROACH

Katrin Buchali
University of Hohenheim

Institute of Economics

02-2021

Discussion Paper 02-2021

Price Discrimination with Inequity-Averse Consumers: A Reinforcement Learning Approach

Katrin Buchali

ISSN 2364-2084

Price Discrimination with Inequity-Averse Consumers: A Reinforcement Learning Approach

Katrin Buchali[a]

[a]University of Hohenheim

Institute of Economics

katrin.buchali@uni-hohenheim.de.

Abstract

With the advent of big data, unique opportunities arise for data collection and analysis and thus for personalized pricing. We simulate a self-learning algorithm setting personalized prices based on additional information about consumer sensitivities in order to analyze market outcomes for consumers who have a preference for fair, equitable outcomes. For this purpose, we compare a situation that does not consider fairness to a situation in which we allow for inequity-averse consumers. We show that the algorithm learns to charge different, revenue-maximizing prices and simultaneously increase fairness in terms of a more homogeneous distribution of prices.

JEL-Codes: D63, D91, L12

Keywords: pricing algorithm, reinforcement learning, Q-learning, price discrimination, fairness, inequity

1 Introduction

Determining the optimal price of a product or service for a particular customer is complex: it requires knowledge of the customer's willingness to pay, estimation of demands, ability to adjust strategies to competition pricing, and more. With the advent of big data, unique opportunities arise for data collection and analysis and thus for personalized pricing. This paper discusses the capacity of reinforcement learning – a type of machine learning in which agents learn from interacting autonomously with their environment – to price discriminate based on a simulation of a price-setting algorithm.[1]

In the context of the increasing use of big data and the use of price-setting algorithms, concerns have been raised, besides the discussion about collusive behavior[2], that self-learning algorithms might engage in price discrimination[3], as they are able to use information about customers to segment them into ever-smaller groups based on certain characteristics related to their willingness to pay. Reports about customers who had to pay different prices for the same product or service have been published in the *Wall Street Journal*, the *Washington Post*, and the German magazine *Wirtschaftswoche*.[4] Consumers tend to be very sensitive to such attempts, as they have a strong sense of fairness.[5] This might explain why price discrimination is still relatively rare in economic reality.[6]

This begs the question of whether it is possible for price-setting algorithms to learn to engage in price discrimination on a basis of fairness, in order to avoid upsetting customers but still maximize expected revenues by charging personalized prices. For this reason, we will not only discuss the possibilities of algorithms to conduct price discrimination but also the possibility of self-learning pricing algorithms to consider customers' fairness preferences while setting prices. If algorithms are able to consider inequity aversion while setting differential prices, price discrimination is more likely to occur.[7] However, the effect of personalized pricing on consumer welfare is ambiguous. In many cases, personalized pricing can be beneficial: for example, the ability to offer targeted discounts might help new entrants to compete (especially in markets with switching costs) and could expand output. On the other hand, in some situations, personalized pricing can lead to consumer harm.[8]

Effective price discrimination relies on the identification of different customer groups. In

[1] Simulations are programmed in Python.

[2] For more details, see e.g. Ezrachi and Stucke (2015), Ezrachi and Stucke (2016), and Woodcock (2017).

[3] See e.g. Reinartz (2002) and UK Office of Fair Trading (2013).

[4] On Orbitz, Mac Users Steered to Pricier Hotels, in: *Wall Street Journal*, August 23, 2012; On the Web, Price Tags Blur, *Washington Post*, September 27, 2000; Der Preis ist heiß, *Wirtschaftswoche*, March 02, 2017.

[5] Kahneman et al. 1986.

[6] As stated, for example, by Odlyzko (2009), the Executive Office of the President of the United States (2015), UK Office of Fair Trading (2013), and the UK Competition and Market Authority (2018).

[7] This is also in line with growing public policy concerns; see, for example, Bourreau and Streel (2018), the UK Competition and Market Authority (2018), and the Executive Office of the President of the United States (2015).

[8] According to the report of the UK Competition and Market Authority (2018), this is the case if there is a lack of competition in the market; if discrimination is particularly complex or opaque to consumers; or if consumers loose trust in the market and, as a result, withdraw their demand or decline to participate in the market.

each group, certain assumptions about price sensitivities must be made. For instance, groups could be defined by customers' location, communication channel, click behaviors, or time spent on a certain website. The price sensitivities within each group should be taken into account by the algorithm in order to maximize expected revenues. In the two simulated scenarios studied here, we allow for differential pricing. We simulate an algorithm setting personalized prices that does and does not consider customers' notions of fairness. For simplification, fairness is maximized if everyone pays the same price. This equality-based fairness approach is widely used in the relevant literature, such as by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000). Under this assumption, the learning procedure of the algorithm should provide homogeneous prices among customer groups while simultaneously taking into account the price sensitivities within each group in order to maximize the expected revenue.[9] Compared to the situation where fairness is not considered, an improvement in fairness is seen in the situation where inequity-averse consumers are considered, and the algorithm maintains the goal of maximizing revenue.

The paper is organized as follows. Section 2 provides a short introduction into the main concepts of price discrimination, fairness, and reinforcement learning. The methodology is then presented in section 3, where the different groups of customers are introduced and reinforcement learning is applied to the differential pricing problem. The simulation results are presented and discussed in section 4. Finally, an outlook on further research is provided in section 5.

2 Basic Concepts

2.1 Price Discrimination

In general, a firm engages in price discrimination when it charges different prices for two units of the same or similar products, wherein the price difference does not reflect any cost difference.[10] Classic economic literature distinguishes between three types of price discrimination.[11] *First-degree price discrimination* occurs when the seller charges each customer the maximum price that he or she is willing to pay.[12] Even if firms do not have enough information to assess each consumer's reservation price, they can still conduct imperfect price discrimination, which is known as *third-degree price discrimination* or

[9] The algorithm has only a limited number of periods to learn the setting of the optimal price, as customers are otherwise assumed to reject buying the product.

[10] Stigler 1987.

[11] Pigou 1920.

[12] Varian 1989.

group pricing. In this practice, sellers segment their customers into broad categories according to observable characteristics; these categories are charged different prices. This is probably the most common form of price discrimination.[13]

However, if firms are able to use information technologies to collect and process a large amount of data, they can improve their knowledge of consumers' preferences. As a result, they might be able to refine the group segmentation, coming close to the ideal situation in which each group consists of a single consumer. Thus, first-degree price discrimination can be seen as an extreme form of group pricing.

First- and third-degree price discrimination rely on the existence of observable and verifiable indicators of consumers' willingness to pay. When it is not possible to identify customer groups with similar levels of willingness to pay, the only opportunity for offering different prices to different consumers is to propose to all consumers the same menu of packages (i.e., some combination of price and product characteristics), among which consumers self-select. This practice is known as *second-degree price discrimination*, nonlinear pricing, menu pricing, or versioning.[14]

Successful price discrimination requires the discriminating firm to have at least a small measure of market power and an ability to segment their customers according to their different price elasticities of demand for goods or services.[15] In addition, the firm must either be able to prevent arbitrage or to exclude arbitrage due to certain product characteristics (e.g. if the product or service has to be consumed directly). If these conditions are fulfilled, the firm is able to increase its revenue by using discriminatory pricing strategies. It should also be mentioned that a unit price can be a special case of price discrimination if every customer has the exact same willingness to pay. Additionally, it is not possible to charge every consumer a different price, as there is a smallest monetary unit and therefore only a limited number of possible prices; therefore, there will always be some consumers who pay the same price as each other.

In this paper, we focus on third-degree price discrimination. We define different customer groups, wherein each group has different sensitivities – i.e., different acceptance probabilities for the price bids of the firm. With the advent of big data, self-learning algorithms are able to differentiate and segment customers into ever-smaller reference groups who have similar price sensitivities and purchase behaviors and who share common biases and levels of willpower. Pricing algorithms are able to use data on how other people in an individual's group react in order to predict the individual's reaction under similar circumstances.

