# Payment platforms and pricing: when does a "one price rule" help consumers?*

Chang Liu† Fengshi Niu‡ Alexander White§

January 10, 2019

**Abstract**

Should a payment intermediary (e.g. American Express, PayPal, etc.) be allowed to require merchants to charge the same price for customers who pay via the platform as they charge other customers? We build a model that highlights certain under-appreciated aspects of this questions. Taking these into account, we find that such platform-imposed requirements may benefit consumers in ways that recent work on the topic has overlooked. A central aspect of our approach is to draw connections between this question, as it relates to payment platforms, and recent, more abstract work on third-degree price discrimination.

**Keywords:** Platforms, Payment intermediaries, Electronic payments, Price Coherence, Third-Degree Price Discrimination, Demand Curvature

Preliminary Draft

*This is a preliminary draft. Comments are welcome.

†Department of Economics, Harvard University; chang_liu@g.harvard.edu

‡Department of Economics, University of California, Berkeley; fniu@berkeley.edu

§School of Economics and Management and National Institute for Fiscal Studies, Tsinghua University; awhite@sem.tsinghua.edu.cn

# 1 Introduction

In 2018, the United States Supreme Court ruled in favor of American Express in a 5-4 decision. The majority opinion allows AmEx to legally impose a rule saying, essentially, merchants that sign up to accept AmEx must charge the same prices, regardless of whether a customer pays by AmEx or another method. In the case, the US Department of Justice argued that merchants should be given the ability to impose a surcharge on customers who paid with AmEx, which levies a relatively high transaction fee on merchants.

This recent case is the latest episode in a longstanding debate surrounding this issue of "price coherence versus price flexibility." Despite the apparent simplicity of the question at hand, it turns out to be rather difficult to analyze the welfare effects of allowing payment intermediaries to impose price coherence. This paper builds a simple model designed to help analyze certain aspects of this question. Broadly speaking, we find that, when a platform is allowed to impose a price coherence policy on merchants, it has the potential to make consumers better off, for reasons that have been under-appreciated.

A common feature of models examining this issue, in which some consumers are "cardholders" and some are not and can pay only by cash, is the following basic property. When the payment platform moves from imposing price price coherence (i.e., requiring a single price), to allowing price flexibility (i.e., permitting a surcharge for paying by card), the new prices diverge, with the cash price lying below the previous, single price and the card price lying above. Thus, compared to the situation where merchants can surcharge purchases made by card, price coherence favors cardholders and harms cash buyers. Much of the analysis then boils down to analyzing which of these effects is larger.

Therefore, the analysis is closely related to the classic paradigm comparing uniform pricing with third-degree price discrimination. This paper draws significantly on the approaches taken in that literature. We attempt to show how this lens can help to simplify the comparison between price coherence and flexibility.

Compared to the prototypical model in the uniform-versus-differential pricing paradigm, however, there are a few important differences. First, unlike in the standard environment, but as in a recent article by Chen and Schwartz (2015), a merchant's marginal costs differ when serving cardholders and cash buyers. Second, unlike in that article, because this difference in marginal costs is driven by the transaction fee chosen by the platform, it is determined endogenously. Third, to some extent, consumers can self-select into one market or the other. That is, when they are ready to make a payment consumers who are already cardholders still retain the option to pay by cash.

Taking these differences into account, we derive three main results. First, we establish a sufficient condition ("Condition 1") under which, given three arbitrary prices (i.e., one cash price, one card price, and one coherence price), consumer surplus is higher under the coherence regime than under the flexible regime. Interestingly, in Chen and Schwartz (2015)'s model, in which consumers *cannot* self-select into markets at all, this same condition is sufficient for consumer surplus to be greater under differential pricing. Using an intuitive graph (see Figure 1), we explain why this difference is makes an important difference.

Second, taking the transaction fee charged by the platform to be exogenous, we provide sufficient conditions under which the merchant's trio of potential equilibrium prices satisfy Condition 1. We show that, under a flexible specification of demand, this occurs so long as demand is weakly concave. Furthermore, we show that, although such a bound on the convexity of demand is necessary in order for Condition 1 to hold, the substantive outcome, in which consumers are better off under price coherence holds for moderately convex demand.

Third, we allow the platform's transaction fees to be set endogenously under each of the two regimes (and we analyze the preferences between the two regimes). We show that, not only does the platform prefer price coherence, but it sets a lower transaction fee under price coherence than it would under price flexibility. This is because, when it is free to set flexibly set the price it charges cardholders, the merchant would be willing to tolerate a higher transaction fee and still accept the card than it is under price coherence. This difference between the transaction fees under the two regimes strengthens the result discussed just above. In other words, when the transaction fees are endogenous, the maximum level of demand convexity such that price coherence helps consumers is higher than when they are exogenous.

