NBER WORKING PAPER SERIES

MARKET POWER AND REDEEMABLE LOYALTY TOKEN DESIGN

Kenneth S. Rogoff
Zhiheng He
Yang You

Working Paper 33201
http://www.nber.org/papers/w33201

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024, Revised October 2025

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. We would like to thank Zhiguo He, Michael Shin, Ke Tang, Susan Jane Thorp, Zhiwei Xu, and seminar and conference participants at Peking University, Workshop on Blockchain and Decentralized Finance at HKUST(GZ), Tokenomics 2024, Tsinghua University, Nanjing University, Univeristy of Sydney, University of Technology Sydney, Dishui Lake International Conference in Finance, and Central Bank Research Association (CEBRA) Annual Meeting for helpful comments. Zhiheng He thanks for the funding support of National Natural Science Foundation of China (723B2013). Yang You thanks for the funding support of Shenzhen Fintech grant SZRI2023-CRF-03. Natalie Kahn, Jinpu Li, Ziwen Sun, and Minghai Xu provide excellent research assistance. All errors are our own.

Market Power and Redeemable Loyalty Token Design
Kenneth S. Rogoff, Zhiheng He, and Yang You
NBER Working Paper No. 33201
November 2024, Revised October 2025
JEL No. G12, G32, G51, M20

ABSTRACT

Digitalization has led to a rapid expansion of loyalty tokens typically bundled as part of product price. An open question is whether, as technology evolves, firms will have a strong incentive to make loyalty tokens tradable, thereby raising regulation issues, including for monetary and banking authorities. Our main finding is that, under fairly general assumptions, firms will still have strong incentives to maintain non-tradability. This may seem surprising, given that an individual consumer would be willing to pay a premium for tradable tokens. However, provided the tokens do not offer a convenience yield to non-platform consumers, non-tradability compensates by expanding the outstanding stock of tokens outstanding at any one time, thereby providing a better overall form of financing. We further show that an issuer with stronger market power tends to make its revenue more token-dependent. Empirical results from airline mileage and hotel reward programs align with our theory.

Kenneth S. Rogoff
Harvard University
Economics Department
and NBER
krogoff@harvard.edu

Yang You
The University of Hong Kong (HKU)
Faculty of Business and Economics
Finance Unit
yangyou@hku.hk

Zhiheng He
Renmin University of China
School of Finance
hez@ruc.edu.cn

1. Introduction

Along with the explosion of cryptocurrency trading, loyalty tokens have also exploded across industries in this era of digitalization. Regulators face an important question of whether they should worry about loyalty tokens, too, will become widely traded, and if so, what restrictions might be required. The main result of this paper is that even though consumers would be willing to pay higher prices for goods bundled with tradable loyalty tokens, there are in fact very strong incentives for firms to make tokens non-tradable, even absent regulation. Indeed, as we show both theoretically and empirically, non-tradable loyalty tokens constitute a greater part of total revenues for firms with more monopoly power.

The world of loyalty programs and their associated tokens represents a highly pervasive and economically integrated form of non-fiat digital value. As a long-lived generation in the evolution of platform assets since the launch of the first airline frequent-flyer program in 1981, loyalty programs have generated substantial high-value business, with some even exceeding the equity value of companies themselves.[1] The reach of these programs is vast: the average U.S. consumer holds 19 (9 active) loyalty memberships,[2] and major corporations boast memberships in the hundreds of millions.[3] These loyalty programs provide an important channel for the issuers to finance themselves from consumers, in addition to traditional debt and equity finance. However, the economic mechanism of loyalty token issuance is unclear, particularly regarding how its finance role depends on functionality designs and market conditions. Furthermore, the technological evolution of these programs — from digital payments and big data to the recent integration of blockchain and artificial intelligence — is transforming traditional voucher-like points into more liquid and versatile digital assets. These features position loyalty tokens as neither a niche subset of the digital economy nor an immature speculative tool, but rather a mainstream application of tokenization with a massive, pre-existing user base across industries.

We present a simple and tractable model where the loyalty token issuer (e.g., a service provider or platform) bundles tokens with unit product sold, and decides the quantity of tokens bundled with each sale and whether tokens are allowed to transfer across token holders. When a customer accumulates enough tokens (with the amount being another choice variable), she can redeem them for one unit of the product. In a nutshell, if each sale comes with $1/M$ "free" tokens, consumers face a "buy M get one free" offer. The more tokens bundled, the more the purchase price of one unit service would rise accordingly to reflect the value of redemption rights. We view loyalty token issuance as a critical yet not well-understood

[1]For example, in 2020, American Airlines' program was valued between $18 B and $30 B, versus its equity market cap of less than $7 B. See Financial Times, https://www.ft.com/content/1bb94ed9-90de-4f15-aee0-3bf390b0f85e.

[2]See the Bond Loyalty Report 2024, https://info.bondbrandloyalty.com/2024-executive-summary.

[3]For example, Amazon Prime exceeded 180 million members as of March 2024. See Bloomberg news, https://www.bloomberg.com/news/articles/2024-04-16/amazon-prime-memberships-in-us-gain-8-to-new-high-after-lull.

channel to raise capital backed by its future service delivery, for example, American Airlines sought a $2.5 billion loan backed by its frequent flyer program during Covid-19.[4] The incentive behind token issuance is that the issuer has greater outside investment opportunities, and so discounts the future more sharply than consumers. Thus, a discount wedge drives the issuer's benefits from pre-sale, even at a fair discount for consumers. We further generalize our model by allowing for the market power of the platform into our economic analysis — product demand decreases as the purchase price rises (e.g., as more tokens are bundled), and the demand elasticity reflects the issuer's market power.

Our first key finding is that the issuer's optimal strategy is always to make tokens non-tradable, regardless of their market power, even if consumers are always willing to pay more for tradable tokens. The intuition lies in the underlying financing mechanism of bundling: it not only bundles tokens with product in the period of sale but, more importantly, effectively bundles current and future purchases as consumers need to accumulate tokens over time for a single redemption — we term this as a *time-series bundling*. Bundling enables a de facto cash flow swap that leverages the discount wedge by front-loading payments to the issuer. If tokens cannot be traded, consumers value early tokens less relative to later tokens as they need to wait for a longer time to redeem them, but they have to pay the same amount for each purchase. Allowing tradability breaks the *time-series bundling* of token holdings, thus eliminating the economic benefits of the cash flow swap. As a result, tradability is not preferred.

In addition to the aforementioned *time-series bundling*, two crucial economic trade-offs of bundling determine the optimal token issuance design. First, bundling drives consumers to a *forced strategy space* of token holdings, which implicitly requires enough enforcement capacity; otherwise, the consumer could give up the whole offer and leave the issuer. Thus, market power enters as a key parameter in the issuer's optimization problem, which in our model is captured by the sensitivity of external (consumer base) losses to unit price increases.[5] Second, bundling tokens, as compared to selling them separately (e.g. issuing platform currencies), has the drawback of *delayed cash flow*, as the issuer can sell tokens only to those with realized demand shocks, rather than to the whole consumer base.

Making tokens tradable disables the time-series bundling as consumers can then buy loyalty tokens and redeem them without waiting, yet partially compensates the issuer by yielding greater short-term cash flow. As a result, allowing tradability increases the token price under any given token bundling issuance policy. Nevertheless, we are able to prove the dominance of non-tradability mathematically; that is, the benefits of time-series bundling always outweigh the cost of delayed cash flow.

[4]CNBC News: https://www.cnbc.com/2021/03/08/american-airlines-plans-5-billion-bond-sale-backed-by-frequent-flyer-program.html.

[5]For an extreme example, consumers inevitably accept the monopolist's offer and cannot leave. Therefore, the monopoly issuer suffers no external loss from enforcing bundling. For oligopolies with less market power, however, the potential loss of the consumer base should be taken into account.

The second key finding is that issuers with greater market power manage to maintain a higher *token dependence*, defined as the revenue share of the token presale. Consequently, a monopoly issuer always favors "buy one and get one free," while a micro business may not have the enforcement leeway to benefit from token presales. This implies that industry giants have a stronger ability to monetize user bases, further translating their market power into profits through loyalty token issuance.

After developing the theory, we proceed to empirical analysis of airline and hospitality loyalty programs. We find that companies do tend to limit trade in tokens, either by adding fixed and/or proportional fees or restricting transfer flexibility. We then find that business scale is a critical determinant of the size of a company's token issuance (normalized by size). In particular, the variation in passenger numbers explains roughly 70% of the standard deviation of token dependence in the airline industry. Similar evidence is also found in hospitality. Last, we discuss airlines' efforts to seek higher convenience and demand probabilities of tokens by forming aviation alliances. Despite no change in business operation, an aviation alliance increases redemption probability, thus increasing consumers' willingness to pay for tokens and the issuer's ability to front-load revenue.

Our paper contributes to the large strand of literature on loyalty programs by offering a novel perspective on tokenization, in particular recognizing its value as a financing mechanism and relevant policy. Existing research comes mainly from marketing and operation management, as recently reviewed by Chen, Mandler, and Meyer-Waarden (2021), focusing on how the adoption of loyalty programs influences demand through factors such as customer retention (e.g., Bolton, Kannan, and Bramlett, 2000; Lewis, 2004; Evanschitzky et al., 2012; Kumar and Reinartz, 2016), consumer psychology and behavior (e.g., Sharp and Sharp, 1997; Liu, 2007; Chun and Ovchinnikov, 2019), complementary to user profiling (e.g., Nunes and Drèze, 2006; Rossi, 2018), and brand reputation evaluation (e.g., Selnes, 1993). The few studies in economics journals are primarily on switching costs (Klemperer, 1987, 1995) where loyalty-including economic arrangements can be thought as artificially generating switching costs (Banerjee, 1987), as well as the moral hazard problems from ticketing agencies (Basso, Clements, and Ross, 2009), where tokens are not directly issued to principal buyers. The contemporaneous work by Luo (2024) considers loyalty programs in a corporate finance model. The fact that loyalty tokens are an asset is recently recognized by Lim, Chun, and Satopää (2024); Chun and Hamilton (2024). Yet, to our knowledge, redemption-based loyalty tokens have not been theoretically modeled, and their optimal token supply, pricing strategy, and secondary market arrangement have not been studied.

Our paper also adds to the literature on the evolution of redeemable platform currencies and tokenization. A strand of literature has examined the role of tokenization in platform finance, where tokens are used as a medium of exchange within a platform (e.g., Cong, Li, and Wang, 2022; Goldstein, Gupta, and Sverchkov, 2024), and token price represents uncertain future convenience, particularly capturing network effects (e.g., Cong, Li, and Wang, 2021). Unlike this literature, we focus on the redemption functionality of tokens. As reviewed in Rogoff and You (2023), the long history of redeemable platform currencies can be traced back to trading stamps in the 1930s, includes the rise of loyalty tokens as the second generation, and has recently evolved into platform currencies (e.g., gift cards and digital cash). Rogoff and You (2023) argues that a major rationale for platform currencies is a financing mechanism, i.e., product presale, and shows that the countervailing force of the secondary market on pricing power undermines the issuer's incentives to make platform currencies tradable. Interestingly, over this long evolutionary process, older loyalty tokens have not been replaced by platform currencies. A fundamental question therefore is why loyalty tokens have proven so enduring.

Here, we show the unique feature of loyalty tokens by exploring the underlying microeconomic mechanism of bundling in redeemable token issuance, in particular uncovering the distinct financing mechanism and non-tradability motive — bundling eliminates the influence of the secondary market on issuers' pricing power; yet a preference for non-tradability remains, as tradability also mutes a crucial upside of loyalty tokens: time-series bundling provides a financing channel that goes beyond token presales and can even bypass native product delivery.

In the rest of the paper, Section 2 introduces the definition of loyalty tokens and clarifies key modeling elements. Section 3 sets up the model and presents main theoretical findings. Section 4 provides a number of interesting extended results and points out the potential impact of integrating new technologies. Section 5 presents empirical tests of our theoretical predictions and Section 6 concludes.

2. Loyalty Tokens: Definition and Mechanisms

A loyalty program is a marketing strategy designed to encourage consumers to shop or use a service frequently or regularly (e.g., Sharp and Sharp, 1997), where the notion of "frequent consumption behavior" gives rise to the concept of "loyalty." To incentivize such frequent (loyal) behaviors, a typical loyalty program offers one complimentary reward after multiple purchases. During this process, loyalty tokens are issued to consumers, effectively serving as (digital) representations of past consumption, and can be redeemed to claim the reward. Put differently, loyalty tokens are vehicles for future redemption rights.

Loyalty tokens are commonly used, though they are termed differently across industries, such as airline miles, hotel reward programs, etc. For example, consumers earn airline miles each time they purchase a ticket and accumulate enough miles (often over multiple transactions) to redeem a free trip. Notably, loyalty rewards are not a "free lunch" but rather services pre-sold (possibly at a discount), as issuers — such as airlines and hotel groups — always maintain pricing power. In this context, we see loyalty tokens are distinct from other forms of presales in three core ways, as visualized in Figure 1.

(1) Forced consumer strategy. Consumers are forced to pay for the service bundled with a certain number of tokens rather than choosing the quantity of tokens to hold. For the same reason, they may have to forgo the entire offer when faced with liquidity constraints.

(2) Delayed cash flow. To serve as a certificate of consumption, loyalty tokens must be sold in conjunction with a fiat-money purchase. Therefore, compared to direct presales, the issuer misses out on the front-loading benefits from consumers who do not purchase in the current period.

(3) Time-series bundling. To enable "rewarding frequent consumptions," loyalty tokens are typically designed to represent only fractional redemption rights, requiring consumers to complete multiple purchases to activate the redemption right. In other words, consumers' fiat-money purchases over time are bundled together by the redemption incentive.

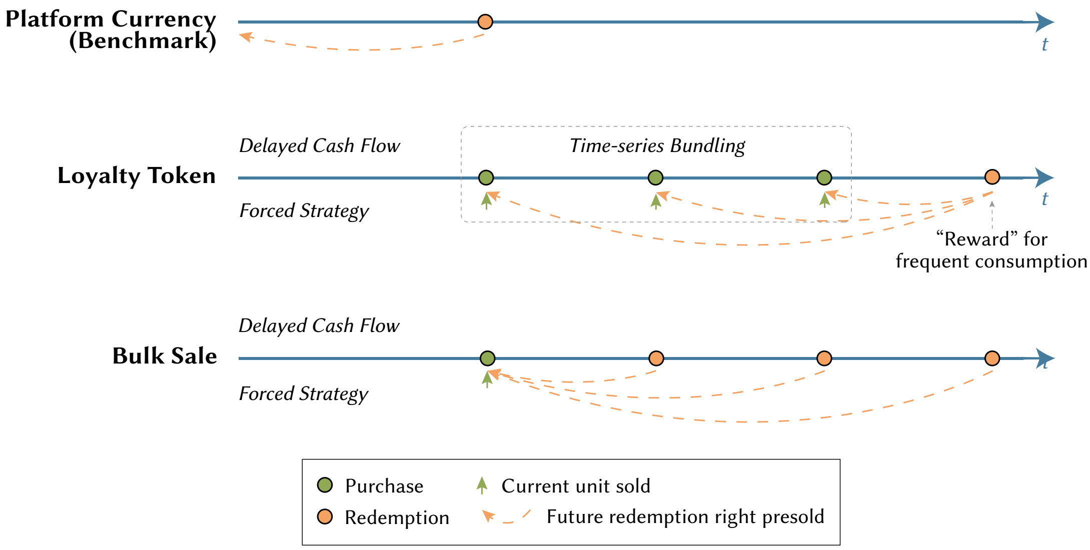

Figure 1. Loyalty Token v.s. Platform Currency v.s. Bulk Sale

This figure compares the timelines of three strategies related to presale, i.e., selling platform currency, issuing loyalty tokens (take "buy three and get one free" as an example), and bulk sale (four-pack bundle as an example), from top to bottom. The yellow dashed arrows indicate presales, and the green dashed arrows indicate current sales. The yellow (green) dots represent services redeemed (paid) by tokens (fiat money).

We term platform currencies as the benchmark (see Rogoff and You (2023) for detailed analysis), which are directly sold to consumers who may not yet demand the services. Compared to this benchmark, the central characteristic of loyalty token is bundling tokens sales with consumption. It thus has the drawback of ***delayed cash flow***, but naturally enables ongoing token issuance together with fiat-money purchases based on the ***forced consumer strategy*** and ***time-series bundling***. In fact, we show later that loyalty token issuance is more beneficial unless the issuer is extremely impatient.

One may also relate loyalty tokens to bulk sales, i.e., the consumer makes one purchase and receives multiple units of products or services to be redeemed in future, which are also termed as a bundle, but fiat-money purchases do not generate a ***time-series bundling***. Loosely speaking, issuing loyalty tokens can be termed as a "buy M get one free" strategy, whilst bulk sales refer to “buy one get N free,” $M, N \geq 1$. A lifetime membership can be regarded as an extreme example of a bulk sale, where $N = \infty$.

Bulk sales appear to be more beneficial, as they sell a greater amount of products in advance. However, why do so many industries and firms still issue loyalty tokens? First, bulk sales are significantly constrained by consumers’ liquidity concerns, especially when the unit price is high. That is, bulk sales impose a greater penalty because of consumers’ forced purchasing strategies. As a result, firms may compromise on the presale volume and instead issue loyalty tokens. Second, and more seriously, the challenge arises from intense regulatory pressure, as over-preselling exposes consumers to a risk of exit scamming. Bundled redemption rights are, in fact, liabilities for firms.[6] Large scale of bulk sales would lead firms to face massive product delivery obligations, thereby having a strong incentive to strategic default.[7]

In what follows, we formally model the issuance of loyalty tokens by capturing bundling token sales while considering the penalty arising from consumers’ liquidity constraints. The tractable framework distinguishes loyalty tokens from the benchmark platform currency — we focus on loyalty tokens (“buy M get one free” cases) in our baseline analysis. Our framework can easily adapt to study the bulk sales design, and we analyze the “buy one get N free” cases in Online Appendix A.

3. Theory

3.1. Model Setup

We consider a discrete-time infinite-period model, in which a token issuer (e.g., a traditional service provider or digital platform) provides a service or product at unit dollar cost (there is no inflation in the fiat money) and issues tokens representing redemption rights for future service. Following Rogoff and You (2023), there is a continuum of risk-neutral consumers. Every period, each consumer needs a service unit with probability p. The issuer and consumers have discount rates β^* and β, respectively. Due to greater outside investment opportunities, the issuer values the present more, yielding a discount wedge, $\beta^* < \beta$.

Issuance policy. The issuer sells one unit of service and provides $1/M$ units of loyalty tokens to the consumer. When a consumer accumulates enough fractional tokens to achieve one full unit, her tokens give her the right to redeem them for one unit of service.[8] Loyalty tokens are bundled with service delivery and are not sold directly to consumers.

[6] In practice, common bulk sales typically involve a one-time delivery of physical goods (e.g., bulk sales of items on Amazon or Costco), whereas in many industries, goods and services are consumed per use, so the bundled items would be (redemption rights for) future goods.

[7] There is even a popular idea that lifetime deals are sometimes used as a last-ditch effort to extract cash before the business struggles or even near bankruptcy. For example, see https://www.reddit.com/r/YouShouldKnow/comments/zf8xut/ysk_a_company_offering_a_lifetime_membership_or/.

[8] Here we focus on $M \geq 1$ as discussed in Section 2. We extend the entire analysis to the $M < 1$ cases, i.e., “buy one, get N free” strategies, in Online Appendix A.

Timeline. In each period, consumers (1) receive tokens (as a reward) for their consumption in the previous period, (2) trade tokens if allowed, (3) receive consumption shocks with probability p, and then (4) purchase or redeem for service. Section 4.1 extends our discussion to an alternative timeline where trade happens in each period after the realization of demand shocks instead of before the realization. Our main results remain robust in the alternative timeline.

Market power and the issuer's payoff. The issuer optimally designs the loyalty token issuance strategy (M, x) to maximize the loyalty payoff U_M^x, $M \geq 1$, $x \in \{NT, T\}$, where NT and T refer to issuing non-tradable and tradable tokens, respectively. By definition, loyalty token issuance is bundled with service purchase. Thus, the issuer has to charge a higher service price to cover the free redemption in the future. The market power reflects how consumer demand reduces in response to service price changes: if the issuer monopolizes the market, consumer demand won't be affected regardless of the quality of loyalty tokens bundled; however, in a more competitive market, charging a higher service price would drive budget-constrained consumers shift to other service providers that bundle fewer loyalty tokens.

