# Learning Fair And Effective Points-Based Rewards Programs

**Chamsi Hssaine**
Marshall School of Business
University of Southern California
Los Angeles, CA 90007
hssaine@usc.edu

**Yichun Hu**
Johnson Graduate School of Management
Cornell University
Ithaca, NY 14853
yh767@cornell.edu

**Ciara Pike-Burke**
Department of Mathematics
Imperial College London
London SW7 2AZ
c.pike-burke@imperial.ac.uk

## Abstract

Points-based rewards programs are a prevalent way to incentivize customer loyalty. In these programs, customers who make repeated purchases from a seller accumulate points, working toward eventual redemption of a free reward. While they can generate significant revenue gains for the seller if implemented correctly, these programs have come under scrutiny of late due to accusations of unfair practices in their implementation. Motivated by these real-world concerns, this paper studies the problem of *fairly* designing points-based rewards programs, with a special focus on two major obstacles that put fairness at odds with their effectiveness: (i) the incentive to exploit customer heterogeneity by personalizing programs to customers' purchase behavior, and (ii) risks of devaluing customers' previously earned points when sellers need to experiment in uncertain environments. To study this problem, we focus on the popular "Buy $N$, Get One Free" (BNGO) rewards programs. We first show that the optimal *individually fair* program that uses the same redemption threshold for all customers suffers from a constant factor loss in revenue of at most $1 + \ln 2$, compared to the optimal personalized strategy which may unfairly offer different customers different thresholds. We then tackle the problem of designing *temporally fair* learning algorithms in the presence of demand uncertainty. Toward this goal, we design a "stable" learning algorithm that limits the risk of point devaluation due to experimentation by only changing the redemption threshold $O(\log T)$ times, over a learning horizon of length $T$. We prove that this algorithm incurs $\widetilde{O}(\sqrt{T})$ regret in expectation; this guarantee is optimal, up to polylogarithmic factors. We then modify this algorithm to ever only decrease redemption thresholds, leading to improved fairness at a cost of only a constant factor in regret. Finally, we conduct extensive numerical experiments to show the limited value of personalization in average-case settings, in addition to demonstrating the strong practical performance of our proposed learning algorithms.

## 1 Introduction and Model

Loyalty programs have long been a way for companies to increase their revenues, beginning with the introduction of grocery store trading stamps in the late 1800s [NJ.com, 2013]. Since then, they

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

have exploded in popularity, with over 90% of companies maintaining some sort of loyalty program in 2016, and the average customer enrolled in over 15 loyalty programs today [Vox, 2024]. One prominent form of loyalty program is the *points-based rewards* program. In a points-based rewards program, customers accumulate points every time they make a purchase; once the balance of points accumulated exceeds a certain redemption threshold, the customer is able to redeem her points for a reward (typically a free item or a discount on the next purchase). Prominent examples of points-based rewards programs include those offered by airlines, hotels, and casinos (Kopalle et al. [2012]), and those offered by the food service industry [Starbucks, 2025, McDonald's, 2025, Wendy's, 2025, Taco Bell, 2025], typically maintained through mobile applications.

Due to their preponderance in practice, the impact of points-based programs on customer behavior has been a topic of extensive study in the marketing literature. In particular, a substantial amount of empirical work has found evidence of a behavioral phenomenon known as the *points pressure effect*, which describes the idea that points-based programs give customers a goal to work toward (i.e., accumulating enough points to obtain a reward). This goal incentivizes them to purchase more frequently than they would without the prospect of obtaining a reward; moreover, the rate at which they make purchases only increases the closer they are to achieving this goal [Kivetz et al., 2006, Hartmann and Viard, 2008, Kopalle et al., 2012]. This points pressure effect engenders the following trade-off for companies (henceforth referred to as *decision-makers*, or *sellers*): while rewards programs generate additional revenue due to the increase in purchase frequency as customers approach the redemption threshold, this increase in revenue is immediately followed by a revenue loss from having to give out a reward. The decision-maker must therefore trade off between setting a lower redemption threshold, which would increase purchase probabilities but result in rewards being offered more frequently, and setting a higher threshold, which would reduce the regularity of rewards but result in a decrease in customers' purchase probabilities. The success of any rewards program hinges on the ability to set thresholds that optimally trade off between these incentives.