## 1.1 Related Literature

Our model takes its inspiration most directly from two earlier works, Edelman and Wright (2015) and Schwartz and Vincent (2006), both of which address the same broad question we do. The crucial added feature of our model compared to Edelman and Wright's is that we consider a setting with variable total demand for the good. In contrast, they consider merchants that compete in a Vickrey-Salop circle. Relaxing the assumption of fixed total demand is one crucial ingredient that allows for our results. An important feature of their model is that users endogenously choose whether or not to sign up for the card in the first place. In order to highlight key results regarding pricing of the good under the two regimes, we initially ignore this issue and then incorporate it in Section 5. They also consider other issues that we do not address, including the incentives for the platform to invest in order to provide better rewards for cardholders.

Unlike in Schwartz and Vincent's model, ours allows for cardholders to retain the option to pay by cash. In their model, each of the two groups is confined to one payment option. Consequently, in their setting, the joint incentives facing the platform and the merchant are quite different, as they not need worry about the card price reaching a level that leads cardholders to pay by cash. It turns out that including this option for cardholders to pay by cash tends to tilt the comparison between the two regimes in favor or price coherence.

A pair of recent papers that focus on similar issues are Bourguignon, Gomes, and Tirole (2018) and Gomes and Tirole (2018). These papers follow quite a different framework in which they assume consumers do not learn about their costs/benefits of paying via the platform until the time of a given sale. Compared to Edelman and Wright (2015) and Schwartz and Vincent (2006), these papers align more with ours in the finding that platform-imposed restrictions of surcharging (or cash discounting) may be efficient. It will be interesting to further explore the extent to which these findings may or may not, in some fundamental sense, rely on the same mechanism. One way in which our approach differs from these is in its attempt to draw the closest connection possible between the questions of price coherence versus flexibility and the classic one of uniform versus differential pricing. We focus most directly on the comparison

Preliminary Draft

to Chen and Schwartz (2015), but also see Aguirre et al. (2010), Cowan (2012), Cowan (2016), among others, which study the fundamental tradeoffs between uniform and differential pricing.

## 2 The Model

Consider the following model in which an intermediary ("the platform") mediates transactions between some, but not all, buyers and a merchant. Each member of the unit mass of buyers has a valuation for the merchant's good, $v$, drawn from a twice continuously differentiable distribution, $g$, with strictly positive support on $(0, \overline{v})$, where $0 < \overline{v} \leq \infty$. A fraction, $\lambda \in (0, 1)$, of buyers have not joined the platform and, thus, may use only cash to purchase the good. If they do so, they receive payoff $v - p$, where $p$ denotes the price of the good.

The remaining $1 - \lambda$ buyers are "cardholders" that have joined the platform. When deciding whether or not to purchase the good, they may also choose whether to pay by card or by cash. If a cardholder pays by card, she receives a benefit, $b > 0$, which can represent time saved from not needing to visit an ATM but can also capture more substantial features, such as warranty protections offered by the platform for purchases made using the card. Cardholders who pay by card receive a payoff of $v + b - p$. If cardholders choose to pay by cash or to not purchase at all, they receive the same payoffs as non-cardholders. When cardholders face the option to pay by card at price $p$, we refer to $p - b$ as their *cash-equivalent price*.

Note that, here, buyers' decision whether or not to be cardholders is exogenously given. We relax this assumption in Section 5. Also, the distribution of buyers' valuations for the good is the same, independently of whether or not they are cardholders.

We are interested in comparing two types of contractual arrangement between the platform and the merchant, which, following Edelman and Wright (2015), we label "price flexibility" (F) and "price coherence" (C). In both arrangements, the platform collects a "transaction fee," $f$, from the merchant, for each sale made using the card. Under price flexibility, the merchant is allowed to charge two different prices – one, $p_m$, for "mediated" purchases made with the card and another, $p_d$, for "direct" purchases made using cash. In contrast, under price coherence, the merchant must charge the same price, $\hat{p}$, to all buyers, regardless of their method of payment.