We model this impact as an external penalty depending on service pricing in the issuer's optimization problem. In what follows, $c \cdot f(P)$ represents the external penalty, where $f(P)$ is a non-negative function of the service price P. $f'(P) \geq 0$, i.e., a higher price generates more losses. Without loss of generality, we normalize $f(1) = 0$, that is, the consumer demand equals to one unit when the issuer does not bundle any loyalty tokens and the service price equals 1.Imagine consumers have no outside options but to take the offer, so the monopoly does not suffer from a loss of its user base, that is, $c = 0$. The issuer-specific parameter $c \geq 0$ captures the market power: a lower c indicates a less harmful implementation of loyalty token issuance.[9]

In general, the loyalty token issuer's payoff U_M^x equals lifetime revenue from service sales minus the penalty, and the issuer's problem reads:[10]

$$\max_{M\geq 1,\ x\in\{T,NT\}} U_M^x \equiv \max_{M\geq 1,\ x\in\{T,NT\}} Rev_M^x - cf(P_M^x), \tag{1}$$

where Rev_M^x is the aggregate discounted infinite-period revenue when adopting the loyalty program (M, x). It can be further decomposed as $P_M^x Q_M^x$, where P_M^x is the service price under strategy (M, x), Q_M^x is the aggregate discounted fiat sales — although there are always p units of products delivered in one period, some of them are redeemed by tokens, therefore Q is dependent of (M, x).

[9]From another perspective, it can be interpreted as an opportunity cost of maintaining the consumer base in an oligopoly with switching costs (e.g., Klemperer, 1987): a large market power (small c) implies greater switching costs for consumers to other competitors, which leaves more room for price increases.

[10]Since the issuer always faces the same realized demand flow, the present value of the total product cost is constant. We omit the production cost term when it is not specified.

A natural benchmark is not to issue tokens, where the issuer's payoff is

$$U_0 = Rev_0 - cf(1) = \sum_{t=0}^{\infty} \beta^{*t} p = \frac{p}{1-\beta^*}.$$

It is not surprising that if c is sufficiently high, the issuer would lose its pricing power, and the optimal service pricing would be directly affected by the penalty. To focus on the central mechanisms in comparing the tradable/non-tradable loyalty token designs, unless otherwise stated, we restrict c such that the issuer can set the monopoly price optimally (e.g., $c \leq Q_M^x / f'(P),\ \forall P \leq P_1^T$), rather than lowering the price to avoid penalty.[11] We prove in the Appendix that all the subsequent propositions hold for any $c \geq 0$.

3.2. Equilibrium Prices

Non-tradable tokens. If tokens are not allowed to be traded, the consumer will enjoy $M+1$ units of services by purchasing M times, where M is a positive integer. Note that the probability that the consumer demands the ith service in period j ($j \geq i$) is given by

$$\mathrm{Pr}_{j,i} = \binom{j-1}{i-1} p^i (1-p)^{j-i}.$$

Therefore, by using a combinatorial identity, the consumer's willingness to pay for the i^{th} good delivery solves

$$\sum_{j \geq i} \beta^j \mathrm{Pr}_{j,i} = \left(\frac{\beta p}{1-\beta(1-p)}\right)^i \equiv a^i,$$

where a can be viewed as the effective discount rate of consumers. Similarly, $a^{*i} \equiv \left(\frac{\beta^* p}{1-\beta^*(1-p)}\right)^i$ is the present value of the Mth service supply from the issuer's view.

The issuer has pricing power over consumers, and charges the unit service price P_M^{NT} to overwhelm the whole willingness to pay for $M+1$ units through M sales,

$$P_M^{NT} \sum_{i=0}^{M-1} a^i = \sum_{i=0}^{M} a^i \quad \Rightarrow \quad P_M^{NT} = \frac{1-a^{M+1}}{1-a^M}. \tag{2}$$

Note that neither the issuer nor consumers discount the current period, the price for the current unit service equals one. Then unit token is effectively priced at $P_M^{NT} - 1 = \frac{(1-a)a^M}{1-a^M}$.

Tradable tokens. When tokens can be traded, consumers can sell them to other buyers — that is, a secondary market for loyalty tokens is introduced. Lemma 1 characterizes how tokens are traded among consumers and the market-clearing condition. Through tradability,

---

[11] For a general $c \geq 0$, the issuer's problem reads $\max_{M \geq 1,\ x \in \{T,NT\}} \max_{P \leq P_M^x} \left\{ Q_M^x P - cf(P) \right\}$, where note that Q_M^x relates to (M, x) yet is not directly affected by price. That is, if c were sufficiently large, the issuer's optimal pricing might be forced away from the monopoly price. In an extreme example where $c > Q_M^x / f'(P),\ \forall P \leq P_1^T$, any price increase would be overall harmful. Thus, any strategy earns less payoff compared to the trivial strategy with no token issuance.

consumers can avoid holding tokens that are insufficient for redemption. Consequently, the token market clears in the secondary market every period, determining the price $\hat{P}$ by ensuring that consumers are indifferent between holding zero and M tokens (where, again, M is the quantity needed to get a free unit of service).[12]

Lemma 1. ***Token allocation through the secondary market.***

(i) For any given issuance strategy M ($M \geq 1$), in each period, any user holds M tokens or zero tokens after trading in the secondary market.

(ii) Denote the service price as P_M^T. The secondary market price for unit token $\hat{P}$ solves

$$\hat{P} = \frac{p + (1-p)a}{M + a} P_M^T. \tag{3}$$

Remark. *The key intuition is that tradability speeds up redemption, as it omits the forced token-accumulating process by introducing the secondary market. Any single token can reliably be redeemed in the next period from its issuance by being combined with tokens from other consumers. In the non-tradable case, however, redemption is not feasible until M tokens are accumulated.*

For a consumer, one single purchase with fiat money will include a unit service plus a $1/M$ share of future service that can be sold in the next period. With pricing power over consumers, the issuer could charge a service price on the primary market that fully captures the resale revenue, i.e.,

$$P_M^T = 1 + \beta \hat{P},$$

which solves

$$\hat{P} = \frac{p + (1-p)a}{M} \quad \text{and} \quad P_M^T = 1 + \frac{a}{M}.$$

Proposition 1. ***Token price properties.***

Regardless of whether tradability is permitted, service prices will be higher if (i) the demand probabilities p are higher; or (ii) fewer tokens are required for one redemption.[13] *That is,*

$$\frac{\mathrm{d}P_M^x}{\mathrm{d}p} > 0, \quad \frac{\mathrm{d}P_M^x}{\mathrm{d}M} < 0, \quad \forall x \in \{T, NT\}.$$

In particular, $\forall M \geq 1$, the tradable tokens are always priced higher, i.e., $P_M^{NT} - 1 \leq P_M^T - 1$, where the equal sign is obtained if and only if $M \in \{1, \infty\}$ or $p = 0$.

Services bundled with tradable tokens are always priced higher, as consumers gain excess option value from tradable tokens. The only exception is when $M = 1$ or $p = 0$, in which

[12] Appendix A provides a detailed discussion of the flows of tokens among consumers and shows why they clear at the same price.

[13] Note that although M is defined as an integer, the price formulas can be viewed as continuous functions w.r.t. M. The implication naturally holds for integer values. In the following, we allow for the derivatives w.r.t. M without causing confusion.

case consumers naturally accumulate enough tokens to redeem, so there is no trade in the secondary market. At the other extreme, $M = \infty$ is equivalent to not issuing loyalty tokens. Whether tradable or not, higher demand probabilities p increase consumers' willingness to pay for redemption rights, leading to a higher price. On the other hand, smaller M means that more units $(1/M)$ of redemption rights are bundled, and also implies that in the non-tradable cases, the rights would be activated earlier.

3.3. Roles of Bundling and Tradability

Revisit the underlying insight of ***time-series bundling***. Equation (2) indicates that the issuer offers a menu that relates the total price paid to the number of purchases during the accumulation process and leaves zero consumer surplus. However, during this token accumulation game, each consumer actually receives different bundling rights in different periods, values them increasingly, but buys at the same price. This effectively enables a ***cash flow swap*** that front-loads payments to the issuer.

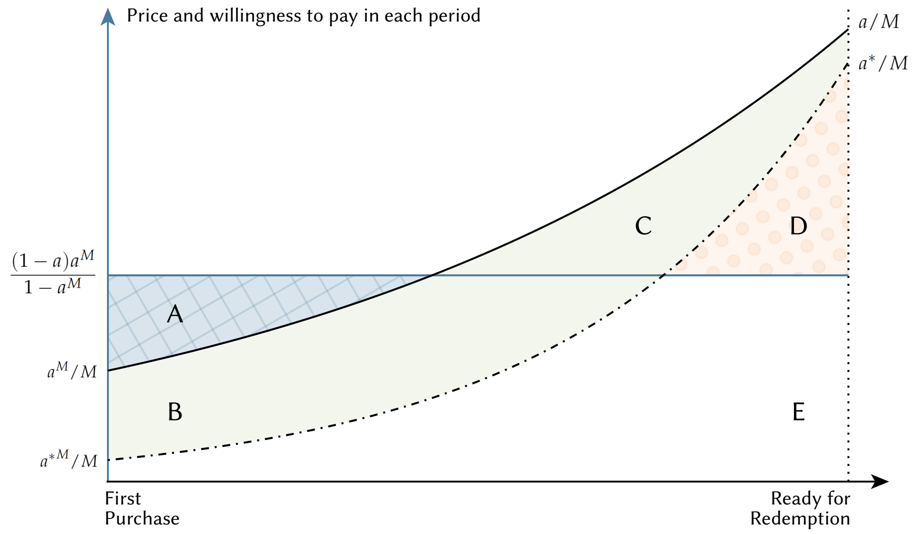

Figure 2. Cash Flow Swap by Time-series Bundling

This figure illustrates how time-series bundling effectively enables a cash flow swap. The solid (dot-dashed) line shows the consumer's willingness to pay for (the issuer's expected value of) the token in the current purchase, starting from the first purchase until the period when enough tokens are accumulated. The blue horizontal line shows the actual token price, $P_M^{NT} - 1$. Squares B and C show the discount wedge. Squares A and D respectively present the issuer's revenue surplus and loss for selling a token in single periods, demonstrating the cash flow swap.

Figure 2 illustrates how time-series bundling achieves a cash flow swap. Let "after-k right" denote the redemption right that becomes effective after k purchases.[14] Then it is straightforward to prove the consumer's willingness to pay for one "after-k right" is a^{k+1}, whereas the issuer's expected value is a^{*k+1}. In the first purchase, the consumer actually buys $1/M$ shares of "after-$(M-1)$ right." Similarly, the last token for one redemption, $1/M$ shares of "after-0 right," is valued a (a^*) by the consumer (issuer). If the tokens were sold separately (this already

[14]Note that the acquired (fragments of) redemption rights remain "inactive" and cannot be used until enough tokens are accumulated.

requires enforcement), the issuer earns the discount wedge, i.e., the present value of square B plus C as per the issuer (hereafter $B^* + C^*$). On the other hand, the time-series bundling actually sells a token at a constant price, $P_M^{NT} - 1 = \frac{(1-a)a^M}{1-a^M}$, which satisfies

$$\frac{a}{M} \geq \frac{(1-a)a^M}{1-a^M} \geq \frac{a^M}{M}, \quad \forall M \geq 1.$$

Therefore, consumers overpay in early periods as square A shows, then reduce spending as squares C and D show. Consumers accept this offer because, by (2), this offer makes them indifferent between buying and not buying, i.e., $A = C + D$. Since the issuer values the present more, from its perspective the present value of A is greater, i.e., $A^* > C^* + D^*$. Consequently, the overall net revenue ($A^* + B^* - D^*$) is larger than $B^* + C^*$.[15] Therefore, the issuer makes a cash flow swap and yields more front loads than the original discount wedge.

Consider the impact of tradability. First, tradability partially compensates for the delayed cash flow by selling additional resale options. From another perspective, the secondary market allows tokens to be held by anyone rather than only those who receive demand shocks. On the other hand, tradability undermines the intertemporal correlation of a consumer's token-holding quantities, thus turning off the time-series bundling. In particular, the issuer fails to sell "after-k rights" by adjusting its issuing strategy M but rather sells only "after-0 rights," since tokens are always gathered after the trade, as Lemma 1 shows. As a result, there is no room for cash flow swap.

3.4. Discounted Revenue of Monopolist Issuer

Recall that the issuer's problem (1) can be decomposed as calculating the discounted revenue Rev_M^x ideally as a monopoly issuer ($c = 0$) while then accounting for a penalty. This section analyzes the properties of Rev_M^x. The disruption to cash flow swap means that while consumers are willing to pay a premium for tradability, tradable tokens do not necessarily have a higher discounted revenue. To be precise, the loyalty token issuance (M, x) generates two effects. First, M shapes the share of token-in-advance, or equivalently, the aggregate discounted demand paid by fiat money. A smaller M generates more front-loaded revenue, leading to more services paid by redemption in the future. Second, x determines whether a reselling option is embedded in the tokens and leaves potential excess earnings from charging the reselling option value.

Revenue through non-tradable tokens. On the issuer side, the revenue comes from a series of "buy M and get one free" rounds, with each redemption starting a new round in the next period. At the beginning of each round, the consumer's i^{th} consumption delivers a present value (PV) of $a^{*i} P_M^{NT}$. The aggregate revenue reads

[15] Square D does not necessarily exist: when β^* is sufficiently small, $\frac{(1-a)a^M}{1-a^M} \geq a^*/M$. The logic applies similarly, i.e., $A = C \Rightarrow A^* > C^*$, whereas the net revenue becomes $A^* + B^* > B^* + C^*$. The issuer with lower β^* benefits even more from the cash flow swap.

$$
\begin{aligned}
Rev_M^{NT} &= \underbrace{\frac{1}{1-a^{*M+1}}}_{\substack{\text{Discount of}\\ \text{All Rounds}}} \underbrace{(p+(1-p)a^*)}_{\substack{\text{Discount of}\\ \text{First Purchase}\\ \text{in a Round}}} \underbrace{\frac{1-a^{*M}}{1-a^*}}_{\substack{\text{PV of}\\ \text{M Purchases}}} P_M^{NT} \equiv Q_M^{NT} \cdot P_M^{NT} \\
&= \frac{p}{1-\beta^*} \frac{1-a^{*M}}{1-a^{*M+1}} \frac{1-a^{M+1}}{1-a^M},
\end{aligned} \tag{4}
$$

where the aggregate discounted demand $Q_M^{NT} = \frac{p}{1-\beta^*}\frac{1-a^{*M}}{1-a^{*M+1}}$.

Revenue through tradable tokens. First, we solve for the number of goods consumers buy with fiat money in each period. Let $Q_{M,t}^T$ represent the fiat sales in period t (i.e., $\sum_{i=0}^{\infty} Q_{M,t}^T = Q_M^T$), and $S_{M,t}^T$ represent the outstanding units of redemption rights, i.e., there are $MS_{M,t}^T$ unredeemed tokens in period t. Naturally, $Q_{M,0}^T = p$, $S_{M,0}^T = 0$. With the total demand always equal to p, we obtain $\forall t \geq 0$,

$$
p = Q_{M,t}^T + pS_{M,t}^T. \tag{5}
$$

On the other hand, the total number of redeemable tokens equals the newly issued tokens plus the unredeemed ones from the previous period,

$$
S_{M,t+1}^T = \frac{Q_{M,t}^T}{M} + (1-p)S_{M,t}^T. \tag{6}
$$

Combining equations (5) and (6) with the initial values, we obtain

$$
S_{M,t}^T = \frac{1-\left[1-\frac{p(M+1)}{M}\right]^t}{M+1}, \quad Q_{M,t}^T = \frac{p}{M+1}\left[1-\frac{p(M+1)}{M}\right]^t + \frac{pM}{M+1}.
$$

The total revenue for tradable token issuance is

$$
Rev_M^T \equiv Q_M^T P_M^T = \sum_{t=0}^{\infty} \beta^{*t} Q_{M,t}^T P_M^T = \frac{p}{1-\beta^*}\frac{M+a}{M+a^*}. \tag{7}
$$

Lemma 2. ***Revenue and token quantity.***

The aggregate discounted revenue decreases with M regardless of tradability:

$$
\frac{\mathrm{d}}{\mathrm{d}M} Rev_M^x < 0, \quad \forall x \in \{T, NT\}.
$$

Since $Rev_M^x = P_M^x Q_M^x$, the impact of changing M can be decomposed into two opposite parts: a smaller M is associated with a higher token price, but more redemption rights granted reduce future services paid for by fiat money. Lemma 2 ensures that given x fixed, the former dominates.

As a result, without consumer base losses (and regulatory concerns), the monopoly issuer would always prefer more redemption rights embedded, and would even optimally choose a lifetime membership ($M \to 0$) if they do not enable tradability, as detailed in Online Appendix A.

It is also straightforward to show the equivalence of tradable and non-tradable tokens when $M \in \{1, \infty\}$,

$$Rev_1^{NT} = Rev_1^T = \frac{p}{1-\beta^*}\frac{1+a}{1+a^*}; \quad Rev_\infty^{NT} = Rev_\infty^T = \frac{p}{1-\beta^*} = U_0,$$

where $M = \infty$ effectively means zero redemption rights embedded, equivalent to non-issuance, and no trade happens when $M = 1$.

Tradable vs. non-tradable tokens. Despite the monotonicity of discounted revenues w.r.t. M, the trade-off of issuance designs in the second dimension, $x \in \{T, NT\}$ is more complex according to the mechanisms discussed in Section 3.3, making the comparison of strategies opaque, as visualized in Figure 3.

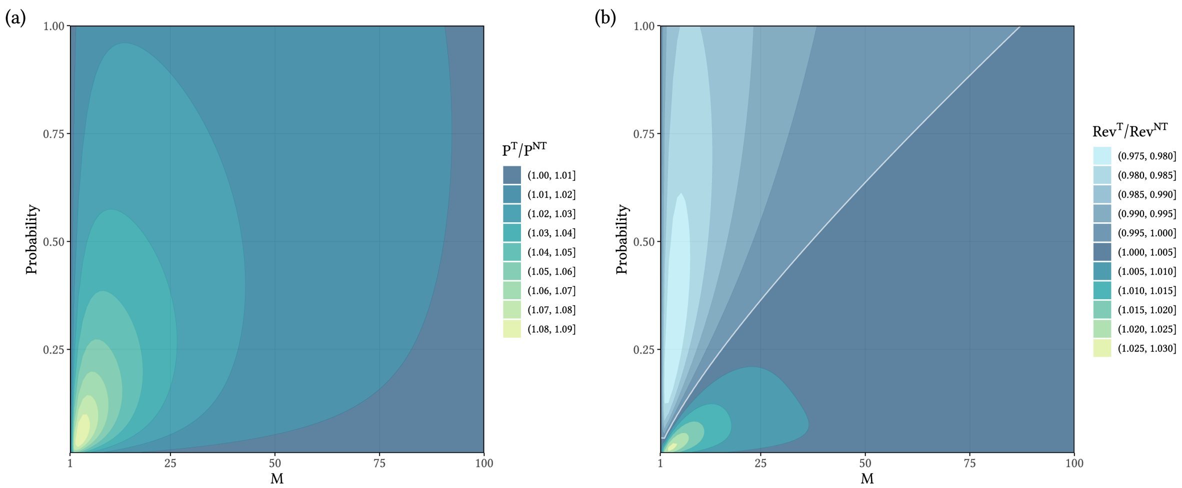

Figure 3. Tradable v.s. Non-Tradable Token Designs

This figure compares tradable and non-tradable tokens over the parameter space (p, M). Panel (a) shows the relative price dominance of tradable tokens, where the coloring scheme visualizes how P^T/P^{NT} changes across the parameter space. Panel (b) shows the revenue of tradable tokens relative to non-tradable tokens, i.e., Rev^T/Rev^{NT}, where $Rev^T = Rev^{NT}$ on the white indifference curve. $\beta = 0.95$ and $\beta^* = 0.9$.

Figure 3 (a) visualizes the comparative statics of the service price w.r.t. probability p and strategy M. Tradable tokens are always higher priced, since consumers obtain excess option values from selling them. The exception appears in $M = 1$ and $p = 0$, where consumers do not benefit from trading. Once the parameters leave these two boundaries, the relative price advantage of tradable tokens increases when p and M become smaller as consumers value tradability more when more redemption rights are bundled or they purchase goods more frequently.