[13] Varian 1989.

[14] Ibid.

[15] Federal Trade Commission and U.S. Department of Justice, *Horizontal Merger Guidelines*.

This method enables the algorithm to adjust prices for products and services according to the estimated willingness to pay; consequently, the more data the algorithm obtains, the closer those personalized prices can be set to a consumer's reservation price.[16]

2.2 Fairness

The fairness concerns of consumers have been a discussion point in economics at least since the works of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). Many experiments have shown that economic agents do not just care about their own payoffs, but rather also about the payoffs of others, and they act accordingly. Fehr and Schmidt (1999) modeled fairness as inequity aversion, which means that people resist inequitable outcomes and are willing to give up some material payoff to move in the direction of more equitable outcomes. Thus, fairness or unfairness with regard to prices is evaluated by a comparison to prices offered to other customers under similar circumstances.[17] A way to measure the fairness of a given distribution is provided by Jain et al. (1984), who take into account consumers' preference for equality, which drives similarly distributed – i.e., homogeneous—prices. The proposed fairness index is:

$$f(x) = \frac{\left[\sum x_i\right]^2}{n \sum x_i^2} \in [0, 1]\,, \tag{1}$$

where f measures the fairness if resources are allocated to n individuals such that the i^{th} individual receives an allocation x_i.[18] For the purpose of this paper, n is interpreted as the number of customer groups, which are labeled by i. The variable x_i can represent either the price that is charged or the quantity that is sold to a particular customer group i. The index measures the equality of the allocation x: if all consumers get the same amount or pay the same price – i.e., all x_i's are equal – then the index is 1, and the policy is 100% fair. If disparity increases, fairness decreases; a policy which favors only a selected few consumers has a fairness index near 0. The index is dimensionless and independent of scale; it is bounded between 0 and 1; and it is continuous, so that any slight change in x_i changes the index.

[16] Ezrachi and Stucke 2016.

[17] The approach of Fehr and Schmidt (1999) also includes the aspect of self-centered inequity aversion, which means that consumers suffer more from inequality that is to their disadvantage (they have to pay a higher price) than from inequality that is to their advantage (they have to pay a lower price). However, for the sake of simplicity, we only consider a "general" inequity aversion.

[18] Jain et al. 1984.

2.3 Reinforcement Learning

As this study examines price discrimination and fairness through self-learning algorithms, we introduce *reinforcement learning* as a basic machine learning concept. Reinforcement learning is learning what to do, or how to map situations to actions so as to maximize a numerical reward signal. The learning agent is not told which actions to take but instead must discover which actions yield the most reward through trial. Reinforcement learning uses the formal framework of Markov decision processes to define the interaction between a learning agent and its environment in terms of states, actions, and rewards.[19]

One of the challenges that arises in reinforcement learning is the trade-off between *exploration* and *exploitation*. To obtain a high reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in producing reward. However, to discover such actions, it has to try actions that it has not tried before. Therefore, the algorithm should use a dynamic action selection policy that balances exploitation (or choosing the optimal action as currently perceived) and exploration (or choosing another action to improve action selections in the future.[20]

Independent *Q-learning*[21] is a simple but well-established reinforcement learning algorithm. Interacting with its environment, the algorithm learns according to a so-called Q-function $Q(s, a)$ that matches the optimal long-run value of choosing any action $a \in \mathcal{A}$ facing any given state $s \in \mathcal{S}$. During this interaction, the algorithm uses a dynamic action selection policy that balances exploitation and exploration. The Q-function can be represented as an $|\mathcal{S}| \times |\mathcal{A}|$ matrix. If the Q-matrix is known, the algorithm can then easily calculate the optimal action for any given state. However, Q-learning is a method for estimating the Q-matrix without knowing the underlying model through an iterative procedure. Starting from an arbitrary initial matrix Q_0, the algorithm chooses action a_t in state s_t, observes r_t and s_{t+1}, and updates the corresponding cell of the matrix $Q(s_t, a_t)$ according to the following recursive relationship:

$$Q(s_t, a_t) \leftarrow (1 - \alpha) Q(s_t, a_t) + \alpha \left(r_t + \delta \max_a Q(s_{t+1}, a) \right), \tag{2}$$

where the updated value $Q(s_t, a_t)$ is a convex combination of the previous value $Q(s_t, a_t)$ and the reward obtained after performing action a_t in state s_t plus the discounted value of the state that is reached in the following period. For all other cells of the matrix, the Q-value does not change. The parameter $\alpha \in (0, 1)$ is the learning rate and $\delta \in [0, 1)$ is the discount factor. Action a denotes the optimal strategy (i.e., the action leading to the

[19] For a comprehensive introduction to single-agent reinforcement learning, see Sutton and Barto (2018).

[20] Ibid.

[21] Watkins and Dayan 1992.

highest reward) until this time step.

To balance exploration and exploitation, the Q-learning algorithm adopts a probabilistic action selection policy. Using what is called a ϵ-greedy strategy, the algorithm follows a random action (exploration) within a given interval $[a_{min}, a_{max}]$ with $\epsilon(t) \in [0,1]$ probability and exploitative action with $1-\epsilon(t)$ probability. In the case of multiple actions sharing the same highest Q-value under exploitation, the algorithm randomizes across these actions.[22]

$$a_t = \begin{cases} [a_{min}, a_{max}] & \text{with probability } \epsilon(t) \\ \arg\max_a Q(s_t, a) & \text{with probability } 1-\epsilon(t) \end{cases} \tag{3}$$

The probability of exploration is determined by:

$$\epsilon(t) = \epsilon(0)(1-\theta)^t, \tag{4}$$

where $\epsilon(0) \in [0,1]$ is the initial exploration probability and $\theta \in [0,1]$ is a decay parameter. Whenever $\theta > 0$, the decay in exploration ensures convergence to a deterministic strategy.

Although other more complex independent-learning algorithms (such as deep learning) are currently being applied to diverse strategic settings, this paper focuses on a simple Q-learning algorithm. In contrast to more sophisticated algorithms, this simple algorithm can be fully described by just two parameters: the learning rate α and the experimentation parameter θ.[23] A disadvantage of this algorithm is the slowness of learning, which increases with the complexity of the environment. However, since the environment in this paper is very simple, Q-learning provides an adequate approach for this problem.

3 Methodology

Customer groups g_i with $i = 1, 2, \ldots, n$ are defined in advance. This can be interpreted as an algorithm that divides customers into groups based on collected and analyzed data. The characteristics used to classify customers must have an effect on their willingness to pay. For the sake of simplicity, customer groups in this paper are exogenous. Therefore, the processes of defining customer groups, allocating customers, and setting prices are

[22] Approach based on Klein (2018).

[23] For example, with deep learning algorithms, in addition to the learning and experimentation parameters, one must specify a functional form for the Q-function, the number of estimation layers, and the structure of the neural network in each layer.

sequential.

Furthermore, we assume that customers are equally uniformly distributed in the interval $[0, 1]$; thus, we assume a unit mass of consumers and unit demand. We start by defining only two customer groups $i = 1, 2$ which can be expanded in later simulations. Every customer belongs to exactly one group (either g_1 or g_2), and each customer responds to each price bid with the following probability function $\phi \in [0, 1]$:

$$\phi(a) = \left[1 + e^{-(s_i + w_i \cdot a)}\right]^{-1}, \tag{5}$$

where a is the action chosen by the algorithm (the price bid), and s_i and w_i are parameters defining the sensitivity of each group, defined in Table 1.[24]

Group	s	w
1	18.229	-23.69
2	4.4757	-15.526

Table 1: Weights for the probability function ϕ

The larger the parameter s is, the higher the acceptance probability of the customer group. For $s = 100$, the acceptance probability equals 1.0 for any given price. For $s < 10$, the acceptance probability for a price of zero is < 1. Parameter w determines how strong and in which direction (positive or negative) the consumer reacts to a change in the price. If $w > 0$, the customers' acceptance rate increases with the price. This happens when consumers judge quality by price. If $w < 0$, the acceptance rate decreases when the price increases.