In practice, there are two major obstacles to the task of optimally setting redemption thresholds: (i) there is significant variability in the points pressure effect across customers [Kopalle et al., 2012], and (ii) the relationship between the number of points a customer has in stock and the probability with which they make a purchase is typically unknown. This paper is concerned with the design of effective learning algorithms for the problem of optimal goal-setting in points-based rewards programs, with a special focus on the *fairness* aspect of these programs. In particular, since many of these programs are implemented over long periods of time (as opposed to being offered as short-term promotions), our work posits that fairness becomes a first-order consideration for decision-makers on two fronts. The first challenge of customer heterogeneity introduces an *individual fairness* consideration: exploiting heterogeneity in customer behavior may lead to unfair outcomes (e.g., higher redemption goals being set for frequent customers), an effect which is exacerbated since customers are exposed to these differences over long periods of time. With respect to the second challenge, there exists a *temporal fairness* consideration: the stability of learning algorithms becomes extremely important in these settings, since customers' purchase decisions in these programs directly depend on the goal that has been set for them. As a result, changes in redemption thresholds (and in particular, *increases* in thresholds), are likely to be viewed as particularly unfair by customers. This claim is well-supported by a number of recent real-world instances wherein companies faced significant backlash after increasing the number of points required for redemption, effectively devaluing customers' hard-earned points. Prominent names associated with these scandals (which were often followed by swift reversals) include: Best Buy, Starbucks and Dunkin' Donuts [CNN, 2023], Chipotle [Reddit, 2023], Chick-Fil-A [PYMNTS, 2023], Microsoft [PCWorld, 2023], and Tesco [The Guardian, 2018]. Such concerns recently reached the highest levels of government, with the United States Department of Transportation (USDOT) launching an investigation into the four largest U.S. airlines' rewards programs. In the announcement of the investigation, the USDOT noted potential unfair practices in the way these companies set point values, highlighting in particular the devaluation of previously earned points [U.S. Department of Transportation, 2024].

Thus motivated, our work asks the following research questions:

*What is the impact of individual and temporal fairness constraints on the design of points-based rewards programs? How should we design stable, devaluation-free learning algorithms for this problem?*

Toward answering these questions, we consider a model in which a seller repeatedly offers a product at a fixed price to a finite population of heterogeneous customers. We conceptualize many of the points-based rewards programs referred to above via the classical "Buy $N$, Get One Free" (BNGO) program. Under this program, customers accumulate one point for each purchase that is made, and may redeem the item for free after they have made their $N$-th purchase. These BNGO programs are popular in practice due in large part to their simplicity, which has additionally made them prime candidates for tractable analysis in the operations literature [Liu et al., 2021]. Real-world examples of rewards given out within the context of BNGO programs include free hotel nights [Kopalle et al., 2012], golf rounds [Hartmann and Viard, 2008], coffee [Kivetz et al., 2006], and grocery items [Lal and Bell, 2003] (see Liu et al. [2021] for an excellent set of examples). Seminal work by Kopalle and Neslin [2003] also noted that frequent-flyer programs can be conceptualized as "Fly $N$ times, Get $(N+1)$-st flight free."

In our model, customers are partitioned into $K$ observable types (e.g., according to characteristics such as age and gender), and make their purchase or redemption decisions in each period according to an unknown, type-specific purchase probability. In line with the points pressure phenomenon, we assume the purchase probability is non-increasing in the number of points remaining until redemption. Importantly, these purchase probabilities are unknown, so the decision-maker must experiment with various redemption thresholds over a finite horizon of $T$ time periods. Our goal is to design an individually and temporally fair learning algorithm that incurs low regret relative to a clairvoyant policy that in each period selects the threshold maximizing the long-run average revenue.

## 2 Main Contributions

### 2.1 On the price of individual fairness in complete-information settings

Our first contribution relates to an important design question for a decision-maker seeking to implement a BNGO program: *To personalize or not to personalize?* More concretely, should a seller attempt to exploit customer heterogeneity by setting different redemption thresholds for different types of customer? In settings where the seller can discriminate between types (e.g., when types correspond to separate, tiered membership statuses), it is easy to argue that the answer is a resounding "yes," from a revenue perspective. However, in many practical settings (e.g., when types are defined according to protected characteristics such as race and gender), such differentiation is likely to be perceived as unfair by customers, potentially also running into ethical and legal issues. Therefore, in order to decide whether or not personalization is a risk worth taking, the seller must be able to quantify the revenue loss associated with a *fair* rewards program, which sets the same redemption threshold for all customers. Thus motivated, we consider the *price of fairness* of BNGO programs, defined as the ratio between the optimal personalized program, which may set a different redemption threshold for each customer type, and the optimal non-personalized program, which is constrained to set the same redemption threshold across all customer types.