Panel (b) shows the revenue comparison. Relative to the non-tradable case, tradability introduces a trade-off driven by two opposite economic forces, *compensating delayed cash flow* with high prices, yet *disabling time-series bundling*. With small M and p, the former effect dominates and $Rev_M^T > Rev_M^{NT}$, as it is far more advantageous to sell more redemption rights to the secondary market than to wait for the next batch of consumers with small probabilities. As

M increases, the compensation is reduced because fewer rights are sold, while the time-series bundling is lengthened, making disabling it more costly. As a result, Rev^T_M/Rev^{NT}_M converges to 1 more quickly than P^T_M/P^{NT}_M in Panel (a). Similarly, a greater p leaves delayed cash flow less of a concern, diminishing the importance of compensation, and can even make non-tradability more profitable. In general, Panel (b) demonstrates the impact of introducing tradability to the aggregate revenue under a given issuance policy — the revenue can both increase and decrease after creating a secondary market depending on the parameter (p, M). [16]

3.5. Dominance of Non-tradability

In this section, we formally study tradability in the general scenario where the issuer chooses the optimal M by maximizing (1) with any market power $c \geq 0$.

Recall that Figure 3 (b) shows tradability adds heterogeneous gains or losses of revenue under different strategies M — the issuer needs to find the global optimal payoff across values of M and tradability decisions, after accounting for enforcement losses. Without solving for the optimal M analytically, we can prove that any optimal strategy is always to issue non-tradable tokens, regardless of consumption frequency p and market power c.

Proposition 2. ***Dominance of non-tradability.***

Given any consumption probability $p \in (0, 1)$ and market power c, the issuer's optimal strategy is always to issue non-tradable tokens, i.e.,

$$U^{NT}_{M^{NT}_*} \geq U^T_{M^T_*}, \quad \forall c \geq 0,$$

where $M^x_* = \arg\max_{M \geq 1} U^x_M$, $x \in \{NT, T\}$. The equal sign holds only if $c = 0$.

The logic of the proof is that the advantage of tradability, higher price, may also be achieved (at least partially) through an alternative non-tradable issuance mechanism that generates a higher revenue. To be precise, denote (M, NT) (M non-tradable tokens for one redemption, $M > 1$) as the benchmark strategy. There are two ways to switch to another strategy with a higher unit price: (i) allowing tradability, i.e., (M, T); (ii) shortening the accumulation process, i.e., (m, NT) with $m < M$. Its unit price also increases as it is bundled with more shares of redemption rights, rather than excess reselling option values. As proved in Appendix A, it is always possible to find a strategy (m, NT) that matches the price of (M, T). Once $m > 1$, the token accumulation has not completely vanished — there is always room to enable a cash flow swap. Therefore, this strategy brings the same advantage (also external loss) as (M, T), but still benefits from time-series bundling.[17]

[16] This also highlights that Proposition 2 in Rogoff and You (2023), which states that revenue is higher for non-tradable platform currencies for any given issuance policy, does not apply to loyalty token.

[17] Note that in this proof, there is no need for m to be the optimal non-tradable plan M^{NT}_*. In certain scenarios, a non-tradable strategy with lower prices could even be more beneficial.

The economic intuition is the following: tradability systematically eliminates time-series bundling by allowing consumers to pool tokens through the secondary market. The reselling option value (also compensation for delayed cash flow) is essentially the premium of immediately unblocking the token's redeemability. Therefore, compensation entirely depends on the total sales of redemption rights. When both non-tradable and tradable designs are forced away from $M = 1$ and reduce redemption right sales, non-tradable issuance simultaneously enables a swap as an offset. Put differently, once the issuer allows tradability, it cannot adjust M to benefit from time-series bundling. Thus, tradability further limits the degree of freedom of the issuer in searching for the optimal M, thus making the issuer worse off.

An alternative interpretation is that tradability speeds up token redemption, and thus reduces aggregate discounted fiat sales. Once time-series bundling is in place, tokens are delayed in being activated for redemption, leaving consumers' up-front demands to be met by fiat purchases; they can only pay with tokens when they have accumulated enough of them. If non-tradable tokens were sold at the same high price as tradable tokens, they would generate more front-loaded cash flow. This is achievable by bundling more redemption rights. In general, both interpretations suggest that time-series bundling is key to the dominance of non-tradability, as it provides a financing tool, increasing front-loading beyond adjusting supply quantities. In addition, the logic and proof in Appendix A are not restricted to the parameter choices of c – even if the issuer were not able to set monopoly prices (as footnote 11 mentions), its optimum would be issuing non-tradable tokens.[18]

Link to redeemable platform currency. The dominance of non-tradability echoes the core conclusion about redeemable platform currencies in Rogoff and You (2023). However, the two are achieved through completely different mechanisms, and the non-tradability result correspondingly depends on an entirely different intuition. Recall that for the platform currency, the resale market (i.e., allowing tradability) makes it possible to buy tokens in the future, undercutting current consumers' willingness to hold large quantities. Therefore, competing with the resale market limits the issuer's pricing power in advance. Under a loyalty token design, such friction is systematically eliminated — tokens are bundled with current services, which means consumers have no need nor option to choose what quantities to hold from the primary market. The issuer's complete control over token supply keeps the resale market from jeopardizing the issuer's power to charge a high price upfront. In fact, the issuer even obtains a higher price, as Proposition 1 shows.

How could non-tradability still be dominant? We have shown that time-series bundling holds the key. This is unique to loyalty tokens and non-existent for platform currencies, where

[18]In the present work, we focus on token functionalities within the issuer's business scope. However, if tokens can be traded for outside use, tradability may be advantageous because of the added external convenience value. In that case, the issuer shares similarities with stablecoin issuers in the cryptocurrency market: it benefits from an over-the-counter (OTC) market in which consumers act as OTC brokers, while simultaneously incurring systemic external risks.

tokens are always redeemable once held by consumers. Recall Figure 1. To sell one "after-k right" by issuing platform currencies, $k > 0$, the issuer has to issue $k + 1$ units at the value of the marginal token, which leaves a positive consumer surplus.

Relative to platform currencies, bundling overcomes the competition with the secondary market, while simultaneously creating a cash flow swap if tokens cannot be traded. Both tradability and non-tradability are improved by bundling; meanwhile, the major cost comes from the delayed cash flow.

3.6. Market Power and Token Dependence

This section considers the role of market power in the optimal issuance policy. Issuers with different market power should have heterogeneous capacities to leverage it with loyalty tokens. To evaluate issuance effectiveness, we introduce an important indicator, token dependence, defined as the revenue share of loyalty token sales under any specific issuance design (M, x), i.e.,

$$\Theta_M^x \equiv \frac{\text{Revenue from Token Sales}}{\text{Total Revenue (from Token and Product Sales)}} = \begin{cases} \frac{a}{M+a}, & x = T; \\ \frac{a^M(1-a)}{1-a^{M+1}}, & x = NT. \end{cases} \tag{8}$$

It is straightforward to show $\mathrm{d}\Theta_M^x/\mathrm{d}M < 0$. Then given the choice x, parameters (p, c) and Lemma 2, a smaller M is equivalent to a higher token dependence, as well as greater revenue. We can compare the issuers' capacities to utilize the loyalty token financing tool by examining their M under optimal strategy.

Proposition 3. ***Market power and token dependence.***

The issuer with greater pricing power (smaller c) is able to maintain a higher token dependence (smaller M) whether or not tradability is allowed. That is,

$$M_*^x(c) \leq M_*^x(c'), \quad \forall 0 \leq c < c',$$

where $M_*^x = \arg\max_{M \geq 1} U_M^x$, $x \in \{NT, T\}$.

Proposition 3 indicates that greater market power is key to maintaining high token dependence. As Section 3.4 shows, the monopolist issuer's first best is $M = 1$, i.e., "buy one and get one free." However, for oligopolies and even small enterprises, a smaller M leads to a higher price but incurs consumer base shrinkage. Therefore, $M = 1$ might not be optimal. This logic is straightforward, yet the implication is noteworthy: issuing loyalty tokens serves as a booster for widening the market position gap. Only industry giants are capable of largely monetizing their user base or network traffic via loyalty programs, thus further consolidating their competitiveness. This echoes Banerjee (1987)'s argument, i.e., loyalty programs can effectively generate switching costs for consumers. Then, we can infer that not all issuers have this privilege: inherent high switching costs create room for loyalty tokens, further expanding switching costs. Importantly, issuers can issue loyalty tokens to leverage their market power for self-financing while stabilizing their customer base.

We next use numerical approaches to quantitatively examine M under optimal strategies over the parameter space (p, c), as shown in Figure 4 (a). The simulated optimal strategies, solved by brute force in the strategy space $M \leq 200$, $x \in \{T, NT\}$ for each point (p, c), show all to be non-tradable, confirming Proposition 2. For any given probability p, lower c (larger market power) is associated with a smaller optimal M, equivalently a greater token dependence Θ shown in Panel (b), corroborating Proposition 3. In addition, probability p affects token dependence non-monotonically, as it affects both the total sales and consumers' willingness to pay. With extremely low demand probabilities, consumers do not value future rights. In industries with low demand probabilities (e.g., international airlines), token presale generates significantly advanced cash flow, whereas in frequent-demand industries (e.g., daily retailing), issuing tokens fails to achieve long-term cash flow swap, making presales relatively less crucial relative to fiat sales.

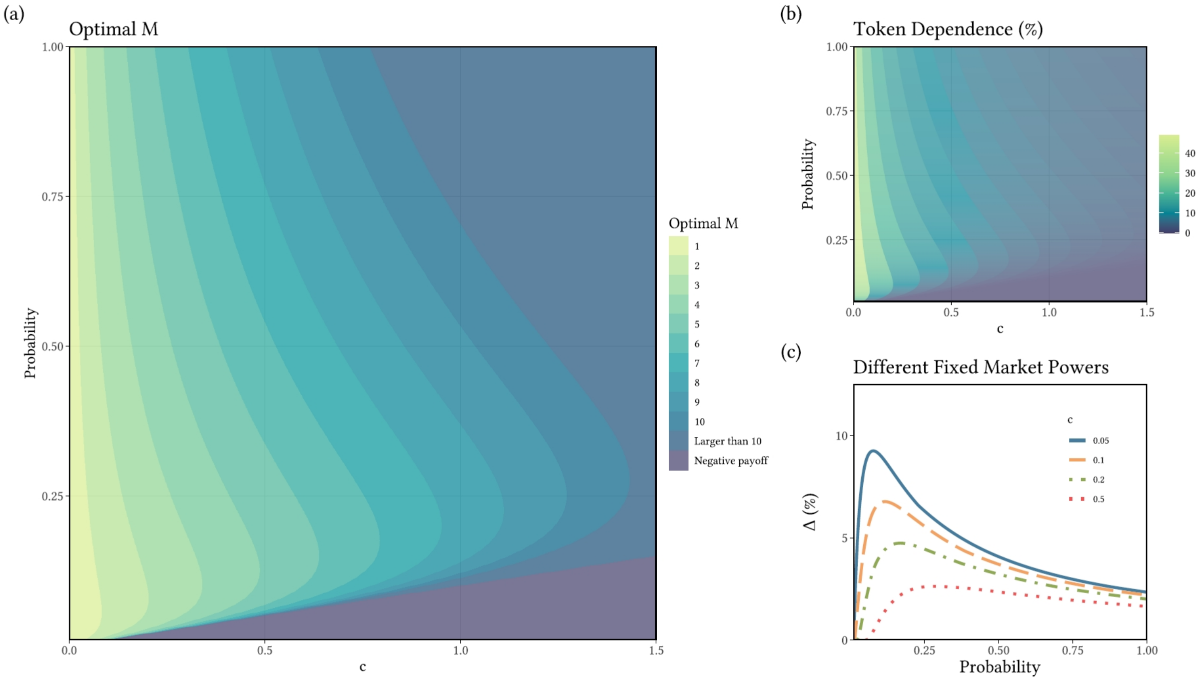

Figure 4. Optimal Issuance and Corresponding Token Dependence

Panel (a) plots the optimal issuance plan M under different probabilities p and market powers c. For each point (p, c), the optimal (discrete) choice of M is solved by brute force ($M \leq 200$, $x \in \{T, NT\}$), confirming the consistent preference for non-tradable tokens in achieving optimal outcomes. "Negative payoff" refers to cases where all the designs fail to yield positive payoffs, as mentioned in footnote 11. Panel (b) shows the corresponding token dependence Θ^* under the optimal issuance $(M_*^{x^*}, x^*)$. Panel (c) shows how $\Delta = (U^* - Rev_0)/Rev_0$ changes with probability p under different market power c. $\beta = 0.95$, $\beta^* = 0.9$, and $f(P) = P - 1$.

Figure 4, Panel (c) shows an alternative evaluation, $\Delta \equiv (U^* - Rev_0)/Rev_0$, where U^* is the payoff under the optimal issuance plan $(M_*^{x^*}, x^*)$, and Rev_0 denotes the benchmark revenue without token issuance. Δ is interpreted as the net payoff growth relative to the benchmark without issuing any tokens. Whereas Θ focuses on revenue generation within an issuance strategy, Δ presents clear theoretical implications on the trade-off of starting a loyalty program.[19] An issuer with greater market power (smaller c) can always generate higher payoff growth by issuing loyalty tokens, regardless of the demand probabilities. Therefore, the ability to maintain a large loyalty program is a unique advantage of greater market power.

Until now, we treat the probability p as an exogenous parameter of the loyalty token issuer. In practice, a higher p also captures a dimension of market power. Issuers may also strive to change demand probabilities through loyalty programs, for example, by collaborating with other companies to expand the realm of redeemable services or issuing loyalty points with credit card companies to attract more consumers. We further illustrate this point using the airline industry in Section 5.3.

4. Model Extension

4.1. Alternative Timeline and Secured Redemption

In our baseline model, the timeline assumes that token trading takes place before consumers know whether they have an immediate consumption need. Suppose, alternatively, consumers could trade tokens (if allowed) after experiencing consumption shocks; then it is not surprising that the token secondary market can efficiently allocate tokens to consumers who will redeem them immediately. This section shows that non-tradability is still dominant under this alternative timeline.

The alternative timeline makes the token secondary market very different from our baseline. Specifically, note that the issuer sells at most p tokens per period, while in the subsequent period, consumers who experience shocks require $pM \geq p$ tokens in total. Thus, all tokens are eventually spent by the end of each period; consequently, the secondary market clears at the price $\hat{P} = 1/M$. For consumers who pay in fiat money and receive one token, they know that they will either use the token with probability p, or sell it with probability $(1-p)$ in the next period. Thus, consumers' willingness to pay for a token in the primary market is always β/M, i.e., they value tokens at a one-period discount β, instead of at an effective discount a in the baseline, where $\beta > a$. Then the service price under the alternative timeline, $\widetilde{P_M^T}$, satisfies:

$$\widetilde{P_M^T} = 1 + \frac{\beta}{M} = P_M^T + \left(\frac{\beta}{a} - 1\right)\left(P_M^T - 1\right) > P_M^T. \tag{9}$$

Analogous to equations (5) and (6), the fiat sales in period t, $\widetilde{Q_{M,t}^T}$, and the outstanding

[19] We use Θ as the main indicator as it is closer to practice. As documented in the empirical section, enterprises' accounting statements disclose redemption revenues and unredeemed loyalty liabilities, while the external loss c is mostly unobserved or even non-financial. Online Appendix B provides more discussion on Δ.

units of redemption rights, $\widetilde{S^T_{M,t}}$, satisfy

$$p = \widetilde{Q^T_{M,t}} + \widetilde{S^T_{M,t}}, \quad \widetilde{S^T_{M,t+1}} = \frac{\widetilde{Q^T_{M,t}}}{M}.$$

The total revenue for tradable token issuance under the alternative timeline, $\widetilde{Rev^T_M}$, is

$$\widetilde{Rev^T_M} = \sum_{t=0}^{\infty} \beta^{*t} \widetilde{Q^T_{M,t}} \widetilde{P^T_M} = \frac{p}{1-\beta^*} \frac{M+\beta}{M+\beta^*}.$$

Equation (9) indicates that tradable tokens are priced even higher compared to the baseline, giving tradability the potential to beat the non-tradability.[20] The intuition is, when token trade occurs after consumption shocks, consumers no longer need to wait to redeem tokens in the next consumption shock arrives. Revisit the overall arrangement of issuing loyalty tokens: a "buy M get one free" game can be divided into two stages — *token accumulation* and *waiting for redemption*. In the baseline timeline, tradability eliminates the accumulation stage: some consumers are ready to redeem after trade but still need to wait for a demand shock. In the alternative timeline, tradability further eliminates the waiting stage, which translates into a higher reselling option value.

Non-tradable design with secured redemption. Following the above intuition, a comparable non-tradable design should also eliminate the *waiting for redemption* stage — consumers would still need to accumulate tokens, but the tokens can be used immediately with certainty once the accumulation process is completed. In essence, there is widespread practice to achieve this goal, e.g., allowing delegated redemption, setting up a reward marketplace, or even announcing cashback policy. We refer to these practical approaches collectively as "secured redemption" and discuss them in detail later.

When tokens are non-tradable with a secured redemption arrangement, the issuer charges the unit service price $\widetilde{P^{NT}_M}$ to overwhelm the willingness to pay for M units plus an immediate deterministic unit "reward" after M purchases,

$$\begin{aligned} &\widetilde{P^{NT}_M} \sum_{i=0}^{M-1} a^i = \sum_{i=0}^{M-1} a^i + a^{M-1}\beta \\ \Rightarrow \quad &\widetilde{P^{NT}_M} = 1 + \frac{a^{M-1}(1-a)\beta}{1-a^M} = P^{NT}_M + \left(\frac{\beta}{a} - 1\right)\left(P^{NT}_M - 1\right). \end{aligned} \tag{10}$$

According to equations (9) and (10), Proposition 1 still holds under the alternative timeline, and tradable tokens are still priced higher than non-tradable tokens.

In essence, the secured redemption arrangement to the non-tradable case is the natural mirror of the trading timing adjustment to the tradable case, as visualized in Figure 5. Mathematically, they induce isomorphic transformations on the service price: $\widetilde{P} = P + (\frac{\beta}{a} - 1)P$.

[20] Online Appendix C provides a detailed comparison. In particular, an extremely impatient ($\beta^* \to 0$) issuer would choose tradable tokens, as it only cares about the first-period sales, $p\widetilde{P^x_M}$.

In terms of economic intuition, they both eliminate the "wait for demand" stage for tokens. In addition, further allowing secured redemption for tradable tokens under the alternative timeline would add no difference, as tokens after trade are inevitably used immediately with certainty.

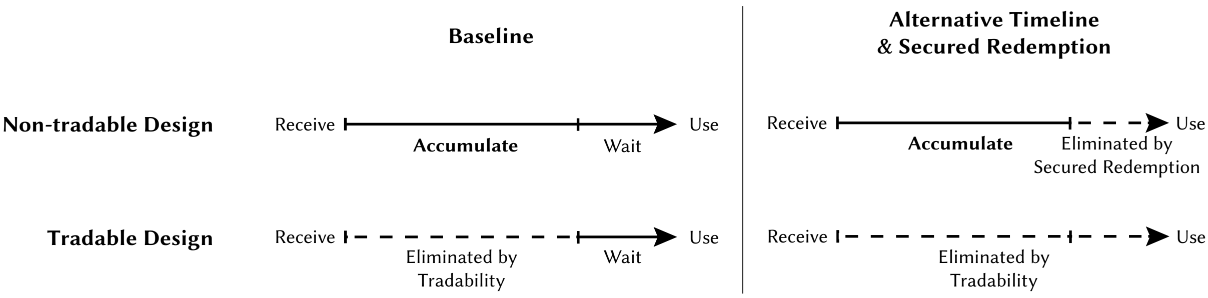

Figure 5. Different Combinations of Timeline and Tradability Settings

This figure uses a 2×2 diagram to compare the earn-use process of tokens under four settings discussed in this paper, i.e., the baseline non-tradable design and tradable design, the tradable design under the alternative (trading) timeline, and the non-tradable design with secured redemption.