We provide a scenario simulating two different types of consumer behavior. Customers in g_1 accept much higher prices (also in terms of the initial acceptance rate) than customers in g_2. Both groups show a decrease in the acceptance probability if the price increases. The acceptance probability functions of the two customer groups for the price interval $[0, 1]$ are shown in Figure 1.

The action space is defined as an interval for the action bid $a \in [a_{min}, a_{max}]$ – i.e., the price. Since Q-learning requires a finite action space, the action bid is a discrete variable scaled between 0 and 1 with k equally-sized intervals; thus, actions are taken from a discrete set $\mathcal{A} = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, 1\}$.

The state space of group i is given by all prices that can be set for group j and vice versa. Consequently, the state space is given by $\mathcal{S} = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, 1\}$.

[24] Similar functions were used by Maestre et al. (2019), who dealt with a more complex reinforcement learning approach in a dynamic pricing context.

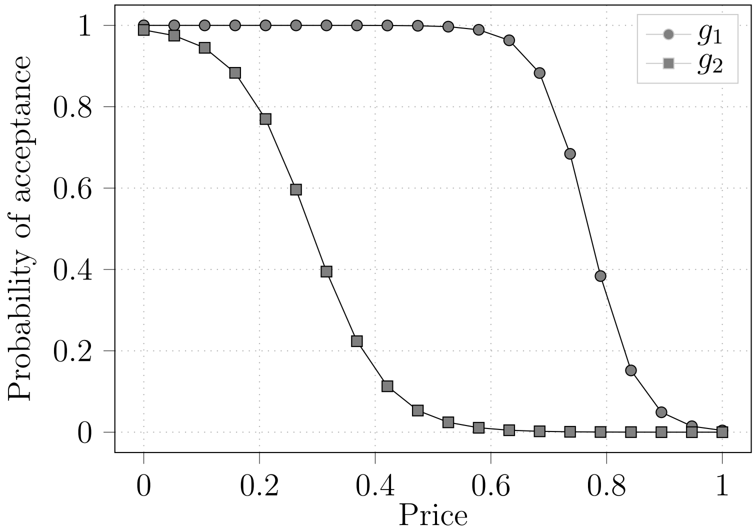

Figure 1: Acceptance probability functions of customer groups g_1 and g_2

Whether a price distribution is considered fair is measured by the following fairness index:

$$f(a) = \frac{\left[\sum_i a_{max} - a_i\right]^2}{n \sum_i (a_{max} - a_i)^2}, \tag{6}$$

where a_{max} is the maximum price that can be set (i.e., the maximum price of the given action set) and a_i is the price set for group g_i. This index is based on the index of Jain et al. (1984) shown in equation (1). Using the original index, a large and homogeneous value of resources x increases the index and thus the fairness. However, in a differential pricing context, large prices are not perceived as fair. With such an index modification, lower, homogeneously-distributed prices lead to a higher fairness index. In order to consider the consumers' notions of fairness, the fairness parameter is integrated into the acceptance probability function such that a low fairness index decreases the acceptance probability and vice versa. The modified acceptance probability is given as:

$$\phi(a) = \left[1 + e^{-(s_i + w_i \cdot a) + \beta(1 - f(a))}\right]^{-1}, \tag{7}$$

where β denotes a parameter for considering fairness ($\beta = 1$) or not considering fairness ($\beta = 0$). The modified acceptance probabilities are shown in Figure 2 for different values of the fairness parameter f.[25]

As customers are equally uniformly distributed within each group g_i, the reward – i.e., the expected revenue – is defined as the price multiplied by the acceptance probability

[25] Customers' aversion to higher prices is captured in both the acceptance probability function (acceptance probability decreases with a higher price) and the fairness index (f is larger for lower, homogeneously-distributed prices). Nevertheless, to see how fairness develops over time, it is necessary to capture this feature as a separate consideration of the fairness index.

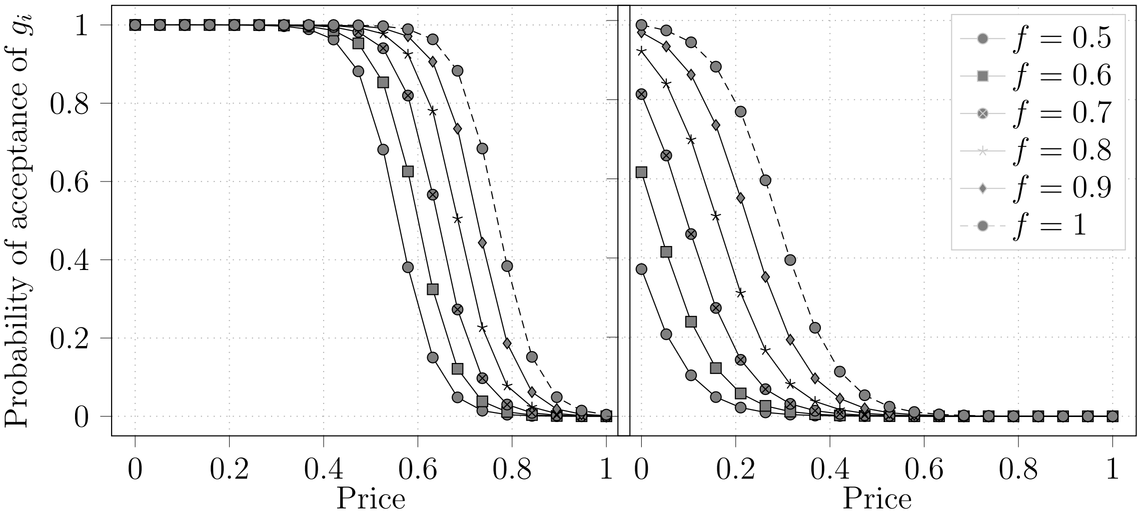

Figure 2: Acceptance probability functions of customer groups g_1 and g_2 for different fairness levels

from equation (7), where $i = 1, 2$ denotes the respective customer group.

$$r_i = a_i \cdot \phi_i \tag{8}$$

The Q-learning is simulated as follows. In $t = 1$, the algorithm sets a price for each group. For all subsequent periods, the algorithm sets prices alternately, wherein the last price set for one group serves as the current state for the other group. Additionally, this price serves as the comparative price for the fairness index. Depending on the selected parameter values, the reward is either independent of the state ($\beta = 0$, if no fairness is considered) or dependent on the current state ($\beta = 1$, if fairness is considered). The algorithm learns to set the optimal price for each group by updating its Q-function for customer group i according to the following recursive relationship:

$$Q_i(a_j, a_i) \leftarrow (1 - \alpha)Q_i(a_j, a_i) + \alpha \left(r_i + \delta r'_j + \delta^2 \max_a Q_i(a'_j, a) \right) \tag{9}$$

where action a_i denotes the price set for group i and a_j denotes the price set for group j, which determines the state for the Q-function of customer group i. The updated value $Q_i(a_j, a_i)$ is a combination of the previous value $Q_i(a_j, a_i)$, the reward obtained for group i, the discounted reward r'_j for group j that will be achieved in the following period, and the discounted value of the new state a'_j that will be reached in the following period.

The pseudocode for the Q-learning algorithm is provided below.

Pseudocode Q-learning
Set probability function ϕ, parameters s, w, β, k, α, δ, $\epsilon(0)$ and θ for group g_i
Initialize Q_1 and Q_2 as empty matrices
Initialize $a_{1,1}$ and $a_{2,1}$ randomly
Initialize $t = 2$, $i = 1$ and $j = 2$
Loop over each period
| Set action bid $a_{i,t}$ according to (3)
| Calculate reward according to (8)
| Update $Q_i(a_j, a_i)$ according to (9)
| Update $t \leftarrow t + 1$
Until $t = T$ (specified number of periods)

4 Results and Discussion

To constrain the number of possible actions, we use the discrete action set $\mathcal{A} = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, 1\}$ with $k = 10$. Consequently, there are a total of eleven possible price bids. The action set might be extended in further simulations.[26] We consider two customer groups, making $i = 1, 2$. We take an initial exploration probability $\epsilon(0) = 1$, decay parameter $\theta = 0.000046$, learning rate $\alpha = 0.1$, and discount factor $\delta = 0.9$. To assess the performance of the algorithm, the statistics are computed over 100 runs. In each run, 100000 price bids (periods) are simulated. Over these 100000 periods, the probability of exploration drops to below 1%. For each period, we average over the 100 simulated runs to see how an average market price and reward develop over time.