Given the limited assumptions imposed on the relationship between points to redemption and purchase probabilities, one may a priori expect that there exist instances where the price of fairness is arbitrarily large. This could occur for instance if implementing a "Buy One, Get One Free" program is optimal for one type of customer, whereas for another type of customer, it is optimal to not implement any rewards program. Moreover, previous work studying the impact of fairness constraints on incentives for retention has found that the price of fairness can be unbounded [Freund and Hssaine, 2025]. However, in our first main contribution, we provide a uniform upper bound on the price of fairness, across all possible instances:

**Theorem 1.** *The long-run average revenue of the optimal personalized BNGO program is no more than* $1 + \ln 2 \approx 1.69$ *times that of the optimal non-personalized program.*

We complement this theoretical finding with extensive numerical experiments that show that the price of fairness may be much lower than this worst-case upper bound in average-case settings. These results yield the important managerial insight that a seller can not extract an arbitrary amount of revenue from heterogeneity in these settings.

### 2.2 Temporal fairness in learning

Having established a small price of fairness in complete-information settings, we turn to the question of designing *temporally fair* algorithms in the learning setting, where the dependence of customers' purchase probabilities on the number of points to redemption is unknown. In line with much of the literature on demand learning [Filippi et al., 2010, Broder and Rusmevichientong, 2012, Ban and Keskin, 2021, Bastani et al., 2021], we assume that customers' type-specific purchase probabilities follow a Generalized Linear Model (GLM) with unknown parameters. Following the previous discussion, we seek to find a single redemption threshold that maximizes the long-run average revenue across all customers.

As a building block towards the design of a temporally fair learning algorithm that never devalues customers' points, we first consider the task of *stable* learning, i.e., learning under a limited number of threshold changes. We propose a greedy epoch-based algorithm, Stable-Greedy, for this task. This algorithm partitions the horizon into epochs of geometrically increasing length. At the beginning of each epoch, given observations of customers' purchase decisions at their respective point balances, it computes the Maximum Likelihood Estimate (MLE) of the unknown GLM parameters, and solves for the revenue-maximizing threshold, given the MLE. To allow for the possibility that not offering a rewards program is optimal, we also compare the (known) revenue without a rewards program to this estimated revenue, terminating the rewards program if this difference exceeds an epoch-specific confidence parameter. Our algorithm achieves the desideratum of stability by only modifying the threshold $O(\log T)$ times throughout the horizon, and has strong regret guarantees:

**Theorem 2.** *For a fixed population of size $M$, Stable-Greedy incurs $\widetilde{O}(\sqrt{MT})$ regret in expectation.*

We show this is optimal up to polylogarithmic factors by proving a matching lower bound of $\Omega(\sqrt{MT})$ on the regret of any (potentially non-temporally fair) policy.

Despite its strong guarantees, the possibility remains that Stable-Greedy may devalue customers' points by increasing the threshold, albeit infrequently. To address this undesirable characteristic, we propose a devaluation-free modification (Fair-Greedy). While this algorithm is still stable in that it proceeds in epochs, instead of choosing the greedy threshold at the beginning of each epoch, it chooses the largest threshold within a consideration set of thresholds. Thresholds are included in this consideration set if and only if their estimated revenue under the MLE is close enough to that of the optimal greedy solution. Importantly, the consideration sets are nested, which guarantees that the sequence of thresholds is non-increasing (i.e., devaluation-free). Leveraging our previous regret analysis, we show that:

**Theorem 3.** *Fair-Greedy incurs only a factor of 2 loss relative to the regret bound of Stable-Greedy in the worst case, and is therefore also order optimal.*

In synthetic experiments, we observe the strong performance of both Stable-Greedy and Fair-Greedy, in addition to numerically demonstrating the trade-off between revenue and devaluation-free learning. Furthermore, we empirically show that both algorithms are robust to misspecification of the GLM.

From a technical perspective, our work uncovers the interesting fact that optimal learning algorithms do not need to explicitly explore in our setting. This lies in stark contrast to the extensively studied problem of pricing under demand uncertainty, for which the suboptimality of greedy algorithms is well-known in non-contextual settings [Broder and Rusmevichientong, 2012, Keskin and Zeevi, 2014, den Boer and Zwart, 2014]. Work on pricing in *contextual* settings has however shown that greedy algorithms may be optimal, under certain regularity conditions on the exogenous distribution from which contexts are drawn [Qiang and Bayati, 2016, Javanmard and Nazerzadeh, 2019]. In contrast to this latter set of results, we require no additional assumptions on customers' purchase probabilities to show the optimality of Stable-Greedy. Rather, our results follow from the fact that customers running through multiple redemption cycles throughout a single epoch induces a form of "natural exploration." This phenomenon guarantees sufficient variability in the points to redemption that the resulting MLE is a high-quality estimate of the unknown parameters. The technical crux of our work lies in demonstrating this fact, which relies on deriving a lower bound on the minimum eigenvalue of the empirical Fisher information matrix (henceforth referred to as the design matrix) of each epoch. Proving this requires a careful analysis that considers a Markov chain representation of a customer's points to redemption and derives a new Chernoff-type bound for the concentration of samples from this Markov chain. This differs from the analysis in related problems, where the assumption of i.i.d. contexts significantly simplifies the concentration results.