To calculate the revenue from non-tradable tokens, $\widetilde{Rev}_M^{NT}$, we note two differences compared to equation (4): (i) a "buy M get one free" round no longer covers $M+1$ consumption periods but instead comprises M consumption periods plus one additional period; (ii) from the issuer's view, fulfilling each secured redemption without a demand shock incurs a unit cost while simultaneously eliminating the liability of M outstanding tokens. Then $\widetilde{Rev}_M^{NT}$ reads

$$\widetilde{Rev}_M^{NT} = \underbrace{\frac{1}{1-a^{*M}\beta^*}}_{\substack{\text{Discount of}\\ \text{All Rounds}}} \underbrace{(p+(1-p)a^*)}_{\substack{\text{Discount of}\\ \text{First Purchase}\\ \text{in a Round}}} \underbrace{\left[\frac{1-a^{*M}}{1-a^*}\widetilde{P}_M^{NT} - a^{*M-1}\beta^*(1-p)\right]}_{\substack{\text{PV of M Purchases and Cost of}\\ \text{Fulfilling Secured Redemption}}}. \tag{11}$$

Proposition 4. ***Dominance of non-tradability under alternative timeline.***

Under the alternative timeline — token trade (if allowed) happens after the realization of consumption shocks — given any consumption probability $p \in (0,1)$ and market power c, the issuer's optimal strategy is always to issue non-tradable tokens with secured redemption, i.e.,

$$U^{NT}_{M^{NT}_*} \geq U^{T}_{M^{T}_*}, \quad \forall c \geq 0,$$

where $M^x_* = \arg\max_{M\geq 1} U^x_M = \widetilde{Rev}^x_M - cf(\widetilde{P}^x_M)$, $x \in \{NT, T\}$. The equal sign holds only if $c = 0$ and $\beta^* = \beta$.

Proposition 4 ensures that our core finding — the dominance of non-tradability — remains robust under the alternative timeline. Recall Figure 5, non-tradability dominates the comparison in both columns, emphasizing that the core strength of non-tradability lies in the accumulation stage, which features time-series bundling, rather than in the waiting stage, which effectively means a pure product presale. That says, non-tradability compensates by expanding the outstanding stock of tokens outstanding at any one time, thereby providing a better overall form of financing beyond presale.

Secured redemption in practice. A natural approach to secured redemption is delegated redemption, for example, allowing clients to redeem a flight ticket for others. To acquire a redemption right, consumers still need to accumulate loyalty tokens. However, once the accumulation process is complete, they can exercise their redemption rights immediately, even when they themselves have no demand. Ticket buyers always value the unit redemption right at 1 as they are indifferent between buying a "token-redeemed ticket" without token accumulation and joining the loyalty program by paying the flight ticket price at a premium.

In practice, individual consumers face matching frictions in delegated redemption, which has fostered a market for redemption agencies.[21] More commonly, issuers also actively enhance delegated redemption by partnering with travel agencies (e.g., Booking.com). Agencies intermediate redemption on behalf of ad hoc consumers by internalizing the loyalty tokens they would otherwise hold as loyalty program members, and passing the benefits through in the form of lower prices (discounts or rebates).[22] This mechanism helps explain the long-run segmentation of purchasing behavior: some consumers maintain loyalty accounts and repeatedly buy and redeem through official channels, whereas others purchase continuously from agencies — typically at a lower unit price that effectively excludes the loyalty token bundle.

An alternative approach is to operate a rewards marketplace, where consumers can use their tokens to redeem a variety of goods. Sufficient variety differentiates consumers' demand probability for token redemption from that for the original service, even pushing the former up to 1. For example, American Airlines has launched the online marketplace AAdvantage Exchange℠, which allows its AAdvantage members to redeem miles for over 4,800 non-travel items from more than 100 brands.[23] Meanwhile, by setting the prices of goods in terms of tokens, the issuer can still determine the token accumulation process, i.e., determine M.

Finally, the most direct way to achieve secured redemption is to announce a loyalty token buyback policy. Recall equation (11). Instead of representing the cost of fulfilling secured redemption, the last term of revenue $\widetilde{Rev_M^{NT}}$ can be equivalently interpreted as the issuer buying back tokens whenever accumulation is completed, or more broadly, whenever a stated threshold is reached. For example, PayPal Rewards can be redeemed for cashback into a PayPal balance;[24] Taobao, the largest e-commerce platform in China, allows its 88VIP members to exchange their points for cash-equivalent checkout credit.

From delegated redemption to marketplaces and buybacks, we see that the circulation

[21] PointsPros constitutes an example. See https://pointspros.com.

[22] This still constitutes a (variant) form of delegated redemption, since the redemption right is exercised by the agency rather than by the individual consumer who originally accumulates tokens.

[23] Detailed information and redeemable items can be found on the official website: https://aadvantageexchange.com.

[24] For details, please see https://www.paypal.com/us/cshelp/article/what-is-paypal-rewards-and-how-does-it-work-help958.

of tokens can be separated from the issuer's main sales business and can even bypass the delivery of physical goods. This explicitly corroborates our finding that loyalty tokens serve as cash flow swap instruments, thereby opening up a financing mechanism beyond front-loading sales. Section 5.3 further discusses the issuer's attempts to obtain higher redemption demands, which effectively achieve the secured redemption design.

4.2. Commitment

In the baseline model, we focus on cases where issuers always follow their pre-announced token issuance scheme. However, in the absence of regulation, issuers may gain by deviating from their issuance policy, leading to commitment problems.

Proposition 5. ***Unreliable commitment to tradability.***

With naive consumers, unregulated issuers always have an incentive to claim to allow tradability and switch back to non-tradability in the future.

The key intuition is that the downside of tradability (the failure of time-series bundling) acts across periods, while the upside (compensation from higher prices) is achieved immediately after sales. Therefore, issuers have an incentive to charge a higher price and then prohibit tradability at any time after issuance, i.e., to directly impose mandatory intertemporal correlation of token holdings. In other words, issuers always hope that tokens will be redeemed long after the sale.

It is straightforward to see that with naive consumers and no regulations, the most beneficial strategy is to claim tradability at the beginning and then move back to the overall optimal strategy (M_*^{NT}, NT) when the number of unredeemed tokens reaches a maximum. In practice, when merchants are about to collapse, they tend to offer extra-long time-limited memberships, which are not designed to fulfill the loyalty rights, but are to sell as many redemption rights as possible in one shot. However, such a strategy is clearly impossible in the long run, since people will realize that the issuer's commitment to tradability is unreliable, and in particular, that tradability is suboptimal for issuers. Of course, with clever consumers, such worries would lower consumers' expected payoff of holding tradable tokens. In our model, this leads to an additional discount in market clearing condition (3).

The importance of credibility suggests a potential upside for the introduction of blockchain and cybersecurity technologies (e.g., as examined by Sönmeztürk, Ayav, and Erten, 2020, from an industrial perspective). While the literature extensively analyzes blockchain embedding for commitment to future token supply,[25] token quantities no longer need to be committed in the context of loyalty tokens. However, the essential goal of the commitment, i.e., token valuation, still needs to be ensured by the commitment to its functionality. Token tradability

[25]For example, Chod and Lyandres (2023) studies the product market competition stage, focusing on the advantages of introducing blockchain technologies and smart contracts to product tokens, showing that commitment to future token sales benefits both giants and new entrants.

as a functional extension requires additional commitment power.[26]

4.3. Perfect Competition

Then we consider the limiting case of perfect competition. Assume the issuers offer homogeneous services, and all services are perfect substitutes. Then, under perfect competition, $\forall M \geq 1$, an infinite number of issuers offer the issuance strategy with parameter M and earn zero profit. That is, the service price of first M purchases cover the cost of $M+1$ services,

$$P_M^{NT} \sum_{i=0}^{M-1} a^{*i} = \sum_{i=0}^{M} a^{*i} \quad \Rightarrow \quad P_M^{NT} = \frac{1-a^{*M+1}}{1-a^{*M}}.$$

Consumers solve the cost-minimization problem, $\min_{M\geq 1} C_M^{NT}$,[27] where

$$C_M^{NT} = \underbrace{\frac{1}{1-a^{M+1}}}_{\substack{\text{Discount of} \\ \text{All Rounds}}} \underbrace{(p+(1-p)a)}_{\substack{\text{Discount of} \\ \text{First Purchase} \\ \text{in a Round}}} \underbrace{\frac{1-a^M}{1-a} P_M^{NT}}_{\substack{\text{PV of} \\ \text{M Purchases}}} = \frac{p}{1-\beta} \frac{1-a^M}{1-a^{M+1}} \frac{1-a^{*M+1}}{1-a^{*M}}. \tag{12}$$

The cost C_M^{NT} here is identical in form to the revenue in equation (4), except that the positions of β and β^* are interchanged. Essentially, they represent a pair of duality problems: the issuer's revenue-maximization under complete monopoly and the consumers' cost-minimization under perfect competition ($c = 0$, as remarked below). In the monopoly case, consumers do not enjoy any surplus, and the price relates to consumers' discount. Conversely, under perfect competition, issuers do not generate any surplus, causing the price to be determined by the issuers' discount. The rest of terms are discounted values and thus relate to β^* in issuers' problems and β in consumers' problems.

Remark. *Revisit the economic interpretation of the parameter c. As defined in Section 3.1, c represents relative competitiveness in terms of the price sensitivity of the consumer base. The implicit assumption is that there are alternative issuers for these consumers to satisfy their demands. Therefore, a monopoly issuer has complete pricing power $c = 0$, i.e., consumers have no choice but to stay on the platform. Whilst under perfect competition, issuers still incur the external penalty $c \cdot f(P)$. However, it is the consumer side to solve the optimization problem so that the penalty term is not included in equation (12) at all.*

Consider issuing tradable tokens. Tokens in the secondary market still clear at the price that satisfies equation (3), while the issuers under perfect competition lose the pricing power

[26] Other forms of commitment problems can also exist. For example, the issuer can change the "price" when consumers finish token accumulation — hotels can require more points to redeem one free night, or hotels might arbitrarily require more points for redemption during public holidays. We do not discuss this possibility formally as they are irrelevant to our core tradability results.

[27] It is easy to show that once consumers adopt one service provider (issuer), they will continue to adopt it even if there are perfectly competing firms — suppose they choose another issuer, they would have to forgo the pre-consumed future service in terms of the loyalty tokens, which can be viewed as a switching cost. Therefore, consumers optimize their lifetime-discounted consumption cost.

to occupy the reselling revenue, and the price equals the service cost,

$$P_M^T = 1 + \frac{a^*}{M}.$$

As the secondary market price ensures that the consumers (sellers and buyers) have the same cost, we only need to solve one of the two, say the sellers. They purchase services through fiat money forever and earn spreads through the secondary market, as the following formula shows, which also appears to be a dual problem of equation (7).

$$C_M^T = \underbrace{\frac{1}{1-a}}_{\substack{\text{PV of All}\\ \text{Payments}}} \underbrace{\left(P_M^T - \beta \hat{P}\right)}_{\substack{\text{Purchase Cost minus}\\ \text{Selling Revenue}}} \underbrace{(p + (1-p)a)}_{\substack{\text{PV of}\\ \text{First Purchase}}} = \frac{p}{1-\beta}\frac{M+a^*}{M+a}. \tag{13}$$

By the nature of duality problems, one would expect that under perfect competition, consumers' cost-minimization would also obtain $M = 1$ in optimal, thereby ruling out the issuer's other choices for M. However, the issuer can still decide whether to introduce loyalty tokens. Proposition 6 formally solves for the equilibrium situation.

Proposition 6. ***Loyalty tokens under perfect competition.***

Under perfect competition, the only equilibrium is not to issue tokens.

The underlying mechanism is that consumers would prefer a greater scale of pre-consumption, i.e., $M = 1$, under perfect competition that minimizes their lifetime cost,

$$C_1^T = C_1^{NT} = \min_{M\geq 1, x\in\{T,NT\}} C_M^x = \frac{p}{1-\beta}\frac{1+a^*}{1+a}.$$

However, for any design $\{M, x\}$, the front-loaded cash flow from token issuance only offsets the issuer's present value of its corresponding production cost for redemption delivery. That is, issuers cannot make profits from bundling loyalty tokens, but only suffer the external loss from increasing unit service price:

$$U_M^x = \frac{p}{1-\beta^*} - cf(P_M^x) < U_0 = \frac{p}{1-\beta^*}, \quad \forall x \in \{T, NT\}, M \geq 1.$$

As a result, loyalty token issuance is not an equilibrium outcome under perfect competition.

Proposition 6 echoes the reality that famous loyalty programs tend to be found in oligopolistic, capital-intense industries, or at least are run by industry leaders, rather than by retailers under near-perfect competition (e.g., similar micro-merchants on Amazon), even though classic marketing incentives for loyalty programs seem universal across market structures. It turns out that successful loyalty token issuers shall have motives beyond marketing, such as creating financing channels, as we show.

5. Empirical Findings

This section provides empirical evidence for core testable model implications: (i) companies tend to limit token transactions, and (ii) companies with greater market power are able to maintain higher token dependence. In addition, we discuss practical efforts to manipulate redemption probabilities.

5.1. Designs for Blocking Tradability

Our model shows that token tradability is dominated by non-tradability, suggesting token issuers should disable token transactions. Proposition 5 proves that it would be optimal for airlines to claim that tokens are tradable but block token transactions when customers actually trade their tokens, as they can benefit from both a higher flight ticket price in the pre-sale and delayed redemption from time-series bundling. We present a descriptive analysis of loyalty point transfer policies consistent with our theoretical predictions. It is worth pointing out that the token transfer restrictions documented in our analysis could change over time, whereas our focus lies in discussing their underlying logic of limiting token usage.[28]

As Table 1 shows, none of the airlines or hotels in our sample allows unlimited free token transfers. The most direct approach is to ban tradability, labeled "N" in Column (1). However, this tactic may appear overly forceful or uncaring to consumers and might limit engagement in the loyalty program. Instead, as predicted in Proposition 5, issuers often employ various transfer restrictions to implement a token trading ban indirectly.

Fixed and/or proportional fees. The first approach to restricting free transfer is to add financial costs for transactions, e.g., processing fees and/or proportional transaction fees, as shown in Table 1, Columns (2)-(3). Fees are usually extremely high relative to the value of the tokens to be transferred. For example, even as one of the most affordable loyalty programs, the Mileage Plan™ run by Alaska Airlines only allows consumers to transfer 1,000 to 30,000 miles (i.e., tokens) in increments of 1,000 miles at a cost of $10.00 per 1,000 miles, plus a $25.00 processing fee per transaction.[29] At their estimated value in US dollars, about 1.5 cents each on average, a 30,000-mile transfer equates to a $450 transaction with a $325 fee — much more expensive than retaining the miles for the token sender's own use or purchasing points from the airline for the token receiver.[30] Columns (2)-(3) show that a costly transfer plan like this is a common practice among airlines and hotels. As a result, token transfers are rarely used by consumers.

[28] The data collected in this section are based on the latest officially-announced terms and conditions in March 2024, and may omit subsidiary-specific or related provisions, e.g., exemption policies, regional differences, etc.

[29] Documented in the official terms and conditions. Please see: https://www.alaskaair.com/content/mileage-plan/use-miles/share-gift-miles.

[30] The estimated value refers to *The Point Calculator*. Please see: https://www.thepointcalculator.com/us/airline/alaska/alaska-miles-value-calculator. The calculation is based on comparing main cabin fares using both cash and miles across several dates and destinations.

Table 1. Transaction Fees and Limitations

Corporation	Tx	$ Cost / 10^3 Units	$ Fee / Tx	Limit	Paper Work	Family Pool	Corporation	Tx	$ Cost / 10^3 Units	$ Fee / Tx	Limit	Paper Work	Family Pool
	(1)	(2)	(3)	(4)	(5)	(6)		(1)	(2)	(3)	(4)	(5)	(6)
Airline							Lufthansa	Y	11.2	0	Y	N	2+5
Alaska	Y	10	25	Y	N	N	Qatar	Y	Y	Y	Y	N	9
American	Y	0	15	Y	N	N	Singapore	W	1	0	Y	N	0+5
Avianca	Y	15	0	Y	N	N	Southwest	Y	10	0	Y	N	N
British	Y	50	0	Y	N	7	Spirit	N					9
Delta	Y	10	30	Y	N	N	Turkish	Y	20	0	N	N	SA
Emirates	Y	15	0	Y	N	8	United	Y	15	30	Y	N	N
AirFrance	Y	Y	Y	Y	N	2+6	VriginAmerica	Y	10	25	Y	N	N
Hawaiian	Y	10	25	Y	N	N	Hotel						
VirginAtlantic	Y	25	22	Y	N	10	BestWestern	W	0	0	Y	Y	SA
WestJet	Y	0	50	N	N	N	Choice	N					N
AirCanada	Y	20	0	Y	N	8	Hilton	W	0	0	Y	N	11
Cathay	Y	17	0	Y	N	N	Hyatt	Y	0	0	Y	Y	N
ANA	N					8	IHG	Y	5	0	N	N	N
Etihad	Y	Y	Y	Y	N	9	Marriott	Y	0	0	Y	Y	N
Frontier	Y	25	5.6	Y	N	8	Radisson	W	0	0	N	Y	SA
JetBlue	Y	12.5	0	Y	N	7	Starwood	Y	0	0	Y	Y	N
KoreanAir	N					5	Wyndham	N					N

Notes. This table reports the narrative features of loyalty token transactions. *Tx* indicates whether token transactions are allowed, with "W" indicating that transfers are confined to family groups. The following columns are applicable only when transfers are allowed. *$ Cost / 10^3 Units* documents the transaction fee in US dollars for every 1,000 airline miles or loyalty points (the unit name varies across companies), with "Y" denoting unreported or elaborate fee structures. *$ Fee / Tx* shows the fixed processing charge per transaction, with "Y" denoting unreported or elaborate fee structures. *Limit* shows whether there are further transaction limits, e.g., quantity caps and floors, maximum yearly transactions, or stepped quantity choices. *Paperwork* indicates whether extra processes are required, e.g., forms and telephone calls. *Family Pool* outlines the maximum permissible members in a family pool (or similar group plan) for shared token use. "N" signifies no family pooling, "SA" means unlimited members with the same address, and + denotes additional members with restrictions, such as limiting pooling to children's accounts. All the contents here is collected from official terms and conditions in March 2024.

Transfer restrictions. It is also common to limit the flexibility of token transfer, e.g., by implementing minimum and maximum amounts and stepped transaction volumes, labeled "Y" in Column (4). This makes it almost impossible for consumers to achieve optimal holding positions through transactions as in a frictionless secondary market. An annual cap on total transaction times or volumes is also common, limiting the market size. Some projects also require additional paperwork (e.g., submitting applications and awaiting approval), labeled "Y" in Column (5), which indicates that transfer can be allowed in special cases, rather than designed to allow tradability.

Sharing within a family pool. Interestingly, family pools are often the highlight of loyalty programs — accounts within a pool can share token accumulation. Given that both appear to be conveying tokens, why is family sharing welcomed when token transfer are largely restricted? The explanation is twofold. The key difference is that people do not earn premia from trading within their families. Within a pool, the reselling option is not valued, since no tokens are actually traded but only accumulated and redeemed more quickly. Therefore, family pooling is equivalent to issuing non-tradable tokens to a collective consumer with a larger demand probability. Recall that in Section 3.6 we show that companies always benefit from increasing p, especially for companies specialized in international routes where p is initially small. We can see from Column (6) that international airlines are more likely to implement family pools and usually set a larger headcount cap. On the other hand, even if there are trading premiums within a family, firms are unable to prohibit them because acquaintances can trade "over-the-counter" based on social connections, just as firms can do little about the gray market. As a result, it is more beneficial to incentivize a collective consumer by allowing family pooling.

5.2. Maintaining High Token Dependence

Our model shows that greater market power helps maintain higher token dependence, which is effectively caused by a decrease in consumer base loss from token issuance. In this context, we use consumer counts as the main independent and examine the relationship to token dependence.[31] Section 5.2.1 summarizes our data. Section 5.2.2 draws a sanity check to corroborate our model settings. Sections 5.2.3 and 5.2.4 document our main regression results in the airline industry and hospitality industry respectively.

5.2.1. Data Summaries and Stylized Facts

We collect data from airlines' and hotels' 10-K financial reports.[32] Despite slight variations in indicators disclosed by companies, airlines and hotels typically generate comparable measurements within their industry. For airlines, we obtain the direct measure of token dependence defined in equation (8), $\Theta = \frac{\text{Redemption Revenue}}{\text{Operation Revenue}}$. Note that Θ is a share concept relative to total revenue that reflects how many "rebates" are attached to each unit of sales, and by definition is independent of business scale as the outcome of firm decisions.