The merchant produces the good at zero marginal cost. Under price flexibility, assuming all cardholders who purchase prefer to pay by card (which, we will show, holds at equilibrium), the merchant faces demand from cash buyers equal to the $\lambda$-share of $Q(p_d) = \int_{p_d}^{\infty} g(x)\,dx$. Demand from cardholders is the $(1-\lambda)$-share of $Q(p_m - b) = \int_{p_m - b}^{\infty} g(x)\,dx$. Thus, the merchant's total profits are

$$\lambda Q(p_d)\,p_d + (1-\lambda)\,Q(p_m - b)(p_m - f). \tag{1}$$

Under price coherence, the demands facing the merchant are analogous, but it can choose only one price, $\hat{p}$, giving rise to profits of

$$\lambda Q(\hat{p})\,\hat{p} + (1-\lambda)\,Q(\hat{p} - b)(\hat{p} - f). \tag{2}$$

The platform has zero costs. It receives $f$ for each purchase made using the card. Therefore, under price flexibility, it earns profits of $(1-\lambda)\,Q(p_m - b)\,f$, and, under price coherence, it earns profits of $(1-\lambda)\,Q(\hat{p} - b)\,f$.

The timing is as follows.

1. The platform sets the transaction fee, $f$, and it chooses which arrangement to use, price flexibility or price coherence.

2. The merchant chooses whether or not to accept the card. If it accepts, then, under price flexibility, it sets $p_d$ and $p_m$, whereas, under price coherence, it sets $\hat{p}$.

3. Buyers choose whether or not to purchase the good. In so doing, cardholders can choose whether to pay by card or by cash.

Our solution concept is subgame perfect equilibrium.

# 3 Pricing and Consumer Surplus with exogenous transaction fees

In this subsection, we focus on the last two stages of the game. In particular, we analyze the pricing incentives and the welfare consequences of price flexibility versus price coherence, holding fixed the transaction fee, $f$, at an exogenously set level that remains constant across the two regimes.

### 3.1 Price Flexibility

In the final stage, non-cardholders purchase the good if and only if $v \geq p_d$. Regarding the choices facing cardholders, first, they prefer to make a purchase using the card rather than cash if and only if $p_m - b \leq p_d$. Second, if this inequality is satisfied, they purchase the good if and only if $v + b \geq p_m$.

Next, consider merchant's price-setting problem in stage 2. Lemma 1 says when the merchant chooses to accept the card.

**Lemma 1.** *Under price flexibility, the merchant accepts the card if and only if $f \leq b$.*

To set prices, the merchant maximizes equation (1) with respect to $p_d$ and $p_m$, yielding

$$p_d^* = \frac{Q\left(p_d^*\right)}{-Q'\left(p_d^*\right)}, \qquad p_m^* = f + \frac{Q\left(p_m^* - b\right)}{-Q'\left(p_m^* - b\right)}. \tag{3}$$

Since $f \leq b$, cardholders' cash-equivalent price, $p_m^* - b \leq p_d^*$, ensuring that cardholders who buy choose to pay by card.

### 3.2 Price Coherence

In the final stage, assuming the merchant accepts the card, non-cardholders purchase the good if and only if $v \geq \hat{p}$. Cardholders purchase if and only if $v + b \geq \hat{p}$. If the merchant accepts the card, to set its price, it maximizes (2) with respect to $\hat{p}$, giving first-order condition

$$\lambda\left(Q\left(\hat{p}^*\right) + \hat{p}^* Q'\left(\hat{p}^*\right)\right) + (1-\lambda)\left(Q\left(\hat{p}^* - b\right) + \left(\hat{p}^* - f\right) Q'\left(\hat{p}^* - b\right)\right) = 0. \tag{4}$$

If it does not accept the card, the merchant sets a cash-only price equal to $p_d^*$ in (3). Lemma 2 states the condition under which the merchant chooses to accept the card.

**Lemma 2.** *Under price coherence, the merchant accepts the card if and only if $f \leq \overline{f}$, where $\overline{f} \in (0, b)$.*

### 3.3 Ranking of prices under the two regimes

We now establish the ranking among a set of relevant prices arising from the two regimes. In order to do this, we impose the following assumption throughout.

**Assumption 1.** *The demand function, $Q(\cdot)$, is globally strictly log-concave.*

This assumption, restricts the "pass-through rate"[1] of the demand function to be strictly less than one. Also, in order to restrict attention to cases in which the merchant optimally accepts the card under either regime, we assume that $f \in \left[0, \overline{f}\right]$. We now state Lemma 3.