## References


Gah-Yi Ban and N Bora Keskin. Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science*, 67(9):5549–5568, 2021.

Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67(3):1329–1349, 2021.

Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

CNN. Best Buy, Dunkin' and Starbucks changed their rewards programs. Then came the backlash. `https://www.cnn.com/2023/01/14/business/best-buy-rewards-dunkin-starbucks-ctpr/index.html`, 2023. Accessed: 2025-01-17.

Arnoud V den Boer and Bert Zwart. Simultaneously learning and optimizing using controlled variance pricing. *Management Science*, 60(3):770–783, 2014.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems*, 23, 2010.

Daniel Freund and Chamsi Hssaine. Fair incentives for repeated engagement. *Production and Operations Management*, 34(1):16–29, 2025.

Wesley R Hartmann and V Brian Viard. Do frequency reward programs create switching costs? a dynamic structural analysis of demand in a reward program. *Quantitative Marketing and Economics*, 6:109–137, 2008.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Journal of Machine Learning Research*, 20(9):1–49, 2019.

N Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

Ran Kivetz, Oleg Urminsky, and Yuhuang Zheng. The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43 (1):39–58, 2006.

Praveen K Kopalle and Scott A Neslin. The economic viability of frequency reward programs in a strategic competitive environment. *Review of Marketing Science*, 1(1):000010220215465616 1002, 2003.

Praveen K Kopalle, Yacheng Sun, Scott A Neslin, Baohong Sun, and Vanitha Swaminathan. The joint sales impact of frequency reward and customer tier components of loyalty programs. *Marketing Science*, 31(2):216–235, 2012.

Rajiv Lal and David E Bell. The impact of frequent shopper programs in grocery retailing. *Quantitative Marketing and Economics*, 1:179–202, 2003.

Yan Liu, Yacheng Sun, and Dan Zhang. An analysis of "buy x, get one free" reward programs. *Operations Research*, 69(6):1823–1841, 2021.

McDonald's. MyMcDonald's Rewards. `https://www.mcdonalds.com/ca/en-ca/getmoremcds/mymcdonaldsrewards.html`, 2025. Accessed: 2025-01-17.

NJ.com. Made in Jersey: S&H Green Stamps - in the sixties, Americans were stuck on them. `https://www.nj.com/business/2013/11/made_in_jersey_sh_green_stamps.html`, 2013. Accessed: 2025-01-17.

PCWorld. Microsoft guts microsoft rewards points, and its fans are outraged. `https://www.pcworld.com/article/2160414/microsoft-guts-microsoft-rewards-points-and-its-fans-are-outraged.html`, 2023. Accessed: 2025-01-17.

PYMNTS. Chick-fil-a waters down rewards and hopes customers stick around. https://www.pymnts.com/news/loyalty-and-rewards-news/2023/chick-fil-a-joins-qsrs-watering-down-rewards-programs-amid-inflation/, 2023. Accessed: 2025-01-17.

Sheng Qiang and Mohsen Bayati. Dynamic pricing with demand covariates. *arXiv preprint arXiv:1604.07463*, 2016.

Reddit. Reward points required for free entree raised to 1625 from 1400. https://www.reddit.com/r/Chipotle/comments/xsydih/reward_points_required_for_free_entree_raised_to/?rdt=64941, 2023. Accessed: 2025-01-17.

Starbucks. Starbucks rewards. https://www.starbucks.com/rewards, 2025. Accessed: 2025-01-17.

Taco Bell. Taco Bell Rewards. https://www.tacobell.com/rewards, 2025. Accessed: 2025-01-17.

The Guardian. Tesco delays Clubcard changes after customer backlash . https://www.theguardian.com/business/2018/jan/17/tesco-delays-clubcard-changes-customer-backlash-reward-loyalty-scheme#:~:text=Tesco%20has%20delayed%20changes%20to,the%20changes%20until%2010%20June., 2018. Accessed: 2025-01-17.

U.S. Department of Transportation. USDOT Seeks to Protect Consumers' Airline Rewards in Probe of Four Largest U.S. Airlines' Rewards Practices . https://www.transportation.gov/briefing-room/usdot-seeks-protect-consumers-airline-rewards-probe-four-largest-us-airlines-rewards, 2024. Accessed: 2025-01-17.

Vox. The golden age of retail loyalty programs is here. https://www.vox.com/money/354191/loyalty-rewards-programs-sephora-vib-amazon, 2024. Accessed: 2025-01-17.

Wendy's. Wendy's Rewards. https://www.wendys.com/rewards, 2025. Accessed: 2025-01-17.

References follow the acknowledgments in the camera-ready paper. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.