As Table 2 summarizes, across our (unbalanced) Company×Year panel data, the token dependence of airlines ranges from 0.65% to 13.81% with a standard deviation of 3.16%. The size of the consumer base, measured by the annual number of passengers (in millions), *#passengers*, has a mean of 77.79 with a standard deviation of 58.77. They both suggest notable variations across the corporations and years, as well as the relative loyalty liability. Additional variables and data in the hotel panel are also summarized and will be introduced later.

Stylized business status. The introduction section has listed facts about the loyalty business model over the world. Here we focus on our sample to draw out further stylized facts.

Aviation industry originated and led the loyalty program since the first airline loyalty program began in 1981. To show its recent development trend, we construct a balanced panel

[31] Market power might come from other strengths: product segmentation, technological barriers, etc. We choose the dimension that closely relates to our model. The comprehensive measurement of market power is beyond our focus.

[32] Companies are required to disclose the operation status of their loyalty programs in ITEM 7, i.e., management's discussion and analysis of financial condition and results of operations. The details for data collection and preprocessing are provided in supplementary materials.

Table 2. Data Summary

Variables	N	Mean	S.D.	Min	25%	50%	75%	Max
Panel A: Airline								
RR, Redemption Revenue ($\times 10^9$\$)	69	1.08	1.02	0.05	0.29	0.61	2.18	3.37
OR, Operation Revenue ($\times 10^9$\$)	69	21.45	13.83	0.84	8.43	21.96	31.51	47.01
Token Dependence, *RR*/*OR* (%)	69	5.05	3.16	0.65	2.05	5.22	6.50	13.81
# Passengers ($\times 10^6$)	64	77.79	58.77	3.36	33.59	55.08	134.26	215.00
Price per Mile (\$)	69	0.23	0.10	0.15	0.16	0.19	0.27	0.77
Average Flight Distance ($\times 10^3 Miles$)	64	1.35	0.41	0.74	1.08	1.25	1.49	2.68
Panel B: Hotel								
LL, Loyalty Liability ($\times 10^9$\$)	77	1.02	1.53	0.01	0.07	0.23	1.43	6.47
HR, Hotel Revenue ($\times 10^9$\$)	75	4.82	5.17	0.20	1.25	2.43	6.39	20.97
Relative Loyalty Liability, *LL*/*HR* (%)	75	15.12	13.20	0.98	5.55	12.02	17.94	59.32
# Guests ($\times 10^6$)	57	0.43	0.24	0.01	0.31	0.45	0.56	1.01
# Properties ($\times 10^3$)	56	4.92	2.53	0.08	3.79	5.29	6.80	9.28
Price per Night (\$)	74	106.67	73.26	22.84	47.38	98.17	131.75	320.38

Notes. This table summarizes our variables of interest from corporation 10-K reports. The mean, standard deviation, and quintiles are calculated unconditionally from the Corporation×Year panel.

from 2017 to 2021, including ten major airlines that continuously report loyalty revenues.[33] Among this group, there is a total of \$12.46 billion in redeemed value, making up 5.17% of annual operations revenues, and leaving \$28.59 billion unredeemed in 2017. In 2018 (2019), the indicators reach \$13.66 (\$14.33) billion, 5.32% (5.47%), and \$29.52 (\$30.52). In 2021 after the reductions of the pandemic, the airlines' aggregate redemption revenue is reduced to \$9.65 billion; however, it was less affected than the overall business and thus these programs accounted for 6.13% of annual operations revenues and had remarkable outstanding unredeemed value: \$36.38 billion. Loyalty tokens to some extent provide a lifeline in times of external shocks.

Loyalty programs are also widely recognized and adopted by the hospitality industry, especially among leading hotel groups (e.g., Marriott, Hilton, and Hyatt). These groups are also among the earliest ones that implemented loyalty programs before 2016. In addition, because the pandemic caused a pause in operations, all hotels experienced a higher backlog of unredeemed loyalty liabilities in 2020-21 (e.g., in 2020 Marriott obtained a loyalty liability of 60% to its operation revenue). Considering that the loyalty liability actually implies liquidity front loaded to corporations, it somewhat mitigated the impact of the huge turnover decline under the external pandemic shock.

5.2.2. Sanity Check of Token Dependence

We start with a sanity check for a deeper understanding of the objective determinants of token dependence. Our model setup features that faster redemption (smaller M) aligns with larger token dependence Θ. Empirical data confirms this assumption holds as the positive correlation shown in Figure 6, where redemption speed is obtained by dividing the annual

[33] The balanced panel drops the samples with missing data on redemption or operations returns in any year from 2017 to 2021, and finally includes Air France, Alaska Airlines, American Airlines, Delta Air Lines, Emirates, Hawaiian Airlines, International Airlines Group (IAG), JetBlue, Southwest Airlines, and United Airlines.

redemption revenue by the outstanding stock (unredeemed liabilities) at the end of last year.[34] This appears to be a general fact that applies to not only comparisons among corporations but also to pooling the years.

In particular, during the pandemic (2020-21), the global airline industry suffered a huge blow. The sudden drop in business led to a natural slowdown in redemption speed. However, the positive correlation holds with an even sharper slope: token dependence did not decrease dramatically over the period, as the pandemic simultaneously shut down consumption and redemption. This gives us a clearer understanding of token dependence: it is not a size-like concept and an exogenous consumption shock should not change it unless the shock affects consumers' choice between using fiat money and tokens.

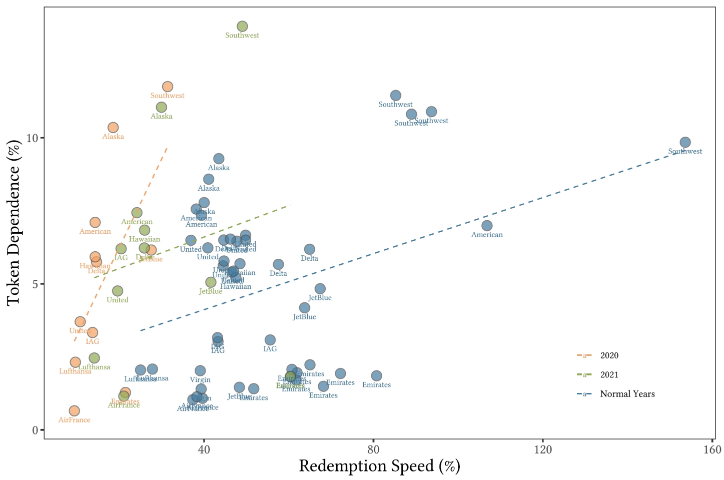

Figure 6. Redemption Speed and Token Dependence

This figure visualizes the sanity check based on the Corporation×Year panel, where the y-axis is token dependence, and the x-axis presents redemption speed, i.e., the annual redemption revenue divided by outstanding stock in the previous year (%). We show the pandemic years (2020 and 2021) with different colors and draw the linear fits separately.

5.2.3. The Token Dependence of Airlines

We formally test that an airline with great consumer counts maintains a high token dependence. We estimate OLS regressions with the following specification:

$$\Theta_{i,t} = \alpha_0 + \alpha \#Passengers_{i,t} + X'_{i,t}\gamma + b_i + d_t + \epsilon_{i,t}, \tag{14}$$

where the coefficient of interest is α, $X'_{i,t}$ includes potential control variables, and b_i and d_t are airline and year fixed effects, respectively.

Table 3, Column (1) shows that there is a strong positive unconditional correlation between business scale and token dependence, where the coefficient is 0.016 ($s.e. = 0.007$) — a one standard deviation greater number of more passengers (58.77 million) is associated with a 0.94 percentage point higher token dependence, explaining about 30% of the standard de-

[34]Note that the redemption speed can exceed 1 (100%), since tokens can be used within the year they are issued, and these are not counted in the previous outstanding stock.

viation of token dependence.[35] Columns (2) and (3) confirm the results with year and airline fixed effects. Notably, with airline fixed effects, R-squared reaches 0.956, suggesting a large share of explanatory power from unobserved time-invariant determinants, including many other dimensions of market power. Column (4) controls other potential correlates of token dependence, including the pandemic-period dummy, average ticket price per mile, and the dummy of whether the airline belongs to an alliance,[36] reflecting heterogeneous implementations of loyalty programs across market segments (cheap or luxury) and alliances. The multivariate regression has a greater explanatory power than Column (1), whereas the coefficient of *#Passengers* equals 0.023 ($s.e. = 0.009$), still economically and statistically significant.

There are still concerns in interpreting the above relationship as causal: unobservable shocks may be correlated with business scale and also affect the calculation of token dependence. Imagine an airline experiences a positive productivity shock and gains a larger market share. This positive shock for airline i in year t attracts more passengers (a larger $Passengers_{i,t}$). Meanwhile, the positive shock leads to higher operation revenue thus mechanically reduces token dependence $\Theta_{i,t}$ as redemption mainly comes from the loyalty tokens accrued in previous years. As mentioned in Section 3.6, an expansion in market share makes revenue from token redemption relatively less important relative to fiat-money sales, reducing token dependence. On the other hand, a negative productivity shock does not increase token dependence correspondingly, as the still-existing users continue to hold their tokens. Therefore, the coefficient α could be underestimated under the OLS specification (14). This also explains why the coefficient in Column (3) is lower than (2) after including airline fixed effects.

We attempt to address these concerns by introducing the average flight distance as an instrument.[37] Average flight distance reflects the airline's (proactive or reactive) decisions in route operations, i.e., whether to focus on short-haul routes with a larger passenger base or longer routes.[38] Therefore, the distance relates to the number of passengers. Meanwhile, the distance captures more cross-sectional variations, e.g., geographical conditions, which is less susceptible to the above endogeneity concern from time-series fluctuations. Also, passenger

[35] We report the standard errors two-way clustered at year and airline×period (normal or pandemic) levels. In particular, the latter clusters the samples from an airline into normal and pandemic clusters. The aggregate shock of the pandemic is not the treatment of interest, but brings about observations significantly different from normal periods, in which the difference vastly outweighs other sample variations. If they were treated as correlated as in the conventional cluster formula, it would result in an unnecessarily conservative confidence interval. A more targeted approach is to calculate the causal cluster variance as suggested by Abadie et al. (2023).

[36] Airlines disclose in 10-K reports the revenue passenger mile (RPM), multiplying the number of paying passengers by the distance traveled. We divide the operation revenue (approximately the gross revenue paid by passengers) by RPM to obtain the average ticket price per mile. The alliance dummy equals one if the airline belongs to one of the three major alliances, i.e., Star Alliance, Oneworld, or Skyteam.

[37] We use the revenue passenger mile (RPM) divided by aggregate passengers to obtain the average flight distance of a passenger. In precise, we denote the number of passengers of flight j as n_j, and the flight distance is m_j. Then the average flight distance reads $RPM\ /\ \#Passengers = (\sum_j n_j m_j)/(\sum_j n_j) \equiv \bar{m}$.

[38] Longer routes may have other advantages. More importantly, airlines may not be able to choose freely due to geographical limitations. We do not study the optimal route choice problem, but rather focus on whether the larger number of consumers resulting from a particular route choice has an effect on token dependence.

changes attributable to airlines' route-distance decisions are likely to be structural, independent of productivity shocks. Therefore, the average flight distance is a potential plausible instrumental variable.

Table 3. Business Scale and Token Dependence of Airlines

Dependent:	Token Dependence, $\Theta^R = \frac{\text{Redemption Revenue}}{\text{Operation Revenue}}$ (%)					
	OLS				2SLS	
Stage:					First	Second
	(1)	(2)	(3)	(4)	(5)	(6)
#Passengers	0.016**	0.023**	0.019*	0.023**		0.037**
	(0.007)	(0.009)	(0.009)	(0.009)		(0.016)
Price per Mile				-20.753***	-29.924	-19.877***
				(5.697)	(50.740)	(5.251)
Alliance Member				-0.795	39.905**	-1.449
				(1.121)	(17.497)	(1.091)
Pandemic				3.597***		
				(0.743)		
Average Flight Distance					-61.636***	
					(11.834)	
Year FE		Yes	Yes		Yes	Yes
Airline FE			Yes			
Observations	64	64	64	64	64	64
R^2	0.081	0.163	0.956	0.532	0.536	
Weak Inst. F-stat					27.1	
Wu-Hausman p-value						0.225

Notes. The dependent and main independent are token dependence (%) and the number of passengers (million), *#Passengers* respectively. Columns (1)-(4) report OLS estimates with year and airline fixed effects and controls including price per mile, the alliance member dummy, and the pandemic dummy. Columns (5)-(6) show the 2SLS specification using average flight distance (10^3 miles) as the instrument. F stats of the Kleibergen-Paap test for weak instruments and p values for the Wu-Hausman endogeneity test are reported. Standard errors two-way clustered at year and airline × periods (normal or pandemic) are in parentheses.

Table 3, Columns (5)-(6) report the 2SLS estimates with the instrument variable, the average flight distance. The first-stage estimation displays a significant predictability with a Kleibergen-Paap F-stat of 27.1, suggesting the distance to be a strong instrument of the number of passengers. As Column (6) shows, the coefficient of *#Passengers* (0.037) is almost double of the OLS estimates (around 0.02). This confirms our conjecture, i.e., the omitted determinant problem causes an underestimation in OLS approaches. A one standard deviation greater number of passengers is associated with a 2.17 percentage point higher token dependence, roughly explaining 69% standard deviation of the token dependence. Overall, Table 3 indicates that airlines with larger consumer bases choose to maintain higher token dependence in their loyalty program design.

5.2.4. The Token Dependence of Hotels

A similar empirical analysis is applied in the hospitality industry. We use the annual number of guests to proxy for the business scale of hotels (one counted guest per occupied room), as summarized in Table 2, and use the number of properties as an instrument. The logic is similar (even simpler): the number of properties reflects a hotel's capacity and goal to host a larger consumer base. We also control for price per room night and its heterogeneous effect during pandemic. The main difference is the dependent variable. Hotels do not report redemption revenues, possibly due to different disclosure requirements. Alternatively, we use an indirect measure, relative loyalty liability, $RLL = \frac{\text{Loyalty Liability}}{\text{Operation Revenue}}$. While token dependence accounts for realized revenue, *RLL* accounts for the present value of unredeemed loyalty tokens. *RLL* can be used similarly to compare the importance of the loyalty program to the airlines in the cross section.

Columns (1)-(4) of Table 4 report the OLS and 2SLS estimates. In particular, Column (4) suggests that a one standard deviation more guests (0.24 million) is associated with 7.82 percentage points higher relative loyalty liabilities, roughly explaining 59% of its standard deviation. These results are based on pooling regressions (or with year fixed effects). The explanatory power is mainly from the cross-sectional variations in *#Guests*, which captures the comparison of market power. However, a hotel's change in *#Guests* also technically affects *RLL*: a guest's redemption would reduce loyalty liabilities but add to token dependence. Column (5) confirms this conjecture by regressing *RLL* on the change rate of *#Guests* and documenting a negative relationship. With hotel fixed effects in Column (6), we also obtain a negative coefficient of *#Guests*, as the variation effectively used in the estimation is relatively more from temporal changes within hotels. Despite the unavoidable use of information on time-series changes, the instrument *#Properties* captures the cross-sectional variation more, as its within-group changes are significantly smaller than across-group variations. Then, in Columns (7)-(8), we still obtain a positive coefficient of *#Guests* using a 2SLS estimation. Therefore, despite challenges from the indirect measurement *RLL*, we still find robust evidence that a larger consumer base helps maintain a higher token dependence.

Relative loyalty liability (*RLL*) provides additional corroboration to our model. It can be seen as a transient state, abstracting away the redemption speed. Then if the redemption or consumption flow is shocked, *RLL* fails to one-to-one correspond to token dependence. For example, as documented in Section 5.2.1, during the pandemic periods, airlines suffered from stoppage of routes. Although the high token dependence held up (even increased), it was a combination of simultaneous crashes in total revenue and redemption — there would have been more loyalty liabilities converted to revenue at the regular redemption speed. As a result, we saw a remarkable increase in outstanding loyalty stocks. Similarly, for hotels, token dependence is overestimated by *RLL* in pandemic years.

Overall, we find evidence in airline and hospitality industries that greater market power enables corporations to maintain higher token dependence, thus better utilizing loyalty token issuance for liquidity creation. While we make efforts to separate the business scale from other potential determinants, some unobserved factors are also in the scope of market power in our theory. For example, issuers can utilize technology barriers and monopolize a submarket, and thus maintain a greater market power. Therefore, representing market power by business scale still economically underestimates its determination of token dependence. Our estimate

Table 4. Relative Loyalty Program Scale of Hotels

Dependent:	Relative Loyalty Liability, $\frac{\text{Loyalty Liability}}{\text{Hotel Revenue}}$ (%)							
	Impact of Pooling Variations of #Guests				Impact of Longitudinal Change in #Guests			
	OLS		2SLS		OLS		2SLS	
Stage:			First	Second			First	Second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
#Guests	19.164***	28.152***		32.586***		-21.086		40.623**
	(5.327)	(5.266)		(6.968)		(15.116)		(17.614)
Revenue per Room Night		0.115***	0.003**	0.116**	0.213**	-0.141	0.000	-0.194
		(0.032)	(0.001)	(0.038)	(0.077)	(0.085)	(0.001)	(0.139)
Revenue per Room Night × Pandemic		0.889**	0.004**	0.882**	0.422***	0.327**	-0.001	0.375*
		(0.303)	(0.001)	(0.305)	(0.070)	(0.136)	(0.001)	(0.184)
#Properties			0.083***				0.076***	
			(0.016)				(0.019)	
Guest Growth					-15.925***			
					(3.825)			
Year FE		Yes	Yes	Yes		Yes	Yes	Yes
Hotel FE						Yes	Yes	Yes
Observations	52	52	52	52	50	52	52	52
R^2	0.098	0.696	0.764		0.426	0.937	0.933	
Weak Inst. F-stat				26.8				16.3
Wu-Hausman, p-value				0.264				0.0004

Notes. The dependent and main independent variables are relative loyalty liability (RLL) (%) and the number of guests (million) (*#Guests*). Column (1) reports OLS estimation of the univariate model. Column (2) controls revenue per room night and its interaction with the pandemic dummy and year fixed effects. Columns (3)-(4) show the 2SLS estimation of the same model as in Column (2), using the number of properties as an instrument. Column (5) shows the unconditional relationship between *RLL* and the change rate of *#Guests*. Columns (6)-(8) report the OLS and 2SLS estimation of the same model as in Column (2) with additional hotel fixed effects. F stats of the Kleibergen-Paap test for weak instruments and p values for the Wu-Hausman endogeneity test are reported. Standard errors clustered by year and hotel×periods are in parentheses.

can be regarded as a lower bound, whereas a more comprehensive analysis should rely on a general metric of market power. In addition, our empirical tests are based on disclosures in 10-K reports, where token revenues are not accounted for until redemption is realized. However, our model emphasizes tokens' crucial role in liquidity creation before being redeemed, which is not fully presented in current related accounting standards.

5.3. Redemption Probability Manipulation

To focus on the central mechanism of loyalty tokens, our model regards demand probability p as an exogenous feature of the located market. In particular, the baseline model implicitly assumes that people have the same demand probabilities p for fiat-money consumption and redemption. In practice, highly user-friendly loyalty token systems have been observed to increase consumers' demand probability p, which can be viewed as an extension of market power through the loyalty business model.[39] Token designs can also increase p to elicit a higher token price by affecting consumers' perception of token redemption convenience, forming alliances across continents, and expanding time-series bundling benefits by setting an expiration date, as some consumers might not be able to accumulate enough tokens before

[39]From the perspective of artificially increasing switching costs (Banerjee, 1987), loyalty programs facilitate repeat consumption.

the expiration date.

Digitalization and Convenience If the tokens were inconvenient to use (e.g., hard to understand, complicated to redeem), consumers would perceive a redemption probability lower than p, thereby decreasing the valuation of tokens. Then, greater market power would be required to fill this discount loss in implementing the loyalty program. In practice, we see large issuers always applying technologies to their loyalty programs as quickly as possible to increase the ease of use. For example, as Table 5 shows, airlines and hotel groups tend to launch mobile apps and enable systems of token management and usage earlier since the iOS App Store launched in July 2008. This becomes apparently standard by the 2020s.