**Lemma 3.** *Given any transaction fee that leads the merchant to accept the card under either price flexibility or price coherence, the following ranking holds:*

$$\hat{p}^* - b < p_m^* - b < p_d^* < \hat{p}^* < p_m^*. \tag{5}$$

To understand this lemma, first consider the ordering $p_m^* - b < p_d^* < p_m^*$. This says that, under price flexibility, the nominal price paid by cardholders is greater than the cash price; however, the cash price exceeds cardholders' cash-equivalent price. Note that $p_d^* < p_m^*$ depends on Assumption 1, which effectively limits the merchant's incentive to discount cardholders' cash-equivalent price by too much, compared to the cash price. The fact that $\hat{p}^*$ lies between $p_d^*$ and $p_m^*$ follows standard logic from third-degree price discrimination that, when a monopolist is constrained to set a uniform price in two markets, this price must be between the optimal price in each of the respective markets (see, e.g., Schmalensee (1981)).

### 3.4 Consumer surplus under the two regimes

Now we compare consumer surplus under price flexibility and price coherence. We first give a result that applies to an exogenous set of prices that follows the ranking established in Lemma 3. Then we move on to the case of endogenous prices.

[1]The pass-through rate says how fast a monopolist facing a given demand curve optimally increases its price in response to an increase in marginal cost. It can be derived by totally differentiating the standard monopoly pricing formula, $p^* = mc + \frac{Q(p^*)}{-Q'(p^*)}$, with respect to $mc$, yielding $dp^*/dmc = 1/\left(2 - \frac{Q''}{Q'} / \frac{Q'}{Q}\right)$. For details, see Bulow and Pfleiderer (1983) and Weyl and Fabinger (2013).

For three prices, $p_d$, $p_m$ and $\hat{p}$, we will refer to the following condition.

**Condition 1.** $\hat{p} \leq \lambda p_d + (1-\lambda)\, p_m$.

Denote consumer surplus associated with demand function $Q(\cdot)$ by $S(p) \equiv \int_p^\infty Q(x)\, dx$. Note that for all $p \in (0, \overline{v})$, $S'(p) = -Q(p) < 0$, $S''(p) = -Q'(p) > 0$. Under the respective regimes, consumer surplus is thus $S^F(p_d, p_m) \equiv \lambda S(p_d) + (1-\lambda)\, S(p_m - b)$ and $S^C(\hat{p}) \equiv \lambda S(\hat{p}) + (1-\lambda)\, S(\hat{p} - b)$. In Proposition 1, we make use of this convexity of $S(\cdot)$ to compare consumer surplus under the two regimes.

**Proposition 1.** *For any set of prices and cash-equivalent prices ranked as in equation* (5)*, if Condition 1 is satisfied, then consumer surplus is greater under price coherence than under price flexibility, i.e.,* $S^C(\hat{p}) > S^F(p_d, p_m)$.

To understand Proposition 1, imagine a shift from a regime with cash price $p_d$ and card price $p_m$ to a regime with a single price $\hat{p}$. As cash buyers now face a higher price and cardholders now face a lower price, the former group is worse off, and the latter group is better off. Condition 1 ensures that the latter effect dominates. To see this, rewrite Condition 1 as $\lambda(\hat{p} - p_d) \leq (1-\lambda)(p_m - \hat{p})$, a bound on cash buyers' price increase compared to cardholders' price decrease, and consider Figure 1. Area A represents cash buyers' loss in a standard way. Cardholders' gain, however, includes not only area B, which would typically arise when $p_m$ falls to $\hat{p}$, but also area C. This is because, among cardholders, new buyers are those with valuations *for the good* that are between the *cash-equivalent* prices $p_m - b$ and $\hat{p} - b$, rather than the nominal prices, $p_m$ and $\hat{p}$.

Contrast this with a prototypical exercise comparing consumer surplus under differential and uniform pricing. In such an example, under differential pricing, the "low market" features price $p_d$, the "high market" features price $p_m$, and the uniform price is $\hat{p}$. Here, following a shift from differential to uniform pricing, the exiting buyers in the "low market" still have valuations between $p_d$ and $\hat{p}$, but the new buyers in the "high market" have valuations between $\hat{p}$ and $p_m$. Thus, the gain in consumer surplus in the high market includes only area B. In such an exercise with exogenous prices, the conditions under which consumers could, on net, gain from a switch