Two different scenarios are conducted in order to compare the outcomes of different objectives. In the first scenario, the algorithm learns to set different prices for each customer group. Fairness is not considered in this case, so the parameter β is set equal to zero. Therefore, the acceptance probability simplifies to equation (5). The reward is determined as $r_i = a_i \cdot \phi_i$. Since the fairness index has been dropped, the reward of a chosen action does not depend on the current state (i.e., the price of the other group).[27] The corresponding action and reward history are shown in Appendix A. In scenario II, fairness considerations are included ($\beta = 1$); thus, prices a_i and a_j depend on each other, due to the fairness index. Customers are assumed to be inequity averse – i.e., they have a strong preference for equitable, homogeneous prices. As a result, the acceptance probability decreases with a decreasing value of the fairness parameter. The

[26] A large action set negatively affects learning convergences; consequently, the algorithm would take much longer to learn the optimal pricing strategy.

[27] This simplified setting implies a one-stationary problem, which could be solved more efficiently by a multi-armed bandit algorithm. However, for a better comparison, the Q-learning algorithm is also applied in this case.

reward is calculated the same way as in scenario I, with the only difference being that the acceptance probability ϕ now depends on the fairness parameter f. Consequently, the algorithm learns to increase the price within each group (maximizing the expected revenue, as seen in the previous scenario) while reducing price differences between groups, in order to achieve both maximum revenue and fairness. It is assumed that maintaining a balance of price distribution among groups of customers leads to fairer prices, as these prices are chosen by considering customers' preference for equality. The corresponding action and reward history are shown in Appendix B; the average prices of the two scenarios are displayed in Figures 3 and 4.

The average market outcomes of both settings of the final 1000 periods are presented in Table 2. In the first scenario, the Q-learning agent is able to conduct price discrimination and does not consider fairness. Accordingly, the value of the fairness parameter is lower than 1, since consumers are charged different prices. If inequity aversion is introduced (scenario II), the price for group 2 remains nearly unchanged, while the price charged to group 1 decreases. As a result, prices are more homogeneously distributed (although there is no unit price), since the price-setting policy balances between maximizing revenue and providing a fair market outcome. This is further illustrated by a higher value of the fairness parameter than in scenario I. The higher reward is obtained in scenario I, where prices are set on a discriminatory basis without considering fairness. However, the reward obtained when setting differential prices while simultaneously taking customers' inequity aversion into account is just slightly lower than that in scenario I.

	Hyperparameters				Results			
Scenario	α	δ	k	β	$\bar{a}_1$	$\bar{a}_2$	$\bar{r}$	$\bar{f}$
I	0.10	0.90	10	0	0.636	0.204	0.742	0.874
II	0.10	0.90	10	1	0.599	0.204	0.740	0.900

Table 2: Parameter values and averaged results of the final 1000 periods

In order to compare the results of the algorithmic learning procedure with the analytical solution of the profit maximization problem[28], the market outcomes for both scenarios are calculated. If fairness is not considered, it can be seen here that it is beneficial for a monopolist to charge a price that is almost equal to the willingness to pay of customer group g_1 in order to extract as much consumer surplus as possible from this group. The results of the analytical calculation are shown in table 3.[29]

[28] Since we assume zero marginal cost, profit equals revenue in this case.

[29] Because we only allow for numbers with one decimal place (in the case of $k = 10$), the simulation results are not able to converge to the exact numbers of the analytical solution.