Table 5. Redemption Demand Probabilities Summary

Corporation	Expiry	App Release	Alliance / # Codeshare Partners	Corporation	Expiry	App Release	Alliance / # Codeshare Partners
Airline				Lufthansa	3C	2008-12-08	Star Alliance
Alaska	2C	2010-02-20	Oneworld	Qatar	3C	2012-12-02	Oneworld
American	2C	2010-07-26	Oneworld	Singapore	3	2012-04-06	Star Alliance
Avianca	2	2013-01-18	Star Alliance	Southwest	N	2009-12-18	0
British	3C	2008-07-11	Oneworld	Spirit	1C	2018-11-15	3
Delta	N	2010-09-01	Skyteam	Turkish	3	2017-09-18	Star Alliance
Emirates	3	2014-09-11	15	United	N	2011-10-20	Star Alliance
AirFrance	2C	2010-09-17	Skyteam	VriginAmerica	1.5C	2016-09-23	Oneworld
Hawaiian	N	2010-11-09	5	Hotel			
VirginAtlantic	N	2010-01-31	Skyteam	BestWestern	N	2009-12-17	
WestJet	N	2014-05-05	20	Choice	1.5C	2012-04-06	
AirCanada	1.5C	2009-08-20	Star Alliance	Hilton	2C	2013-04-19	
Cathay	1.5C	2009-02-23	Oneworld	Hyatt	2C	2011-10-31	
ANA	3	2010-07-26	Star Alliance	IHG	1C	2010-04-22	
Etihad	1.5C	2016-04-11	24	Marriott	2C	2011-08-05	
Frontier	1C	2015-10-15	3	Radisson	2C	2019-08-08	
JetBlue	N	2012-02-03	7	Starwood	1C	2009-06-05	
KoreanAir	10	2010-07-28	Skyteam	Wyndham	1.5C	2013-12-31	

Notes. The table aggregates key features of loyalty token design pertinent to the probabilities of token redemption. *Expiry* documents the token's validity period, where "N" indicates that tokens never expire, a numeric value represents years of validity, and "C" indicates expiration dates only apply if accounts are inactive. *App Release* records the initial launch date of IOS applications for booking and (potentially) redemption. *Alliance / # Codeshare Partners* is only applicable for airlines and identifies the alliance membership or alternatively lists the number of code-sharing partners. Token accumulation is typically shareable within the alliance or partners. *Expiry* and *Alliance* are collected from official terms and conditions in March 2024.

Alliance Formation. A direct way to boost redemption probability is to increase the redeemable services. In this way, issuers sell tokens with broader uses without expanding their main business (which would be remarkably costly). Here we focus on a prime example in airlines: forming alliances or entering into code-sharing cooperation, as shown in Table 5. After flying on one airline, consumers can credit tokens in their accounts with partner carriers, or directly interoperate within the alliance. Therefore, consumers have a higher demand probability of redemption than they would for the original airline's flights, and would pay more for the flight tickets.[40]

Figure 7 shows the airline networks of three major alliances, Star Alliance, Oneworld, and

[40]From another perspective, the original airline earns a premium for attracting traffic to its partner.

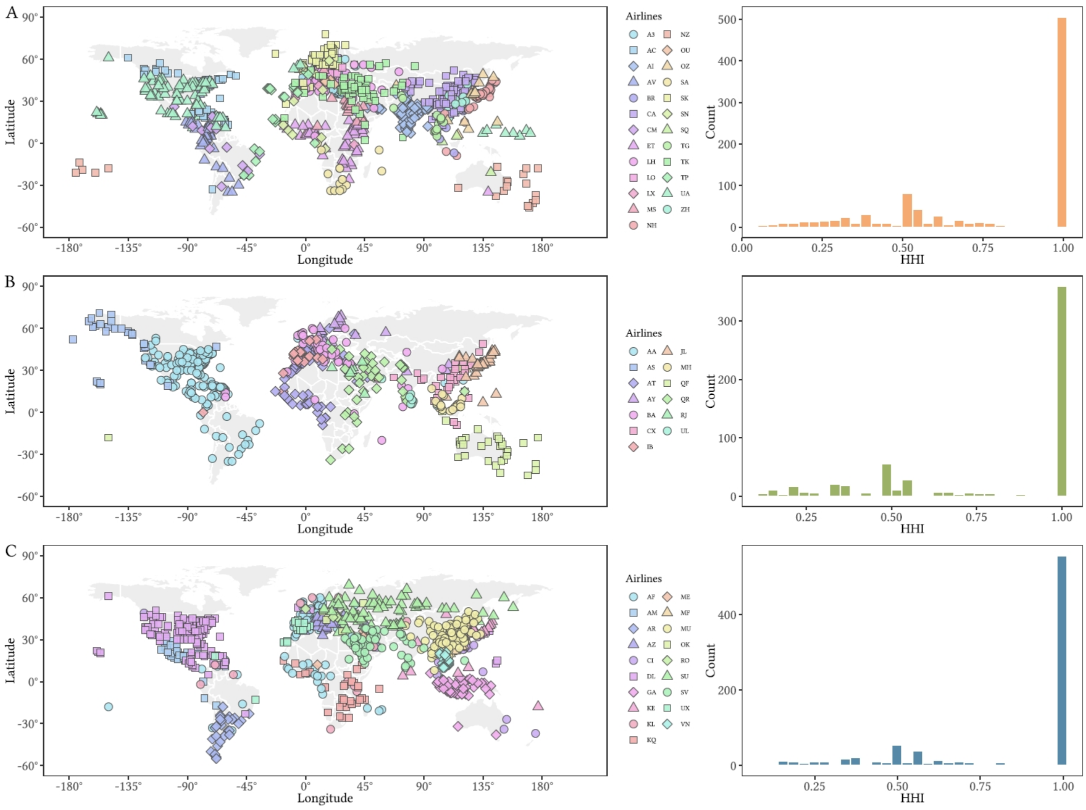

Figure 7. Global Airport Networks of the Alliances

Panels A, B, and C visualize the global airport distribution of Star Alliance, Oneworld, and Skyteam, respectively. In each panel, the map depicts the airports covered by the airline alliance's routes. Each airport is shown in different colors and shapes according to the "dominant member" within the alliance who operates the most routes at that airport. The within-alliance Herfindahl-Hirschman Index (HHI) of operating routes at each airport is also calculated and aggregated in the histogram. Data source: OpenFlights Airports Database.

Skyteam. We define the "dominant member" of an airport as the alliance member who runs the most routes at that airport. Then for each alliance, the map points out all the accessible airports, i.e., those covered by the alliance's routes, in different colors and shapes according to their dominant members. Furthermore, we calculate the Herfindahl-Hirschman Index (HHI) of routes in operation at each airport, and summarize the distribution of the accessible airports' HHI, plotted in the histogram.

There are two findings. First, each of the three alliances has formed a global airline network to take advantage of the global demand for flights to boost their redemption probabilities. Second, the alliance usually does not include two members from the same region to avoid competition within the alliance. Essentially, the worldwide airport network consists of air hubs and branch airports. As histograms show, a large share of airports are connected by one single member in the alliance, making the HHI equal to one. From the airline's perspective, it is better not to cooperate if it does not effectively increase the redemption demand. As documented in Table 5, Southwest Airlines constitutes a special example: as an airline specializing in short-haul travels, it faces a unique user group and has a monopoly. As a result, building a global network is not a top priority. This logic also applies to those targeting specific user bases, e.g., low-cost and regional airlines. In general, both the alliance and the airline favor partners who are able to increase redemption functionality without adding competition.

Loyalty point expiration. As we discussed in Section 4.2, token redemption might suffer from commitment issues. One particular example is that airlines can further exploit time-series bundling and consumer inattention by setting a token expiration term, as consumers might not have sufficient points for redemption, or forget to redeem. Expired tokens can effectively reduce costs associated with service delivery by not committing to a permanent redemption policy. However, the expiration date design might dampen the token demand and lower the willingness to pay for the flight tickets. Therefore, the expiration setup is only profitable when most tokens are eventually unredeemed such that the "free lunch" outweighs the drawback of discounted token valuation. For this reason, we see that international-focused airlines appear to favor expiration terms (e.g., Avianca, Emirates Airlines, ANA, Singapore Airlines, and Turkish Airlines) as there are fewer competitors on the same route. Otherwise, airlines commit to never-expiry, labeled by "N" in Table 5. In addition, with possible consumer exits, issuers may adopt a compromise solution, i.e., expire tokens conditional on the holder being inactive for x years, labeled as "xC" in Table 5. This picks up the 'free lunch' from inactive consumers without discounting the valuation of tokens by frequent consumers.[41]

6. Conclusion

Our paper studies loyalty token issuance as a financing mechanism and highlights the economic trade-offs when token issuance is bundled with consumption. We show that the issuer has an incentive to limit the token secondary market as tradability eliminates the benefit of time-series bundling, and this loss strictly dominates the welfare gain by selling tokens at a higher price. Loyalty tokens can be viewed as a privileged usage of a financial toolkit to leverage the issuer's market power: more market dominance enables higher token dependence; if all issuers are under perfect competition, loyalty token issuance becomes impossible. These predictions are well supported by analyzing airline mileage and hotel point data.

Our result that firms, especially large ones with significant monopoly power, have strong incentives make their loyalty tokens non-tradable, is clearly relevant to regulators concerned with placing boundaries on their use, beyond ensuring that advertising claims are transparent and accurate, for example on ease of use and transferability. Of course, the results here assume that loyalty tokens are not widely used as a transactions currency for a much broader class of goods, albeit it is not clear how loyalty tokens could compete for such use with fiat and cryptocurrencies. If, of course, loyalty tokens began to incorporate blockchain design features such that outside use became much more significant, the seigniorage benefits could outweigh the benefits of non-tradability that we have demonstrated. In this case, the issue of optimal

[41]This implies that consumer heterogeneity matters in expiration designs. See Sun and Zhang (2019), who developed a model to rationalize the trade-off between short and long expiration terms.

regulation of loyalty tokens would need to be re-visited, but for the moment, this does not seem a first-order issue.

Appendix

A. Omitted Proofs

Proof and Intuition of Lemma 1. (i) First, it is obvious that at any time, no one would hold more than M tokens. Otherwise, the user is not able to fully redeem them in one period. Then she should prefer to buy these excess tokens in the next period with a lower present value.

Consider that the user holds m tokens, $0 \leq m < M$. When $M = 1$, the claim naturally holds, that is, any user holds either zero or $M = 1$ tokens.

When $M \geq 2$, we prove that for any period t_1 and any consumer with m_1 tokens before trading in t_1 and plans to hold m_2 tokens after trading in $(t_1 + 1)$, $0 \leq m_1, m_2 \leq M$, so the consumer prefers to hold zero than m tokens after trading in t_1, $0 \leq m < M$. Denote the secondary market price as $\hat{P}$. For any possible (m_1, m_2), the cost of holding m tokens after trading in t_1 reads

$$\underbrace{(m - m_1)\hat{P}}_{\text{trade in } t_1} + \underbrace{p\left[P_M^T + \beta(m_2 - m - 1)\hat{P}\right]}_{\text{trade in } t_2 \text{ if shocked in } t_1} + \underbrace{(1-p)\beta(m_2 - m)\hat{P}}_{\text{trade in } t_2 \text{ otherwise}}, \tag{A.1}$$

where the coefficient of m is $(1 - \beta)\hat{P} > 0$; note that $m < M$ tokens are not enough for one redemption in t_1. Therefore, a larger m increases the cost of any possible holding strategies. That is, holding $m \in (0, M)$ tokens after trade is strictly dominated by holding zero tokens after trade.

(ii) Note that each user can only receive one token from purchase in one period. Combined with the above implication, we find that at the beginning of each period, there are three types of users who hold either 0, 1, or M tokens, labeled u_0, u_1, and u_M. After trade, there are only two types, u_0 and u_M. Then the market clearing condition should ensure that u_1 is indifferent between becoming u_M and u_0. That is, when purchasing and receiving a new token, users are indifferent to either selling it or buying extra $M - 1$ tokens:

$$\underbrace{P_M^T - \beta\hat{P}}_{\substack{\text{pay for product} \\ \text{and resell token}}} + \underbrace{a(P_M^T - \beta\hat{P})}_{\substack{\text{present value of} \\ \text{next consumption} \\ \text{excluding the value of token}}} = \underbrace{P_M^T + \beta(M-1)\hat{P}.}_{\substack{\text{pay for product and buy token} \\ \text{(and redeem for next consumption)}}} \tag{A.2}$$

This solves a unique price $\hat{P}$,

$$\hat{P} = \frac{p + (1-p)a}{M + a} P_M^T. \tag{A.3}$$

Essentially, $\hat{P}$ simultaneously ensures that any type of user is indifferent between holding M tokens and zero tokens:

$$(M - m)\hat{P} = -m\hat{P} + (p + (1-p)a)(P_M^T - \beta\hat{P}), \quad m \in \{0, 1, M\}. \tag{A.4}$$

Therefore, $\hat{P}$ is the equilibrium secondary market price of one token.

Proof of Proposition 1. (i) Calculate the derivatives directly:

$$\frac{\mathrm{d}P_M^{NT}}{\mathrm{d}p} = \frac{a^{M-1}(1-a)\left(M - a\frac{1-a^M}{1-a}\right)}{(1-a^M)^2} \frac{\beta(1-\beta)}{(1-\beta+\beta p)^2} = \frac{a^{M-1}(1-a)\beta(1-\beta)\sum_{i=1}^M(1-a^i)}{(1-a^M)^2(1-\beta+\beta p)^2} > 0. \tag{A.5}$$

$$\frac{\mathrm{d}P_M^T}{\mathrm{d}p} = \frac{1}{M}\frac{\beta(1-\beta)}{(1-\beta+\beta p)^2} > 0. \tag{A.6}$$

(ii)

$$\frac{\mathrm{d}P_M^{NT}}{\mathrm{d}M} = \frac{(1-a)a^M \log(a)}{(1-a^M)^2} < 0; \quad \frac{\mathrm{d}P_M^T}{\mathrm{d}M} = -\frac{a}{M^2} < 0. \tag{A.7}$$

(iii) $P_M^{NT} < P_M^T$, i.e.,

$$\frac{1-a^{M+1}}{1-a^M} \leq 1 + \frac{a}{M} \underset{Rearrange}{\iff} \frac{a(1-a^M)}{1-a} \geq Ma^M. \tag{A.8}$$

Note that $a < 1$, $\frac{a(1-a^M)}{1-a} = \sum_{i=1}^M a^i \geq \sum_{i=1}^M a^M = Ma^M$. The equals sign is obtained if and only if $M = 1$ or $M \to \infty$.

Proof of Lemma 2. Treat Rev_M^{NT} as a continuous function of $M \in [1, \infty)$.

$$\frac{\mathrm{d}Rev_M^{NT}}{\mathrm{d}M} = \frac{p\left[(1-a)\log(a)a^M(1-a^{*M})(1-a^{*M+1}) - (1-a^*)\log(a^*)a^{*M}(1-a^M)(1-a^{M+1})\right]}{(1-\beta^*)(1-a^M)^2(1-a^{*M+1})^2}. \tag{A.9}$$

Note that $a, a^* \in (0,1)$, and $\beta^* < \beta \Leftrightarrow a^* < a$. Define a temporary function w.r.t. $a \in (0,1)$, $g(a) = \frac{a^M(1-a)\log(a)}{(1-a^M)(1-a^{M+1})}$. Then $\frac{\mathrm{d}Rev_M^{NT}}{\mathrm{d}M} < 0$ for any $M \geq 1$ is equivalent to $g(a) < g(a^*)$ for any $M \geq 1$, for which the sufficient condition is $g'(a) < 0$, $\forall a \in (0,1)$. Further,

$$\begin{aligned} g'(a) &= \frac{a^{M-1}(1-a)\log(a)}{(1-a^M)^3(1-a^{M+1})^2}\left[\frac{M(1-a^{2M+1})}{1-a^M} - \frac{a(1-a^M)}{1-a} + \frac{1-a^{M+1}}{\log(a)}\right] \\ &\equiv \frac{a^{M-1}(1-a)\log(a)}{(1-a^M)^3(1-a^{M+1})^2}G(a). \end{aligned} \tag{A.10}$$

$$G(a)=\frac{M(1-a^{2M+1})}{1-a^M}-\sum_{i=1}^{M}a^i+\frac{1-a^{M+1}}{\log(a)} \underset{a^i<a}{>} M\left(\frac{1-a^{2M+1}}{1-a^M}-a\right)+\frac{1-a^{M+1}}{\log(a)}$$
$$=M\left(\frac{1-a}{1-a^M}+a^{M+1}\right)+\frac{1-a^{M+1}}{\log(a)} \underbrace{\geq}_{\text{F increases in M}} 1+\frac{1-a^2+a^2\log(a)}{\log(a)} \tag{A.11}$$
$$\underbrace{>}_{\left(\frac{1-a^2+a^2\log(a)}{\log(a)}\right)'=\frac{a\log(a)-1/a}{(\log(a))^2}<0} 1+\lim_{a\to 1^-}\frac{1-a^2+a^2\log(a)}{\log(a)}=0.$$

Therefore, $G(a)>0$. Recall the formula of $g'(a)$ above; there is only one negative multiplied term, $\log(a)$. Therefore, $g'(a)<0$, and therefore $\frac{\mathrm{d}Rev_M^{NT}}{\mathrm{d}M}<0$.

Consider Rev_M^T. Since $a^*<a$,

$$\frac{\mathrm{d}Rev_M^T}{\mathrm{d}M}=\frac{p}{1-\beta^*}\frac{a^*-a}{(M+a^*)^2}<0. \tag{A.12}$$

Proof of Proposition 2.

Lemma 3. ***Tradable and non-tradable tokens at the same price.***

$\forall M\geq 1$, there exists a unique non-tradable issuance strategy $m\geq 1$, s.t. $P_m^{NT}=P_M^T$.

Proof. $1+\frac{a}{M}=P_M^T=P_m^{NT}=\frac{1-a^{m+1}}{1-a^m} \Rightarrow M=\frac{a}{1-a}\left(\frac{1}{a^m}-1\right)$. Note that the R.H.S. increases in m and, when $m=1$, $\frac{a}{1-a}\left(\frac{1}{a^m}-1\right)=1$, $\lim_{m\to\infty}\frac{a}{1-a}\left(\frac{1}{a^m}-1\right)=\infty$. Therefore, $\forall M\geq 1$, there exists a unit $m\in[1,\infty)$ s.t. $P_m^{NT}=P_M^T$. □

We then show that $\forall M>1$, $Rev_M^T<Rev_{m(M)}^{NT}$, where $m(M)$ is the unique corresponding m s.t. $P_m^{NT}=P_M^T=P$. Recall (4) and (7). It is equivalent to prove $Q_M^T<Q_m^{NT}$,

$$\frac{p}{1-\beta^*}\frac{1-a^{*m}}{1-a^{*m+1}}>\frac{p}{1-\beta^*}\frac{M}{M+a^*} \quad\Leftrightarrow\quad M<\frac{a^*}{1-a^*}\left(\frac{1}{a^{*m}}-1\right). \tag{A.13}$$

By Lemma 3, we only need to show that

$$\frac{a^*}{1-a^*}\left(\frac{1}{a^{*m}}-1\right)>\frac{a}{1-a}\left(\frac{1}{a^m}-1\right). \tag{A.14}$$

Treat it as a function of $a\in(0,1)$ and note that $m>1$; by Lemma 3, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}a}\left[\frac{a}{1-a}\left(\frac{1}{a^m}-1\right)\right]=-\frac{(1-a)m-(1-a^m)}{a^m(1-a)^2}<0. \tag{A.15}$$

Then the above inequality holds since $a^*<a$. This implies that $Rev_{M_*^T}^T<Rev_{m(M_*^T)}^{NT}$, i.e., there is always a non-tradable token-issuance plan that generates more revenue than the optimal tradable plan. Therefore, the overall optimal choice is non-tradable. Note that there is no need to satisfy $m(M)=M_*^{NT}$. Also note that we do not require any restrictions on c – we consider all possible prices, even those lower than the monopoly price. Therefore, this proof applies to the general optimization problem discussed in footnote 11.

Proof of Proposition 3. Similarly, treat U_M^x as continuous functions of $M \in [1, \infty)$.

Consider U_M^{NT}. When $c = 0$, $M_*^{NT} = 1$, as Lemma 2 shows. Suppose there exists $0 \le c_1 \le c_2$, s.t., $M_*^{NT}(c_1) > M_*^{NT}(c_2)$. Denote them as m_1 and m_2, respectively. According to Proposition 1 and Lemma 2, $P_{m_2}^{NT} > P_{m_1}^{NT}$ and $Rev_{m_2}^{NT} > Rev_{m_1}^{NT}$. Then by the process of optimization,

$$\begin{cases} U_{m_2}^{NT}(c_2) > U_{m1}^{NT}(c_2) \Rightarrow c_2 < \dfrac{Rev_{m_2}^{NT} - Rev_{m_1}^{NT}}{f(P_{m_2}^{NT}) - f(P_{m_1}^{NT})}, \\ U_{m_2}^{NT}(c_1) < U_{m1}^{NT}(c_1) \Rightarrow c_1 > \dfrac{Rev_{m_2}^{NT} - Rev_{m_1}^{NT}}{f(P_{m_2}^{NT}) - f(P_{m_1}^{NT})}. \end{cases} \tag{A.16}$$

Then $c_1 > c_2$, leading to a contradiction. Therefore, $\forall 0 \le c < c'$, $M_*^{NT}(c) \le M_*^{NT}(c')$.

In the following analysis, we focus on the cases where the optimal plan gains at least zero payoff, which limits the parameter c to satisfy $c \le Q_M^x / f'(P_M^x)$.

Consider U_M^T. The above proof is also applicable for the tradable case. Here we alternatively prove by calculating the exact values of M_*^T:

$$\frac{\mathrm{d}U_M^T}{\mathrm{d}M} = \frac{acf'(P_M^T)}{M^2} - \frac{p(a-a^*)}{(1-\beta^*)(M+a^*)^2} > 0 \Leftrightarrow \left(1 + \frac{a^*}{M}\right)^2 > \frac{p(a-a^*)}{(1-\beta^*)acf'(P_M^T)}$$

$$\Leftrightarrow \begin{cases} M < \dfrac{a^*}{\sqrt{\frac{p(a-a^*)}{(1-\beta^*)acf'(P_M^T)}} - 1}, & c < \dfrac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)}; \\ M > \dfrac{a^*}{1 - \sqrt{\frac{p(a-a^*)}{(1-\beta^*)acf'(P_M^T)}}}, & c \ge \dfrac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)} \quad (M = \infty \text{ when equal}). \end{cases} \tag{A.17}$$

Note that when $c > \frac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)}$,

$$U_1^T = \frac{p}{1-\beta^*}\frac{1+a}{1+a^*} - acf'(P_M^T) < U_\infty^T = \frac{p}{1-\beta^*}, \tag{A.18}$$

then the optimal issuance plan always satisfies $M_*^T = \infty$ when $c > \frac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)}$. When $c < \frac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)}$, $M_*^T = \frac{a^*}{\sqrt{\frac{p(a-a^*)}{(1-\beta^*)acf'(P_M^T)}}-1}$ increases in c. Thus, M_*^T increases in c and tends to infinity when $\frac{p(a-a^*)}{(1-\beta^*)af'(P_M^T)} \to 1$. Combine the two cases, we obtain $\forall 0 \le c < c'$, $M_*^T(c) \le M_*^T(c')$.

Proof of Proposition 4.

Lemma 4. ***A useful inequality.***

$\forall(\beta, \beta^*, p, M)$ that satisfies $0 < \beta^* \le \beta < 1$, $p \in (0,1)$, $M \ge 1$,

$$\frac{1-a^{*M}}{1-a^{*M}\beta^*} - \frac{1-a^*}{1-a^{*M}\beta^*}\frac{(1-p)a^{*M-1}\beta^*}{\widetilde{P_M^{NT}}} \ge \frac{1-a^{*M}}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*}, \tag{A.19}$$

where $\widetilde{P_M^{NT}}$ is defined in equation (10), i.e., $\widetilde{P_M^{NT}} = 1 + \frac{a^{M-1}(1-a)\beta}{1-a^M}$, $a = \frac{\beta p}{1-\beta(1-p)}$, $a^* = \frac{\beta^* p}{1-\beta^*(1-p)}$, and the equal sign holds if and only if $\beta^* = \beta$.

Proof. Note that $\widetilde{P_M^{NT}}$ increases in β, and the L.H.S. of equation (A.19) increases in $\widetilde{P_M^{NT}}$. Therefore, the L.H.S. reaches the minimum when $\beta = \beta^*$, i.e.,

$$\begin{aligned}
\text{L.H.S.} &\geq \frac{1-a^{*M}}{1-a^{*M}\beta^*} - \frac{1-a^*}{1-a^{*M}\beta^*}\frac{(1-p)a^{*M-1}\beta^*(1-a^{*M})}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*} \\
&= \frac{1-a^{*M}}{1-a^{*M}\beta^*}\left[1-\frac{(1-p)a^{*M-1}\beta^*(1-a^*)}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*}\right] = \frac{1-a^{*M}}{1-a^{*M}\beta^*}\frac{1-a^{*M}+pa^{*M-1}\beta^*-pa^{*M}\beta^*}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*} \\
&= \frac{1-a^{*M}}{1-a^{*M}\beta^*}\frac{1}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*}\left[1-a^{*M-1}\beta^*\left(\frac{p}{1-\beta^*(1-p)}-p+\frac{\beta^*p^2}{1-\beta^*(1-p)}\right)\right] \\
&= \frac{1-a^{*M}}{1-a^{*M}\beta^*}\frac{1-a^{*M}\beta^*}{1-a^{*M}+a^{*M-1}\beta^*-a^{*M}\beta^*} = \text{R.H.S. of (A.19)},
\end{aligned}$$

and the equal sign holds only when $\beta = \beta^*$. □

Then we prove Proposition 4 using the same idea as in proving Proposition 2. According to equations (10) and (9), Lemma 3 still holds under the alternative timeline, i.e., $\forall M \geq 1$, $\exists m \geq 1$ s.t. m is unique and satisfies $\widetilde{P_m^{NT}} = \widetilde{P_M^T} = P$. Denote such unique corresponding m as $m(M)$. Then we show that $\forall M > 1$, $\widetilde{Rev_M^T} < \widetilde{Rev_{m(M)}^{NT}}$, i.e.,

$$P\frac{p}{1-\beta^*}\frac{1}{1+\frac{\beta^*}{M}} < P\frac{1}{1-a^{*M}\beta^*}(p+(1-p)a^*)\left[\frac{1-a^{*M}}{1-a^*}-\frac{a^{*M-1}\beta^*(1-p)}{\widetilde{P_M^{NT}}}\right]. \tag{A.20}$$

Note that

$$p+(1-p)a^* = \frac{a^*}{\beta^*} = \frac{p}{1-\beta^*}(1-a^*)$$

and apply Lemma 4 to equation (A.20), we only need to show that

$$\frac{1}{1+\frac{\beta^*}{M}} < \frac{1-a^{*m}}{1-a^{*m}+a^{*m-1}\beta^*-a^{*m}\beta^*} \Leftrightarrow \frac{1}{1+\frac{a^*}{M}\frac{\beta^*}{a^*}} < \frac{1}{1+\frac{a^{*m}(1-a^*)}{1-a^{*m}}\frac{\beta^*}{a^*}} \Leftrightarrow M < \frac{a^*}{1-a^*}\left(\frac{1}{a^{*m}}-1\right),$$

which is exactly the same inequality to prove as in equation (A.13). Then the rest of the proof follows the proof of Proposition 2.

Proof of Proposition 5. Consider a simple violation strategy, i.e., the issuer issues tradable tokens in period 0 and disables tradability from period 1. The advantage is that the price is P_M^T (higher than P_M^{NT}) in period 0 but without tradability in actuality. The disadvantage comes from extra loss of consumer base. Since no trade happens, the subsequent token accumulation process is not affected. The excess revenue compared to the corresponding non-tradable token issuance reads

$$\Delta Rev = p\Delta P - c\Delta P + \beta^* c\Delta f(P) = \left[p-(1-\beta^*)c\frac{\Delta f(P)}{\Delta P}\right]\Delta P, \tag{A.21}$$

where $\Delta P = P_M^T - P_M^{NT} > 0$. As mentioned, the implicit assumption ensures the issuer has a chance to earn non-negative returns, i.e., $c < Q_M^x/f'(P_M^x)$. Similar to the proof process above, $p-(1-\beta^*)c\frac{\Delta f(P)}{\Delta P} > 0$. Then $\Delta Rev > 0$. In fact, there is a series of applicable strategies that allows issuers to earn excess revenue from pretended tradability, as long as the appropriate number of tokens exist that have been bought at a price that includes tradability but have not yet been traded.

Proof of Proposition 6. We prove the following two observations to show that the only equilibrium is not to issue tokens. (i) Regardless of whether tradability is permitted, consumers prefer $M = 1$ ("buy one, get one free"), i.e.,

$$C_1^T = C_1^{NT} = \min_{M \geq 1, x \in \{T, NT\}} C_M^x = \frac{p}{1-\beta} \frac{1+a^*}{1+a};$$

(ii) The issuer's payoff is always lower than not issuing tokens, i.e.,

$$U_M^x < U_0 = \frac{p}{1-\beta^*}, \quad \forall x \in \{T, NT\}, M \geq 1.$$

(i) When we compare (12)-(13) with the revenue formulas (non-tradable and tradable) discussed in Appendix A, we note that they have the same forms, respectively, but exchange the positions of β and β^*, e.g., $Rev_M^T = \frac{p}{1-\beta^*} \frac{M+a}{M+a^*}$, and $C_M^T = \frac{p}{1-\beta} \frac{M+a^*}{M+a}$. Recall the proof procedure in the above section. One can imagine the duality steps to prove the corresponding (opposite) properties of derivatives. However, if we only need to know the optimum, i.e.,

$$C_1^T = C_1^{NT} = \min_{M \geq 1} \left\{ C_M^{NT}, C_M^T \right\} = \frac{p}{1-\beta} \frac{1+a^*}{1+a}, \tag{A.22}$$

we can consider the following simpler proofs. First, we prove that $\forall M > 1$, $C_M^{NT} > \frac{p}{1-\beta} \frac{1+a^*}{1+a}$.

$$C_M^{NT} > \frac{p}{1-\beta} \frac{1+a^*}{1+a} \Leftrightarrow \frac{1-a^M}{1-a^{M+1}}(1+a) > \frac{1-a^{*M}}{1-a^{*M+1}}(1+a^*). \tag{A.23}$$

Note that $a^* < a$, and

$$\begin{aligned} \frac{\partial}{\partial a}\left[\frac{1-a^M}{1-a^{M+1}}(1+a)\right] &= \frac{1-a^2}{a^2(1-a^{M+1})^2}\left[a^2\left(\frac{1-a^{2M}}{1-a^2}\right) - Ma^{M+1}\right] \\ &= \frac{1-a^2}{a^2(1-a^{M+1})^2}\left(\sum_{i=1}^{M} a^{2i} - Ma^{M+1}\right) \underbrace{\quad > \quad}_{a^{2i}+a^{2(M+1-i)} > 2a^{M+1}} 0, \end{aligned} \tag{A.24}$$

the above inequality holds. Therefore, $C_M^{NT} > \frac{p}{1-\beta} \frac{1+a^*}{1+a}$.

On the other hand, it is easy to show $\frac{\mathrm{d}C_M^T}{\mathrm{d}M} > 0$. Thus, $C_M^T \geq C_1^T = \frac{p}{1-\beta} \frac{1+a^*}{1+a}$. Also, $C_1^{NT} = \frac{p}{1-\beta} \frac{1+a^*}{1+a}$. Therefore,

$$C_1^T = C_1^{NT} = \min_{M \geq 1} \left\{ C_M^{NT}, C_M^T \right\} = \frac{p}{1-\beta} \frac{1+a^*}{1+a}. \tag{A.25}$$

(ii) Now consider the issuer's payoff. When tokens are non-tradable,

$$Rev_M^{NT} = \underbrace{\frac{1}{1-a^{*M+1}}}_{\substack{\text{Discount of}\\ \text{All Rounds}}} \underbrace{(p+(1-p)a^*)}_{\substack{\text{Discount of}\\ \text{First Purchase}\\ \text{in a Round}}} \underbrace{\frac{1-a^{*M}}{1-a^*} P_M^{NT}}_{\substack{\text{PV of}\\ \text{M Purchases}}} = \frac{p}{1-\beta^*} \frac{1-a^{*M}}{1-a^{*M+1}} \frac{1-a^{*M+1}}{1-a^{*M}} = \frac{p}{1-\beta^*}.$$

Similarly, when tokens are tradable, $Rev_M^T = \frac{p}{1-\beta^*}$. Therefore, $\forall x \in \{T, NT\}, M \geq 1$, the payoff U_M^x satisfies (note that for issuers under perfect competition are not monopolists, $c > 0$)

$$U_M^x = \frac{p}{1-\beta^*} - cf(P_M^x) < \frac{p}{1-\beta^*} = U_0.$$

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics* 138, 1–35.

Banerjee, Abhijit. 1987. *On frequent flyer programs and other loyalty-inducing economic arrangements.* Harvard University.

Basso, Leonardo J, Matthew T Clements, and Thomas W Ross. 2009. "Moral hazard and customer loyalty programs." *American Economic Journal: Microeconomics* 1, 101–123.

Bolton, Ruth N, P KandBramlettMatthewD Kannan, and Matthew D Bramlett. 2000. "Implications of loyalty program membership and service experiences for customer retention and value." *Journal of the academy of marketing science* 28, 95–108.

Chen, Yanyan, Timo Mandler, and Lars Meyer-Waarden. 2021. "Three decades of research on loyalty programs: A literature review and future research agenda." *Journal of Business Research* 124, 179–197.

Chod, Jiri, and Evgeny Lyandres. 2023. "Product market competition with crypto tokens and smart contracts." *Journal of Financial Economics* 149, 73–91.

Chun, So Yeon, and Rebecca W Hamilton. 2024. "Paying with Money or Paying with Points: How Variable Versus Fixed Exchange Rates Influence Loyalty Point Redemption." *Journal of Marketing Research* 00222437241255650.

Chun, So Yeon, and Anton Ovchinnikov. 2019. "Strategic consumers, revenue management, and the design of loyalty programs." *Management Science* 65, 3969–3987.

Cong, Lin William, Ye Li, and Neng Wang. 2021. "Tokenomics: Dynamic adoption and valuation." *The Review of Financial Studies* 34, 1105–1155.

Cong, Lin William, Ye Li, and Neng Wang. 2022. "Token-based platform finance." *Journal of Financial Economics* 144, 972–991.

Evanschitzky, Heiner, Balasubramanian Ramaseshan, David M Woisetschläger, Verena Richelsen, Markus Blut, and Christof Backhaus. 2012. "Consequences of customer loyalty to the loyalty program and to the company." *Journal of the academy of marketing science* 40, 625–638.

Goldstein, Itay, Deeksha Gupta, and Ruslan Sverchkov. 2024. "Utility tokens as a commitment to competition." *The Journal of Finance* 79, 4197–4246.

Klemperer, Paul. 1987. "Markets with consumer switching costs." *The quarterly journal of economics* 102, 375–394.

Klemperer, Paul. 1995. "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade." *The review of economic studies* 62, 515–539.

Kumar, Viswanathan, and Werner Reinartz. 2016. "Creating enduring customer value." *Journal of marketing* 80, 36–68.

Lewis, Michael. 2004. "The influence of loyalty programs and short-term promotions on customer retention." *Journal of marketing research* 41, 281–292.

Lim, Freddy, So Yeon Chun, and Ville Satopää. 2024. "Loyalty Currency and Mental Accounting: Do Consumers Treat Points Like Money?" *Manufacturing & Service Operations Management* 26, 1878–1896.

Liu, Yuping. 2007. "The long-term impact of loyalty programs on consumer purchase behavior and loyalty." *Journal of marketing* 71, 19–35.

Luo, Dan. 2024. "Corporate Finance Through Loyalty Programs." *Available at SSRN* August 29.

Nunes, Joseph C, and Xavier Drèze. 2006. "Your loyalty program is betraying you.." *Harvard Business Review* 84, 124–31.

Rogoff, Kenneth, and Yang You. 2023. "Redeemable platform currencies." *The Review of Economic Studies* 90, 975–1008.

Rossi, Federico. 2018. "Lower price or higher reward? Measuring the effect of consumers' preferences on reward programs." *Management Science* 64, 4451–4470.

Selnes, Fred. 1993. "An examination of the effect of product performance on brand reputation, satisfaction and loyalty." *European Journal of marketing* 27, 19–35.

Sharp, Byron, and Anne Sharp. 1997. "Loyalty programs and their impact on repeat-purchase loyalty patterns." *International journal of Research in Marketing* 14, 473–486.

Sönmeztürk, Osman, Tolga Ayav, and Yusuf M Erten. 2020. "Loyalty program using blockchain." In *2020 IEEE international conference on blockchain (Blockchain)*,, 509–516, IEEE.

Sun, Yacheng, and Dan Zhang. 2019. "A model of customer reward programs with finite expiration terms." *Management Science* 65, 3889–3903.

For Online Publication

October 4, 2025

OA. Online Appendix

A. Extended Discussion: Bulk Sale

In the main text, we focus on loyalty token issuance designs, termed as "buy M, get one free," where $M \geq 1$. Now we consider the relevant bulk sale cases as shown in Section 2, corresponding to "buy one and get N ($= 1/M, M \in (0,1]$) free." For the sake of expression, we let one token represent one unit of redemption right, and the issuer sells one service bundled with N tokens, $N \in \mathbb{Z}^+$.

Again, we start with the three roles of bundling as shown in Section 3.3. First, the issuer sells a unit of service bundled with more future rights, thus requiring significantly stronger capacity (market power) for implementation. Second, regarding delayed cash flow, the non-tradable case has a similar intuition, while the secondary market structure changes significantly: token holders sell "superfluous" tokens to unshocked consumers and never buy themselves, becoming de facto competitive token dealers. When $M \geq 1$, they either sell or buy to collect fragmented rights, so that tokens can clear even without unshocked consumers. As a result, the total token supply becomes crucial to pricing and affects the issuer's cash flow. Third, the core difference is that consumers no longer need to accumulate tokens — they always receive enough tokens for the next redemption upon purchase. That is, there is no time-series bundling even when tokens are non-tradable.

One may still expect that greater market power will help maintain high token dependence, but the dominance of non-tradability may be broken, because the cash flow swap based on time-series bundling is disabled. Now we formally solve the prices and revenues for the "buy one, get N free" case. We limit our focus to the monopoly case, since it requires strong commitment power to maintain such an implementation.

Non-tradable tokens. The monopoly issuer charges the unit price P_N^{NT} to overwhelm the whole willingness to pay for $N+1$ units through one purchase,

$$P_N^{NT} = \sum_{i=0}^{N} \left(\frac{\beta p}{1-\beta(1-p)} \right)^i = \frac{1-a^{N+1}}{1-a}. \tag{OA.1}$$

Similar to (4), the revenue reads

$$\begin{aligned} Rev_N^{NT} &= \underbrace{(p+(1-p)a^*)}_{\substack{\text{PV of First}\\ \text{Purchase}}} \underbrace{\frac{1}{1-a^{*N+1}}}_{\substack{\text{Discount since}\\ \text{First Purchase}}} \underbrace{P_N^{NT}}_{\text{Unit Price}} \equiv Q_N^{NT} \cdot P_N^{NT} \\ &= \frac{p}{1-\beta^*} \frac{1-a^*}{1-a^{*N+1}} \frac{1-a^{N+1}}{1-a}. \end{aligned} \tag{OA.2}$$

There is no difficulty in showing $\frac{\mathrm{d}Rev_N^{NT}}{\mathrm{d}N} > 0$, i.e., the issuer always gains greater revenue by bundling more tokens, since it is always able to leave zero consumer surplus. As a result, the monopoly issuer would optimally announce a "buy one to get all free forever" strategy, i.e., a lifetime membership at the price $\lim_{N\to\infty} Rev_N^T = \frac{p}{1-\beta^*}\frac{1-a^*}{1-a}$.