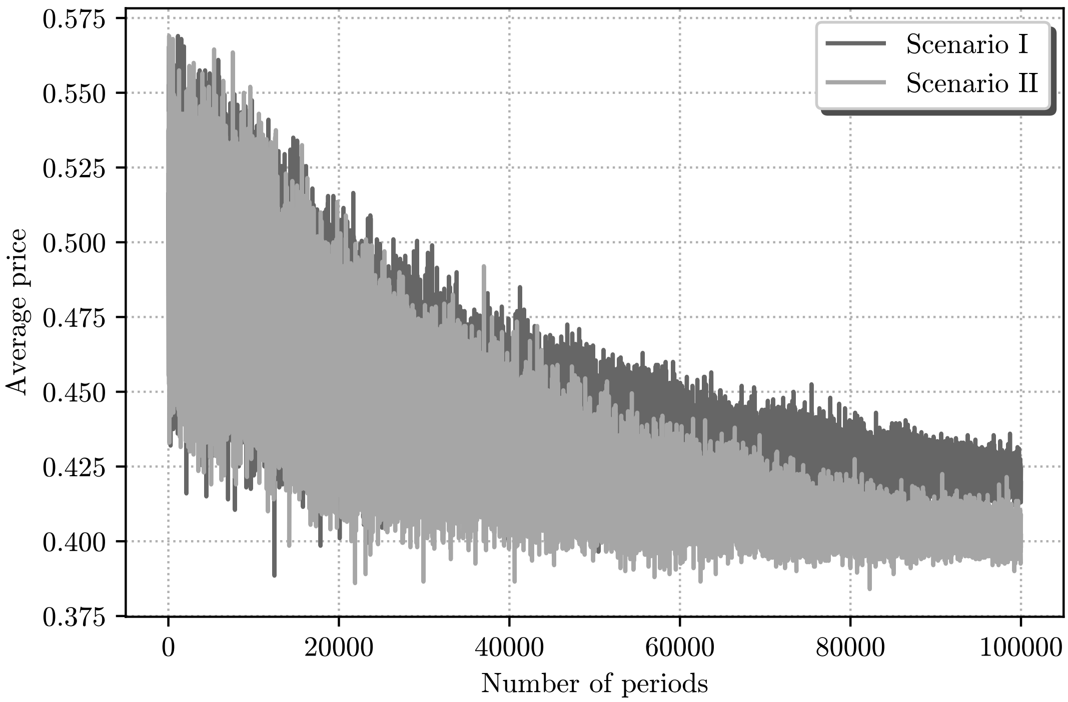

Figure 3: Average prices

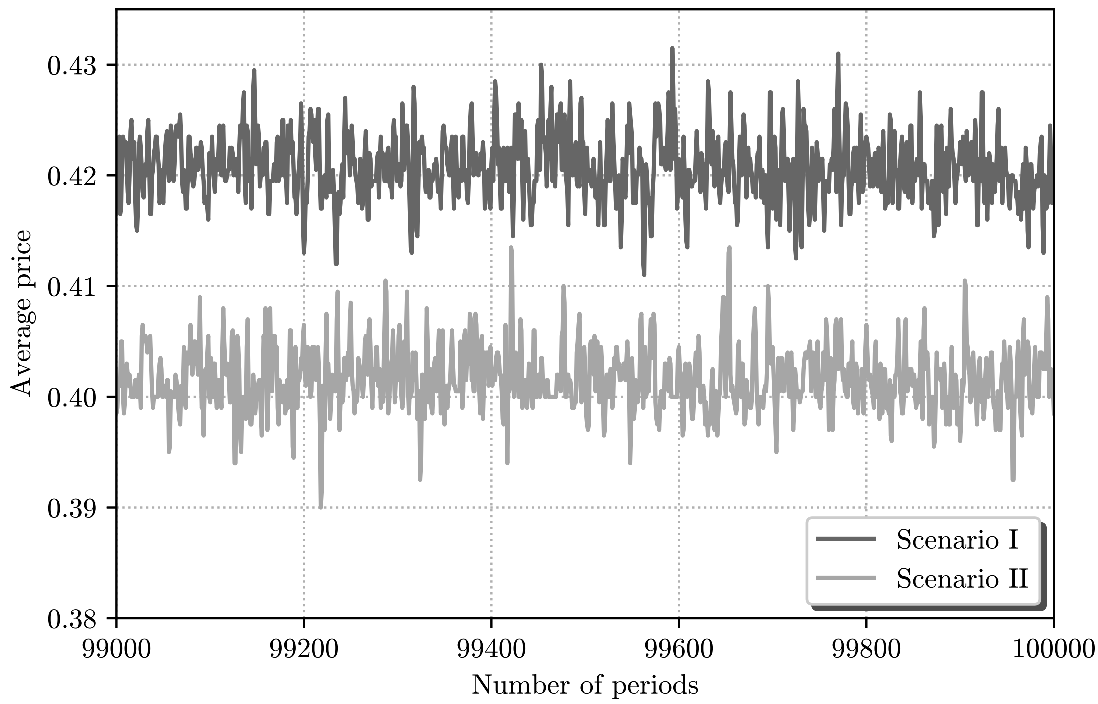

Figure 4: Average prices of final 1000 periods

	Prices		Profits (Revenues)	
Scenario	a_1^*	a_2^*	r_1^*	r_2^*
I	0.656	0.228	0.614	0.164
II	0.649	0.225	0.609	0.158

Table 3: Analytical solution

Certainly, this approach has some limitations. For example, the assumptions made above do not reflect the economic reality. In a perfectly competitive market, price discrimination is not feasible, as firms undercut each other until they end up in a (unit) price that equals marginal cost. However, prices might be differentiated due to search costs, lock-in effects, transportation costs, or asymmetric information. As a result, firms compete with each other, but markets do not seem to be perfectly competitive, and differential prices are likely to occur.[30] Another assumption that applies in reality is that of a limited action space: since firms have a price limit under which they would never sell the product (typically under average or marginal cost) the price interval has a lower bound[31], and the upper bound of the price interval is potential customers' maximum willingness to pay. However, it should be mentioned that each customer's actual willingness to pay cannot be determined by firms, as they only obtain imperfect information about each consumer. For this reason, personalized prices are not identical to perfect price discrimination.

Additionally, the implications of personalized pricing for consumer and total welfare are ambiguous, welfare could move in many directions relative to the benchmark of a unified market. Bergemann et al. (2015) showed that while obtaining additional information about consumers can never hurt the seller, it can increase both total and consumer surplus, decrease both, or increase one and decrease the other. Which form of market segmentation is used in practice is influenced by many factors, which may include both technological and legal limitations on how information can be collected and used. In our simulation, customer group 2 benefits from fair personalized pricing since they pay a lower price, while customer group 1 is worse off since they pay a higher price. With these specific numbers and parameters, the decrease in consumer surplus of customer group 1 is higher than the increase in consumer surplus of customer group 2. However, this distribution might change if the values for s and w (which determine the acceptance probability ϕ) change,

30 Much empirical work tests for the presence of price discrimination in imperfectly competitive environments, e.g. Shepard (1991) in the gasoline market, Goldberg (1995) in the market for European automobiles, and Leslie (2004) in the Broadway theater market. For a survey of price discrimination in imperfectly competitive markets, see Armstrong (2006).

31 Despite the willingness to obtain negative profits in particular situations, retailers in the EU are not allowed to sell below cost (called predatory pricing) according to TFEU, Art. 102. The reason given is that such behavior, exercised over a longer period, can drive smaller competitors out of the market.

or if more customer groups with different acceptance probabilities are introduced.[32] In our setting, both the average market price and the expected revenue increase whenever personalized prices are introduced. This is also in line with economic literature in the monopolistic (single-agent) setting.

5 Further Research

Using reinforcement learning, this paper shows how to maximize prices while maintaining a balance between revenue and fairness. Compared to the "standard" case of price discrimination, the results show an improvement in fairness while also obtaining the goal of maximizing revenue. However, compared to uniform pricing, the average market price increases.

Future research might include possible extensions of this study. First, the parameters (ϵ, f) could be refined in order to do justice to the complexity of this approach. For example, Jain's index does not account for self-centered inequity aversion, which occurs when people do not care about inequity that exists among other people and are only interested in the fairness of their own material payoff relative to the payoffs of others. According to Fehr and Schmidt (1999), the judgment and feelings associated with advantaged and disadvantaged price inequality are different, because consumers suffer more from inequality that is to their disadvantage (which is not considered in the current index). Second, the action space and the state space could be extended to allow for more prices (an increase in k). Additionally, more customer groups may be added. Finally, the comparison with other reinforcement learning methods, such as deep learning, and with evolutionary algorithms could be useful for obtaining further insights. The complexity of the algorithm might also be increased by including the processes of defining and refining customer groups as well as setting prices simultaneously. Using different algorithms might also allow for the investigation of multi-agent learning by simulating a competitive environment instead of a monopolistic structure, the outcomes of which are unclear in economic literature.

[32] Dubé and Misra (2019) conducted an experiment with an online recruiting company, comparing the existing uniform price, an optimized uniform price, and targeted prices. They found that customer surplus declines slightly with personalized pricing relative to uniform pricing. However, over 60% of customers benefit from personalized prices that are lower than the optimal uniform price.

References

Armstrong, M. (2006): Recent Developments in the Economics of Price Discrimination, in: *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by Blundell, R., Newey, W. K., and Persson, T., vol. 2, Econometric Society Monographs, Cambridge University Press, pp. 97–141, DOI: 10.1017/CBO9781139052276.006.

Bergemann, D., B. Brooks, and S. Morris (2015): The Limits of Price Discrimination, *American Economic Review*, Vol. 15, No. 3, pp. 921–957, DOI: 10.1257/aer.20130848.

Bolton, G. and A. Ockenfels (2000): ERC: A theory of equity, reciprocity, and competition, *The American Economic Review*, Vol. 90, No. 1, pp. 166–193, DOI: 10.1257/aer.90.1.166.

Bourreau, M. and A. de Streel (2018): The regulation of personalised pricing in the digital era, DOI: 10.2139/ssrn.3312158.

Dubé, J.-P. and S. Misra (2019): Personalized pricing and customer welfare, Chicago Booth School of Business Working Paper, DOI: 10.2139/ssrn.2992257.

Executive Office of the President of the United States (2015): Big Data and Differential Pricing, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf (visited on 01/24/2020).

Ezrachi, A. and M. E. Stucke (2015): Artificial Intelligence and Collusion: When Computers Inhibit Competition, DOI: 10.2139/ssrn.2591874.

Ezrachi, A. and M. E. Stucke (2016): *Virtual Competition: the promise and perils of the algorithm-driven economy*, 2nd ed., Cambridge, Massachusetts: Harvard University Press, DOI: 10.4159/9780674973336.

Fehr, E. and K. M. Schmidt (1999): A theory of fairness, competition, and cooperation, *The Quarterly Journal of Economics*, pp. 817–868, DOI: 10.2139/ssrn.106228.

Goldberg, P. K. (1995): Product differentiation and oligopoly in international markets: The case of the US automobile industry, *Econometrica*, Vol. 63, No. 4, pp. 891–951, DOI: 10.2307/2171803.

Jain, R. K., D.-M. W. Chiu, and W. R. Hawe (1984): A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems, Hudson, MA.

Kahneman, D., J. L. Knetsch, and R. Thaler (1986): Fairness as a Constraint on Profit Seeking: Entitlements in the Market, *American Economic Review*, Vol. 76, No. 4, pp. 728–741, DOI: 10.2307/j.ctvcm4j8j.13.

Klein, T. (2018): Assessing Autonomous Algorithmic Collusion: Q-Learning Under Short-Run Price Commitments.

Leslie, P. (2004): Price discrimination in Broadway theater, *RAND Journal of Economics*, Vol. 35, No. 3, pp. 520–541, DOI: 10.2307/1593706.

Maestre, R., J. Duque, R. Alberto, and J. Arevalo (2019): "Reinforcement Learning for Fair Dynamic Pricing," in: *Intelligent Systems and Applications*, ed. by Arai, K., Kapoor, S., and Bhatia, R., Springer International Publishing, pp. 120–135, DOI: 10.1007/978-3-030-01054-6_8.

Odlyzko, A. (2009): Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets, *Review of Network Economics*, Vol. 8, No. 1, pp. 40–60, DOI: 10.2139/ssrn.1095350.

Pigou, A. C. (1920): *The Economics of Welfare*, London: Macmillan, DOI: 10.2307/2222816.

Reinartz, W. (2002): Customizing prices in online markets, *Emerging Issues in Management*, No. 1, pp. 55–65, DOI: 10.4468/2002.1.05reinartz.

Shepard, A. (1991): Price Discrimination and Retail Configuration, *Journal of Political Economy*, Vol. 99, No. 1, pp. 30–53, DOI: 10.1086/261739.

Stigler, G. J. (1987): The Theory of Price, New York: Macmillan, DOI: 10.2307/1231731.

Sutton, R. S. and A. G. Barto (2018): *Reinfordement Learning: An Introduction*, Cambridge, Massachusetts: MIT Press, DOI: 10.1109/TNN.1998.712192.

UK Competition and Market Authority (2018): Pricing algorithms: Economic working paper on the use of algorithms to facilitate collusion and personalised pricing, CMA 94.

UK Office of Fair Trading (2013): The economics of online personalised pricing, OFT 1488.

Varian, H. R. (1989): Price Discrimination, in: *Handbook of Industrial Organization*, ed. by Schmalensee, R. and Willig, R. D., Vol. 1, pp. 597–654, DOI: 10.1016/S1573-448X(89)01013-7.

Watkins, C. and P. Dayan (1992): Q-Learning, *Machine Learning*, Vol. 8, No. 3, pp. 279–292, DOI: 10.1007/BF00992698.

Woodcock, R. (2017): The Bargaining Robot, *CPI Antitrust Chronicle*, https://ssrn.com/abstract=2972228.