This is formally consistent with our main focus, $M \geq 1$: when $M \to 0$, the issuance of loyalty tokens converges to a membership model that yields the largest front load. Proposition 3 still holds, indicating that it requires more market power to implement a "buy one get N free" strategy than "buy M get one free." In practice, regulatory issues seem to be more crucial, as consumers are exposed to a significant risk of being scammed by firms promising future consumption and then walking away or going bankrupt. On the other hand, in the extreme case, the membership model effectively makes the consumer a shareholder, raising considerations of capital gains and tax issues, which is beyond our focus.

Tradable tokens. If $N = 1$, it is equivalent to $M = 1$, i.e., there is no trade. When $N > 1$, consumers (after purchase) have "superfluous" tokens for the next redemption, thus becoming de facto token dealers. It is easy to show that: (i) if a consumer has two tokens, selling them to two people is more beneficial than selling them to one person, because people's willingness to pay for the second token is lower; (ii) there is a trade if and only if there exist two consumers whose number of held tokens differs by at least 2, as otherwise consumers have indifferent valuation on the "superfluous" token. As a result, the secondary market will make tokens as evenly allocated as possible across the entire consumer base. Given that there are $S_{N,t}^T$ outstanding tokens in period t, then each consumer holds $\lfloor S_{N,t}^T \rfloor$ or $\lceil S_{N,t}^T \rceil$ tokens after trade.

Consider the price. Note that token sellers, those who hold "superfluous" tokens, are competitive and do not have pricing power. The market clearing price equals the value of the marginal token. In particular, each seller sells $(N - \lceil S_{N,t}^T \rceil)$ tokens at the market-clearing secondary market price,

$$\hat{P}_t = a^{\lfloor S_{N,t}^T \rfloor}(p + (1-p)a). \tag{OA.3}$$

Turn to the primary market. If $\lfloor S_{N,t}^T \rfloor \geq 1$, then there would be no fiat money purchase at time t, i.e., the issuer does not sell tokens. In other periods, the issuer sells $Q_{N,t}^T$ services and can take away all the resellers' surplus. However, this leaves the buyers' surplus, due to the lack of monopoly power of the resellers. In addition, the issuer must maintain a stable service price, although $S_{N,t}^T$ could vary across periods. Therefore, the price satisfies:

$$\begin{aligned} P_N^T &= \min_{\{t|S_{N,t}^T \geq 1\}} \left\{ \underbrace{\sum_{i=0}^{\lceil S_{N,t}^T \rceil} a^i}_{\text{self-use for redemption}} + \underbrace{\beta(N - \lceil S_{N,t}^T \rceil)\hat{P}_t}_{\text{reselling option value}} \right\} \\ &= \min_{\{t|S_{N,t}^T \geq 1\}} \left\{ \sum_{i=0}^{\lceil S_{N,t}^T \rceil} a^i + a^{\lceil S_{N,t}^T \rceil}(N - \lceil S_{N,t}^T \rceil) \right\}. \end{aligned} \tag{OA.4}$$

Note that $\forall k \in \mathbb{Z}$, $k \geq 0$,

$$\left\{ \sum_{i=0}^{k+1} a^i + a^{k+1}[N-(k+1)] \right\} - \left\{ \sum_{i=0}^{k} a^i + a^k(N-k) \right\} = -a^k(N-k)(1-a) < 0. \quad \text{(OA.5)}$$

Therefore,

$$P_N^T = \sum_{i=0}^{\max_{t\geq 0}\lceil S_{N,t}^T \rceil} a^i + a^{\max_{t\geq 0}\lceil S_{N,t}^T \rceil}(N - \max_{t\geq 0}\lceil S_{N,t}^T \rceil). \quad \text{(OA.6)}$$

Consider $\max_{t\geq 0}\lceil S_{N,t}^T \rceil$. Note that

$$S_{N,t+1} \begin{cases} = S_{N,t}^T - p < S_{N,t}^T, & S_{N,t}^T \geq 1; \\ = [1-p(N+1)]S_t + pN \leq pN, & S_{N,t}^T < 1, \text{ and } 1-p(N+1) < 0; \\ = [1-p(N+1)]S_t + pN < 1-p, & S_{N,t}^T < 1, \text{ and } 1-p(N+1) \geq 0. \end{cases} \quad \text{(OA.7)}$$

Therefore, $S_{N,t}^T \leq \max\{1, pN\} \Rightarrow \max_{t\geq 0}\lceil S_{N,t}^T \rceil = \lceil pN \rceil$, and

$$P_N^T = \sum_{i=0}^{\lceil pN \rceil} a^i + a^{\lceil pN \rceil}(N - \lceil pN \rceil). \quad \text{(OA.8)}$$

Similar to Section 3.4, we solve $\{Q_{N,t}^T\}_{t\geq 0}$ and $\{S_{N,t}^T\}_{t\geq 0}$ jointly. Initially, no one has redemption rights, $S_{N,0}^T = 0$. In each period, only those who do not have tokens choose to purchase by fiat money (note that tokens are allocated evenly), therefore

$$p = Q_{N,t}^T + p\min\{1, S_{N,t}^T\}. \quad \text{(OA.9)}$$

The outstanding unredeemed tokens in the next period equals current outstanding minus current redemption plus new issuance, i.e.,

$$S_{N,t+1}^T = NQ_{N,t}^T + S_{N,t}^T - p\min\{1, S_{N,t}^T\}. \quad \text{(OA.10)}$$

These two equations can be viewed as generalized versions of (5) and (6), where $S_{N,t}^T$ is always smaller than 1. Then the revenue reads

$$Rev_N^T = P_N^T \sum_{t=0}^{\infty} \beta^{*t} Q_{N,t}^T. \quad \text{(OA.11)}$$

Although it is hard to obtain the generic formula of $\{Q_{N,t}^T\}_{t\geq 0}$, we can numerically calculate Rev_N^T based on the recurrence formulas (OA.9) and (OA.10).

Before delving into numerical explorations, we qualitatively discuss how Rev_N^T changes with N. There are two opposing forces that make P_N^T potentially nonmonotonic with N. First, a larger N corresponds to more tokens bundled, yielding a positive force for the service price P_N^T, as the first term in (OA.8), $\sum_{i=0}^{\lceil pN \rceil} a^i$, shows. This is in line with the role of tradability in the "buy M get one free" case, partially compensating for delayed cash flow. Second, a larger N brings a new negative force by changing the value of the marginal token in the secondary market, so it may lower the service price, as the second term in (OA.8), $a^{\lceil pN \rceil}(N - \lceil pN \rceil)$, shows. This is de facto the other tale of leaving a positive buyer surplus, while in the main case, the marginal token always corresponds to the first redemption right, where people trade only to collect the "fragmented" rights, rather than sell "superfluous" rights.

Figure OA1 shows the numerical results of Rev_N^T under different issuance strategy N and consumption probability p, and compares to the non-tradable case, Rev_N^{NT}. In Panel A where p is relatively large, it appears that Rev_N^T is increasing monotonically with N $(1/M)$. This is consistent with Lemma 2, $\frac{\mathrm{d}Rev_M^T}{\mathrm{d}M} < 0$. This is because the first effect is dominant. Meanwhile, non-tradability dominates tradability, since tradability allows competitive resellers in the secondary market, leaving a positive consumer surplus.

In Panel B, non-tradability is still the dominant strategy. However, Rev_N^T is no longer monotonic, since the negative force, $a^{\lceil pN \rceil}(N - \lceil pN \rceil)$, becomes stronger. With a lower p, this term decreases faster with N mathematically. From an economic perspective, a lower p implies a longer interval between two redemptions. Then as the marginal token increases by one, the valuation of tokens suffers from a greater discount.

When p is extremely low, as shown in Panel C, the dominance of tradability breaks, i.e., $\max_N Rev_N^T > \max_N Rev_N^{NT}$,[42] where the global optimum solves a relatively small N, rather than $N \to \infty$. The reason is twofold. First, when N is small, the negative force is less pronounced, as mentioned. In particular, as long as $pN < 1$, the marginal token is always the first token, which means that buyers do not obtain a positive surplus. Therefore, the drawback of tradability is disabled. Second, with an extremely low p, consumers largely discount the future willingness to pay. This makes lifetime membership less valuable than immediate transfer rights. As a result, $Rev_{N^*}^T > \lim_{N\to\infty} Rev_N^{NT}$, i.e., it could be worse to sell a membership to a small group than to sell a few tokens to them and let them resell to the others.

Similarly, if the issuer is impatient ($\beta^* \to 0$), tradability could always beat non-tradability. In this extreme case, the issuer only earns p purchases in the initial period, and can always charge a higher price from the reselling option value when allowing tradability. This counterexample does not apply to our main focus, $M \geq 1$. Although $\forall M > 1$, the issuer takes away the reselling option values, these strategies are all suboptimal relative to $M = 1$, where no trade happens, and the token price equals the non-tradable one.

In summary, although rare in reality and subject to higher regulatory requirements, the "buy one and get N free" ($M < 1$) case can also be consistently included in our framework. Compared to our main focus, $M \geq 1$, this case naturally disables time-series bundling, and somehow appears to be more like the benchmark case of platform currencies. Whereas the non-tradable issuance converges to a lifetime membership model in optimal, tradability yields a non-monotonic service pricing with token bundling. In particular, due to losing the unique

[42] Here we focus on the monopoly case, much as we did in Section 3.4. Therefore, $Rev_N^x = U_N^x$, $\forall N \geq 1$, $x \in \{T, NT\}$.

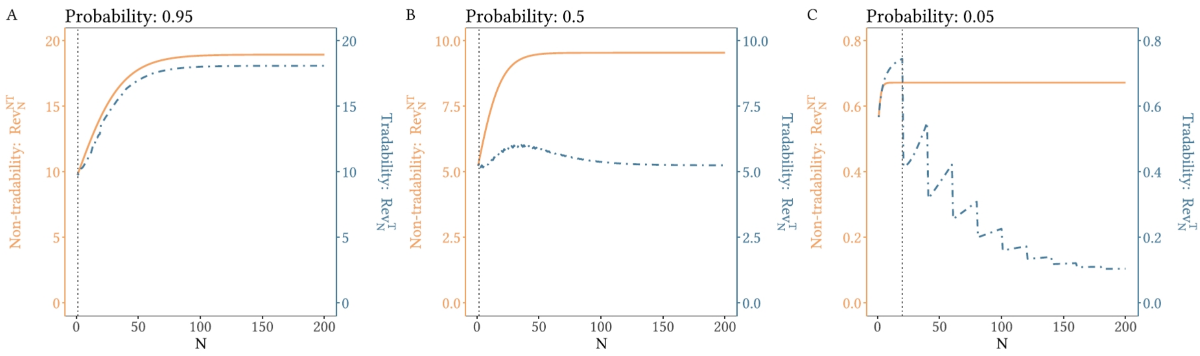

Figure OA1. Revenue Comparison under "Buy One - Get N Free"

This plot compares the revenues of "buy one and get N free tradable / non-tradable tokens," i.e., Rev_N^T and Rev_N^{NT}, under a different N and probability p. In Panels A-C, p equals 0.95, 0.5, and 0.05, respectively. In each panel, the x-axis is the issuance strategy N, the yellow solid line and its corresponding left y-axis represents revenues in non-tradability cases, Rev_N^{NT}, and the blue dashed line and the right y-axis show the revenues with tradability, Rev_N^T. The gray dotted line plots $N = 1/p$ to illustrate the first step-wise segment. $\beta = 0.95$, $\beta^* = 0.9$.

advantage of time-series bundling, a non-tradable issuance may not take dominance when immediate purchases are vitally important.

B. Alternative Evaluation of Loyalty Token Issuance

In the main text, we mainly use the token dependence Θ to evaluate the effectiveness of loyalty token issuance. This represents the revenue share of presale by loyalty tokens, and corresponds well to empirical proxies. Here we discuss the alternative evaluation Δ, defined as the net payoff growth relative to the benchmark of not issuing any tokens, i.e.,

$$\Delta \equiv (U^* - Rev_0)/Rev_0, \tag{OA.12}$$

where U^* is the payoff under the optimal issuance plan $(M_*^{x^*}, x^*)$, and Rev_0 denotes benchmark revenue (without tokenization).

Figure OA2 explores the variation in optimal loyalty payoffs U^* over the parameter space (p, c), and, importantly, examines how much the payoffs are enhanced by adopting loyalty token issuance. Panel (a) visualizes $U^*(p, c)$. The difference in the probability of consumption demand dominates the variation in revenue, which usually reflects the varying characteristics among industries.[43] That is, $Rev_0(p)$ naturally changes with p and in turn increases aggregate revenues.

Figure OA2 (b) examines Δ from a different perspective than Figure 4 (c). Loyalty tokens are most beneficial for relatively small demand probabilities, which partially compensates for the negative impact of the low consumption frequency. However, when the consumption frequency is extremely low, only issuers with absolute power (such as pure monopolies) can

[43]In practice, the unit price may also affect the revenue comparison among industries, which is not our focus. Also, when looking at relative concepts, i.e., Rev^*/Rev_0 and the token dependence Θ, the influence of unit price is eliminated. Thus, for simplicity, we normalize the current unit production cost to 1.

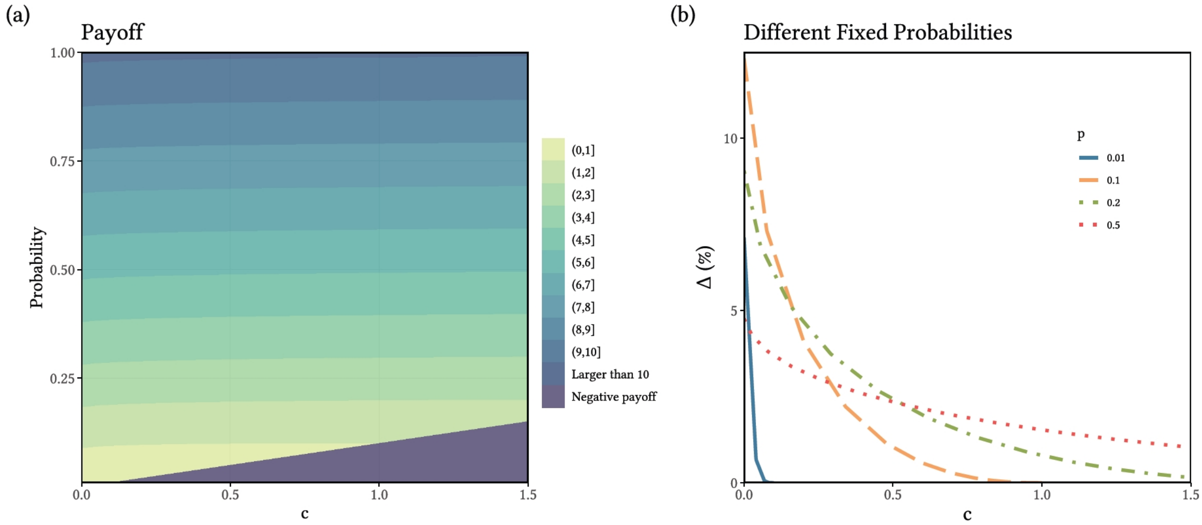

Figure OA2. Payoff Enhanced by Optimal Loyalty Token Designs

Panel (a) plots the aggregate payoff U defined in (1) when adopting the optimal issuance plan for each point (p, c), where the optimal issuance $(M_*^{x^*}, x^*)$ is illustrated in Figure 4. Panel (b) visualizes Δ, the payoff growth rate (accounting for external losses) driven by the optimal issuance, benchmarked against the non-use of any loyalty token offering. Different p are fixed, and the colored curves show how growth changes with market power. $\beta = 0.95$, $\beta^* = 0.9$, and $f(P) = P - 1$.

profitably implement a loyalty program. In summary, issuers with great market power in industries with relatively low-frequency consumption are most favored for issuing loyalty tokens.

C. Supplementary Analysis under the Alternative Timeline

As discussed in Section 4.1, the alternative timeline makes tradability more valuable, as consumers' heterogeneity in willingness to pay for tokens after shock is translated into additional reselling option value, thereby increasing the service (token) price in the primary market. However, even compared to the non-tradable token design without a secured redemption policy, tradability does not necessarily take the dominance.

First, the new timeline does not necessarily yield higher revenues compared to the same tradable-token issuance strategy (M, T) in the baseline. This is because "trading after demand shocks are realized" not only increases the price but also accelerates redemption, thereby delaying the cash flow from fiat-money consumption. As shown in Figure OA3 Panel (a), when the consumption probability is moderate ($p = 0.5$), the new timeline hardly generates higher revenues unless the issuer is highly impatient ($\beta^* \to 0$). In contrast, when the consumption probability is low, as in Panel (b) where $p = 0.05$, an issuer with a slightly higher β^* can potentially earn more from tradable tokens under the new timeline. In this case, the lower consumption probability offsets the drawback of accelerated redemption.

Second, outperforming the same strategy under the baseline timeline is only a necessary condition for achieving tradability dominance. However, solving the optimization problem (1)

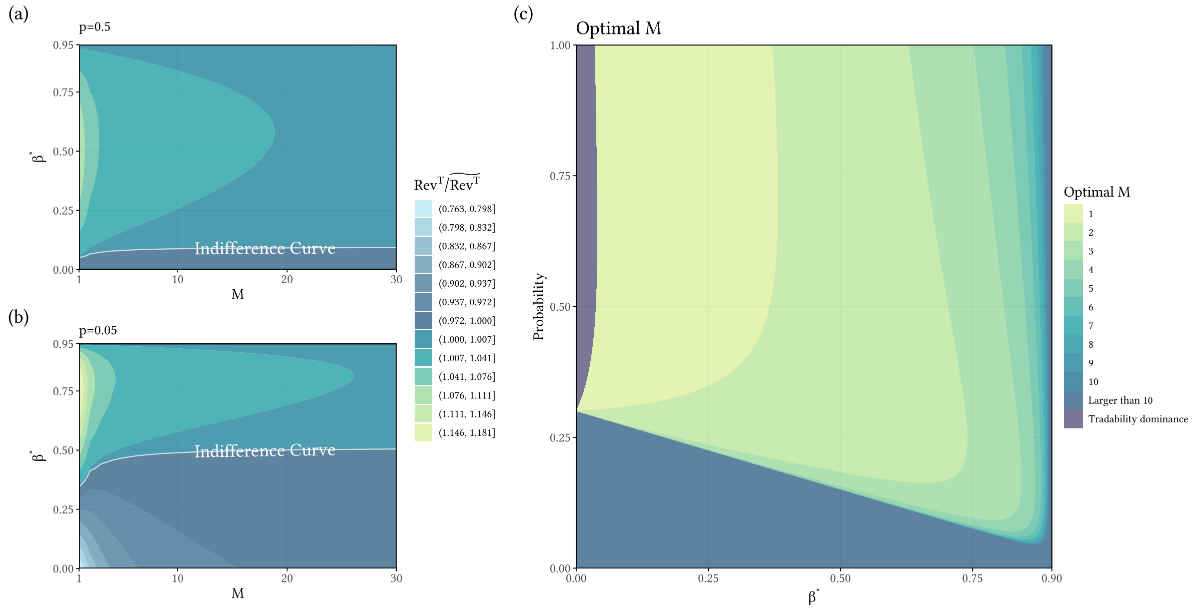

Figure OA3. Dominance of Non-tradability under the Alternative Timeline

This figure illustrates the dominance of non-tradability under the alternative timeline in which trade occurs after the realization of demand shocks. In Panels (a) and (b), we compare the revenues from tradable token issuance (M, T) in the baseline timeline (Rev_M^T) and in the alternative timeline $(\widetilde{Rev_M^T})$ for different values of p ($p = 0.5$ in Panel (a) and $p = 0.05$ in Panel (b)). The white curves represent the indifference curves where $Rev_M^T = \widetilde{Rev_M^T}$. Panel (c) visualizes the issuer's optimal strategy (M, x) for different values of (p, β^*), while keeping $\beta = 0.9$ and $c = 0.3$ fixed. When the optimal strategy is to issue non-tradable tokens, the grid is colored according to the resulting optimal M^*. When the optimal strategy is to issue tradable tokens, the grid is shaded in purple (as shown in the upper left).

under the new timeline — as visualized in Panel (c) — shows that tradability dominates only when the issuer's discount β^* is sufficiently low and the consumption probability p is relatively high. When $\beta^* \to 0$, the issuer only cares about token sales in the first period, therefore only token price matters. Meanwhile, although a lower p allows tradability to benefit more from the timeline adjustment, it inherently confers a greater advantage to non-tradable tokens, as earlier shown in Figure 3 Panel (b).