Appendices

A Results Scenario I

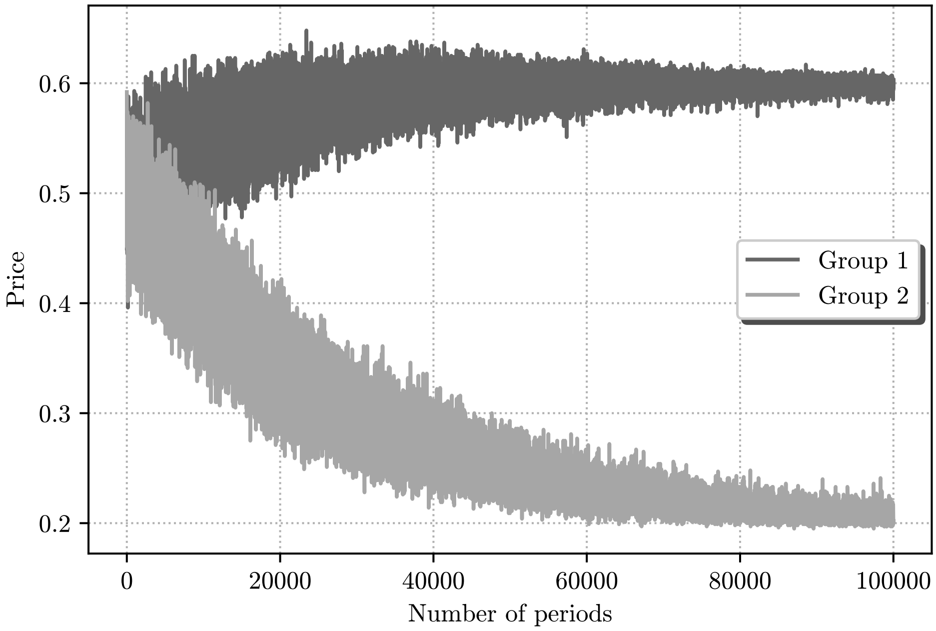

Figure 5: Action history scenario I

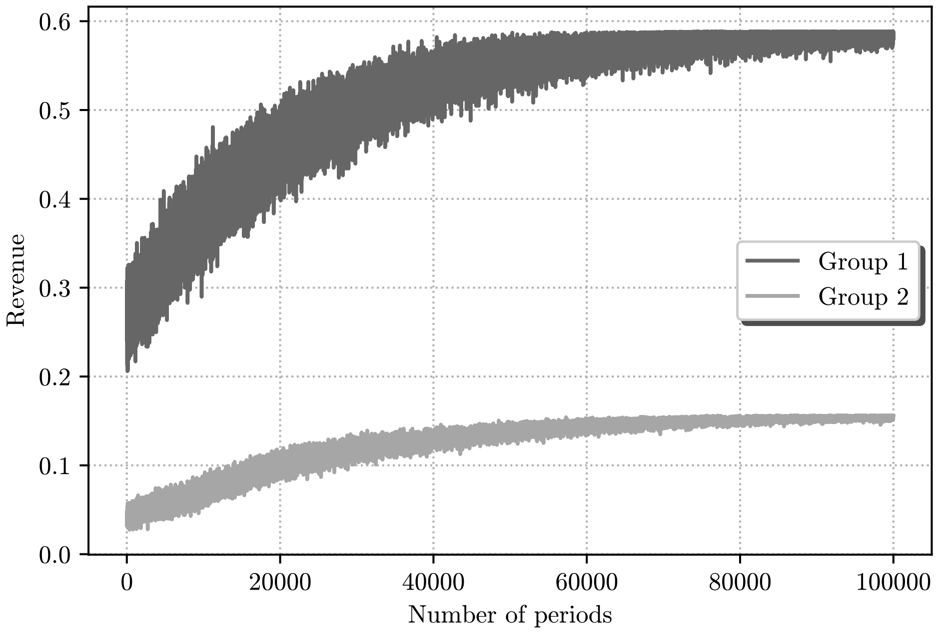

Figure 6: Reward history scenario I

B Results Scenario II

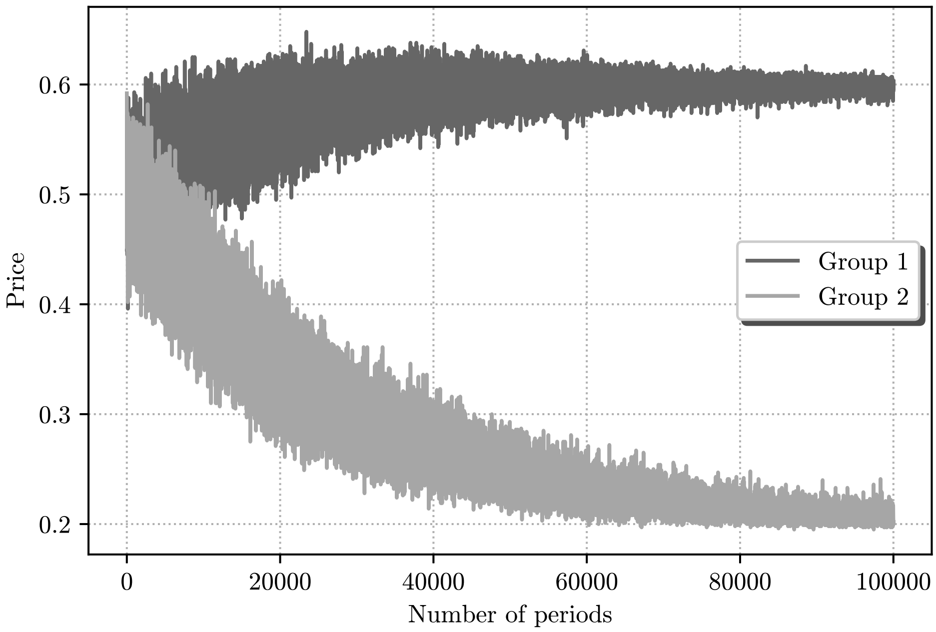

Figure 7: Action history scenario II

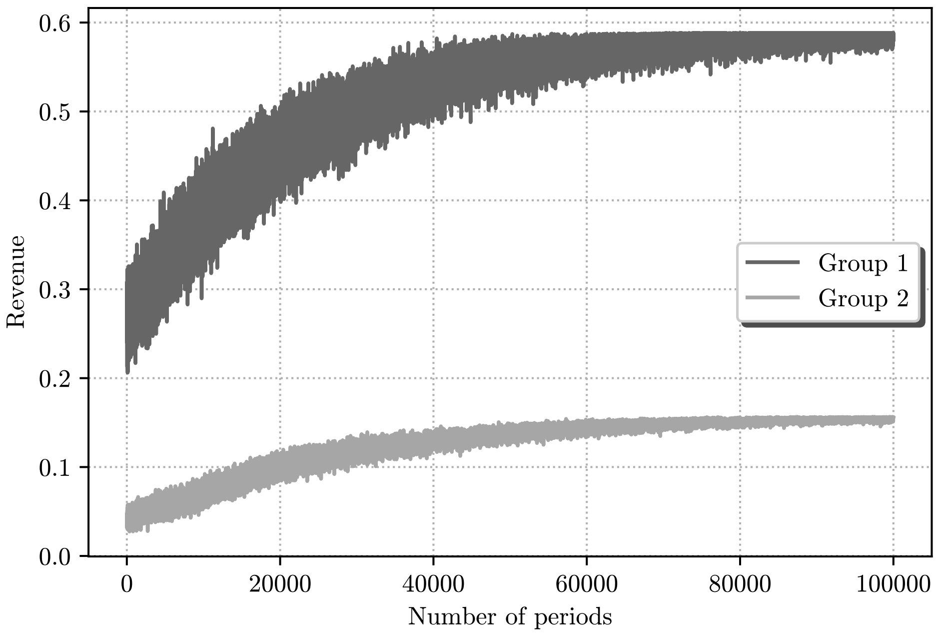

Figure 8: Reward history scenario II

Hohenheim Discussion Papers in Business, Economics and Social Sciences

This paper series aims to present working results of researchers of all disciplines from the Faculty of Business, Economics and Social Sciences and their cooperation partners since 2015.

Institutes

510 Institute of Financial Management
520 Institute of Economics
530 Institute of Health Care & Public Management
540 Institute of Communication Science
550 Institute of Law and Legal Sciences
560 Institute of Education, Labour and Society
570 Institute of Marketing & Management
580 Institute of Interorganizational Management & Performance

Research Areas

INEPA “Inequality and Economic Policy Analysis”
TKID “Transformation der Kommunikation – Integration und Desintegration”
NegoTrans “Negotiation Research – Transformation, Technology, Media and Costs”
INEF “Innovation, Entrepreneurship and Finance”
WDS “Work, Digitalization and Sustainability”

The following table shows recent issues of the series. A complete list of all issues and full texts are available on our homepage: https://wiso.uni-hohenheim.de/papers

No.	Author	Title	Inst
01-2018	Michael D. Howard Johannes Kolb	FOUNDER CEOS AND NEW VENTURE MEDIA COVERAGE	INEF
02-2018	Peter Spahn	UNCONVENTIONAL VIEWS ON INFLATION CONTRAOL: FORWARD GUIDANCE, THE NEO-FISHERIAN APPROACH, AND THE FISCAL THEORY OF THE PRICE LEVEL	520
03-2018	Aderonke Osikominu Gregor Pfeifer	PERCEIVED WAGES AND THE GENDER GAP IN STEM FIELDS	INEPA
04-2018	Theresa Grafeneder-Weissteiner Klaus Prettner Jens Südekum	THREE PILLARS OF URBANIZATION: MIGRATION, AGING, AND GROWTH	INEPA
05-2018	Vadim Kufenko Vincent Geloso Klaus Prettner	DOES SIZE MATTER? IMPLICATIONS OF HOUSEHOLD SIZE FOR ECONOMIC GROWTH AND CONVERGENCE	INEPA
06-2018	Michael Trost	THE WHOLE IS GREATER THAN THE SUM OF ITS PARTS – PRICING PRESSURE INDICES FOR MERGERS OF VERTICALLY INTEGRATED FIRMS	520
07-2018	Karsten Schweikert	TESTING FOR COINTEGRATION WITH TRESHOLD ADJUSTMENT IN THE PRESENCE OF STRUCTURAL BREAKS	520
08-2018	Evanthia Fasoula Karsten Schweikert	PRICE REGULATIONS AND PRICE ADJUSTMENT DYNAMICS: EVIDENCE FROM THE AUSTRIAN RETAIL FUEL MARKET	520

No.	Author	Title	Inst
09-2018	Michael Ahlheim Jan Neidhardt Ute Siepmann Xiaomin Yu	WECHAT – USING SOCIAL MEDIA FOR THE ASSESSMENT OF TOURIST PREFERENCES FOR ENVIRONMENTAL IMPROVEMENTS IN CHINA	520
10-2018	Alexander Gerybadze Simone Wiesenauer	THE INTERNATIONAL SALES ACCELERATOR: A PROJECT MANAGEMENT TOOL FOR IMPROVING SALES PERFORMANCE IN FOREIGN TARGET MARKETS	570
11-2018	Klaus Prettner Niels Geiger Johannes Schwarzer	DIE WIRTSCHAFTLICHEN FOLGEN DER AUTOMATISIERUNG	INEPA
12-2018	Martyna Marczak Thomas Beissinger	COMPETITIVENESS AT THE COUNTRY-SECTOR LEVEL: NEW MEASURES BASED ON GLOBAL VALUE CHAINS	520
13-2018	Niels Geiger Klaus Prettner Johannes Schwarzer	AUTOMATISIERUNG, WACHSTUM UND UNGLEICHHEIT	INEPA
14-2018	Klaus Prettner Sebastian Seiffert	THE SIZE OF THE MIDDLE CLASS AND EDUCATIONAL OUTCOMES: THEORY AND EVIDENCE FROM THE INDIAN SUBCONTINENT	INEPA
15-2018	Marina Töpfer	THE EFFECT OF WOMEN DIRECTORS ON INNOVATION ACTIVITY AND PERFORMANCE OF CORPORATE FIRMS - EVIDENCE FROM CHINA –	INEF
16-2018	Timo Walter	TRADE AND WELFARE EFFECTS OF A POTENTIAL FREE TRADE AGREEMENT BETWEEN JAPAN AND THE UNITED STATES	INEPA
17-2018	Jonas Frank	THE EFFECTS OF ECONOMIC SANCTIONS ON TRADE: NEW EVIDENCE FROM A PANEL PPML GRAVITY APPROACH	INEPA
18-2018	Jonas Frank	THE EFFECT OF CULTURE ON TRADE OVER TIME – NEW EVIDENCE FROM THE GLOBE DATA SET	520
19-2018	Dario Cords Klaus Prettner	TECHNOLOGICAL UNEMPLOYMENT REVISITED: AUTOMATION IN A SEARCH AND MATCHING FRAMEWORK	INEPA
20-2018	Sibylle Lehmann-Hasemeyer Andreas Neumayer	THE PERSISTENCE OF OWNERSHIP INEQUALITY – INVESTORS ON THE GERMAN STOCK EXCHANGES, 1869-1945	INEPA
21-2018	Nadja Dwenger Lukas Treber	SHAMING FOR TAX ENFORCEMENT: EVIDENCE FROM A NEW POLICY	520
22-2018	Octavio Escobar Henning Mühlen	THE ROLE OF FDI IN STRUCTURAL CHANGE: EVIDENCE FROM MEXICO	520

No.	Author	Title	Inst
24-2018	Peng Nie Lanlin Ding Alfonso Sousa-Poza	OBESITY INEQUALITY AND THE CHANGING SHAPE OF THE BODYWEIGHT DISTRIBUTION IN CHINA	INEPA
25-2018	Michael Ahlheim Maike Becker Yeniley Allegue Losada Heike Trastl	WASTED! RESOURCE RECOVERY AND WASTE MANAGEMENT IN CUBA	520
26-2018	Peter Spahn	WAS WAR FALSCH AM MERKANTILISMUS?	520
27-2018	Sophie Therese Schneider	NORTH_SOUTH TRADE AGREEMENTS AND THE QUALITY OF INSTITUTIONS: PANEL DATA EVIDENCE	INEPA
01-2019	Dominik Hartmann Mayra Bezerra Beatrice Lodolo Flávio L. Pinheiro	INTERNATIONAL TRADE, DEVELOPMENT TRAPS, AND THE CORE-PERIPHERY STRUCTURE OF INCOME INEQUALITY	INEPA
02-2019	Sebastian Seiffert	GO EAST: ON THE IMPACT OF THE TRANSIBERIAN RAILWAY ON ECONOMIC DEVELOPMENT IN EASTERN RUSSIA	INEPA
03-2019	Kristina Bogner	KNOWLEDGE NETWORKS IN THE GERMAN BIOECONOMY: NETWORK STRUCTURE OF PUBLICLY FUNDED R&D NETWORKS	520
04-2019	Dominik Hartmann Mayra Bezerra Flávio L. Pinheiro	IDENTIFYING SMART STRATEGIES FOR ECONOMIC DIVERSIFICATION AND INCLUSIVE GROWTH IN DEVELOPING ECONOMIES. THE CASE OF PARAGUAY	INEPA
05-2019	Octavio Escobar Henning Mühlen	DECOMPOSING A DECOMPOSITION: WITHIN-COUNTRY DIFFERENCES AND THE ROLE OF STRUCTURAL CHANGE IN PRODUCTIVITY GROWTH	INEPA
06-2019	Dominik Hartmann Cristian Figueroa Mary Kaltenberg Paolo Gala	MAPPING STRATIFICATION: THE INDUSTRY-OCCUPATION SPACE REVEALS THE NETWORK STRUCTURE OF INEQUALITY	INEPA
07-2019	Stephan Fichtner Herbert Meyr	BIOGAS PLANT OPTIMIZATION BY INCREASING ITS FLEXIBILITY CONSIDERING UNCERTAIN REVENUES	580
08-2019	Annika Lenz Muhammed Kaya Philipp Melzer Andreas Schmid Josepha Witt Mareike Schoop	DATA QUALITY AND INFORMATION LOSS IN STANDARDISED INTERPOLATED PATH ANALYIS – QUALITY MEASURES AND GUIDELINES	NegoTrans
09-2019	Thilo R. Huning Fabian Wahl	THE FETTERS OF INHERITANCE? EQUAL PARTITION AND REGIONAL ECONOMIC DEVELOPMENT	520

10-2019	Peter Spahn	KEYNESIAN CAPITAL THEORY, DECLINING INTEREST RATES AND PERSISTING PROFITS	520
11-2019	Thorsten Proettel	INTERNATIONAL DIGITAL CURRENCIES AND THEIR IMPACT ON MONETARY POLICY – AN EXPLORATION OF IMPLICATIONS AND VULNERABILITY	520
12-2019	Franz X. Hof Klaus Prettner	RELATIVE CONSUMPTION, RELATIVE WEALTH, AND LONG-RUN GROWTH: WHEN AND WHY IS THE STANDARD ANALYSIS PRONE TO ERRONEOUS CONCLUSIONS?	INEPA
13-2019	Vadim Kufenko Vincent Geloso	WHO ARE THE CHAMPIONS? INEQUALITY, ECONOMIC FREEDOM AND THE OLYMPICS	INEPA
14-2019	Laura-Kristin Baric Niels Geiger	POLITICAL IMPLICATIONS OF ECONOMIC INEQUALITY: A LITERATURE SURVEY	INEPA
15-2019	Jens Grüb	MERGERS AND PARTIAL TACIT COLLUSION	520
01-2020	David E. Bloom Victoria Y. Fan Vadim Kufenko Osondu Ogbuoji Klaus Prettner Gavin Yamey	GOING BEYOND GDP WITH A PARSIMONIOUS INDICATOR: INEQUALITY-ADJUSTED HEALTHY LIFETIME INCOME	INEPA
02-2020	Michael Kuhn Klaus Prettner	RISING LONGEVITY, INCREASING THE RETIREMENT AGE, AND THE CONSEQUENCES FOR KNOWLEDGE-BASED LONG-RUN GROWTH	520
03-2020	Simiao Chen Zhangfeng Jin Klaus Prettner	THE RETIREMENT MIGRATION PUZZLE IN CHINA	520
04-2020	Andreas Orland Davud Rostam-Afschar	FLEXIBLE WORK ARRANGEMENTS AND PRECAUTIONARY BEHAVIOUR: THEORY AND EXPERIMENTAL EVIDENCE	520
05-2020	Michael Fritsch Martin Obschonka Fabian Wahl Michael Wyrwich	THE DEEP IMPRINT OF ROMAN SANDALS: EVIDENCE OF LONG-LASTING EFFECTS OF ROMAN RULE ON PERSONALITY, ECONOMIC PERFORMANCE, AND WELL-BEING IN GERMANY	520
06-2020	Sibylle Lehmann-Hasemeyer Klaus Prettner Paul Tscheuschner	THE SCIENTIFIC EVOLUTION AND ITS ROLE IN THE TRANSITION TO SUSTAINED ECONOMIC GROWTH	520
07-2020	Vladan Ivanovic Vadim Kufenko	IT’S A MAN’S WORLD? THE RISE OF FEMALE ENTREPRENEURSHIP DURING PRIVATIZATION IN SERBIA	INEPA

08-2020	Stephanie Briel Aderonke Osikominu Gregor Pfeifer Mirjam Reuter Sascha Satlukal	OVERCONFIDENCE AND GENDER DIFFERENCES IN WAGE EXPECTATIONS	520
09-2020	Thomas Beissinger Joël Hellier Martyna Marczak	DIVERGENCE IN LABOUR FORCE GROWTH: SHOULD WAGES AND PRICES GROW FASTER IN GERMANY?	520
10-2020	Benjamin Jung	THE HETEROGENEOUS EFFECTS OF CHINA‘S ACCESSION TO THE WORLD TRADE ORGANIZATION	INEPA
11-2020	Timo Dimitriadis Xiaochun Liu Julie Schnaitmann	ENCOMPASSING TESTS FOR VALUE AT RISK AND EXPECTED SHORTFALL MULTI-STEP FORECASTS BASED ON INFERENCE ON THE BOUNDARY	520
12-2020	Timo Dimitriadis Andrew J. Patton Patrick W. Schmidt	TESTING FORECAST RATIONALITY FOR MEASURES OF CENTRAL TENDENCY	520
13-2020	Maria Koumenta Mario Pagliero Davud Rostam-Afschar	OCCUPATIONAL LICENSING AND THE GENDER WAGE GAP	INEPA
14-2020	Franziska Brall Ramona Schmid	AUTOMATION, ROBOTS AND WAGE INEQUALITY IN GERMANY: A DECOMPOSITION ANALYSIS	INEPA
01-2021	Paul Tscheuschner	ENDOGENEOUS LIFE EXPECTANCY AND R&D-BASED ECONOMIC GROWTH	520
02-2021	Kathrin Buchali	PRICE DISCRIMINATION WITH INEQUITY-AVERSE CONSUMERS: A REINFORCEMENT LEARNING APPROACH	520

IMPRINT

University of Hohenheim
Dean’s Office of the Faculty of Business, Economics and Social Sciences
Palace Hohenheim 1 B
70593 Stuttgart | Germany
Fon +49 (0)711 459 22488
Fax +49 (0)711 459 22785
wiso@uni-hohenheim.de
wiso.uni-hohenheim.de