Preliminary Draft

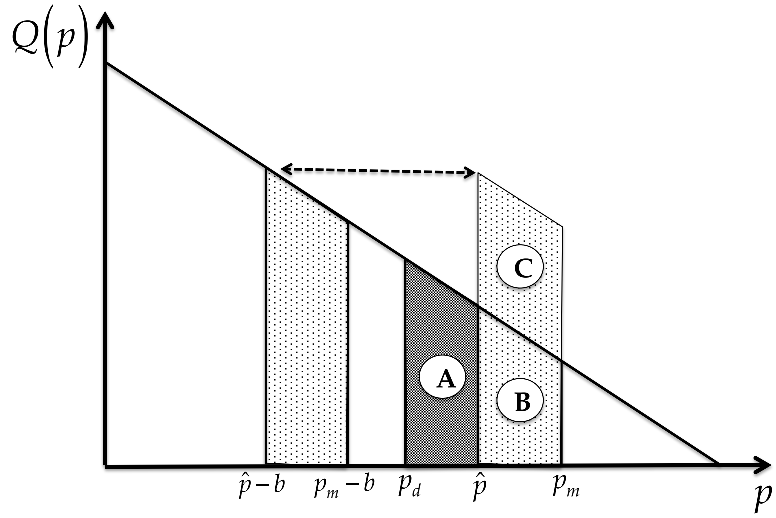


***Figure 1:*** *Change in consumer surplus following a shift in pricing regime. Area A represents cash buyers' loss and areas B and C, together, represent card buyers' gain, when nominal prices shift from $p_d$ and $p_m$ to a single price $\hat{p}$.*



from differential to uniform pricing are more restrictive than in our setting.[2]

We now move to the case where the merchant optimally chooses $p_d^*$, $p_m^*$ and $\hat{p}^*$. Here, to simplify the analysis, we focus on the following *constant pass-through* family of demand functions:

$$Q(p) = (1 - p/\overline{v})^{\gamma}, \quad \gamma > 0, \quad 0 \le p \le \overline{v}. \tag{6}$$

Note that in this specification, the pass-through rate equals $\gamma/(1+\gamma)$, and $\gamma = 1$ corresponds to the special case of linear demand. This form of demand leads to a straightforward result which we state in Proposition 2. In doing so, we assume that $b$ is not so large as to incentivize the merchant to fully exclude cash buyers in the price coherence regime.

**Proposition 2.** *With demand in the constant pass-through family, the merchant chooses prices that satisfy Condition 1 if and only if $\gamma \le 1$. Therefore, if $\gamma \le 1$, price coherence gives rise to greater consumer surplus than price flexibility.*

To interpret Proposition 2, note the following implications.

[2]Indeed, Chen and Schwartz (2015) show that, in this prototypical model comparing uniform pricing and differential pricing, the violation of Condition 1 is sufficient for consumer surplus to be greater under the latter. See Lemma 1(i) of that article.

- Under linear demand, consumer surplus is higher under price coherence than under price flexibility. This also holds for all demand in the constant pass-through family that is strictly concave.
- Linear demand tightly satisfies Condition 1, but this condition stronger than necessary to guarantee that consumers are better off under price coherence.

The two subcases of Example 1 illustrate these points.

**Example 1.** *Let* $\lambda = 1/2$, $\overline{v} = 100$, $b = 5$, *and* $f = 2$.

*(a)* Linear Demand ($\gamma = 1$): $p_d^* = 50$, $p_m^* = 53.5$, $\hat{p}^* = 51.75$, $S^F = 12.88 < S^C = 12.91$.

*(b)* Convex Demand ($\gamma = 2$): $p_d^* = 33.33$, $p_m^* = 36.33$, $\hat{p}^* = 34.89$, $S^F = 10.33 < S^C = 10.34$.

Figure 2 illustrates the mechanism that allows convex demand to yield greater consumer surplus under price coherence, even though it gives rise to prices that violate Condition 1. For the sake of argument, retain the assumption from Example 1 that there are an equal number of cardholders and non-cardholders. The crucial point is that, when $\gamma > 1$ and the regime switches from price coherence to price flexibility, although the price drop enjoyed by cardholders, $p_m^* - \hat{p}^*$, is smaller than the price hike suffered by cash buyers, $\hat{p}^* - p_d^*$, due to the convexity of $S(\cdot)$, consumer surplus changes faster over the interval $[\hat{p}^* - b, p_m^* - b]$ than it does over $\left[p_d^*, \hat{p}^*\right]$. In other words, although area A is wider than area B+C, the latter is taller.

Two points of specific comparison to other works are worth noting here. First, relating our model to that of Edelman and Wright (2015), who also compare consumer surplus under regimes of price coherence and price flexibility, a crucial difference is that, in our setting, total demand for the good is variable. In contrast, their setting uses the Vickrey-Salop "circular city" model of merchant competition in which total demand remains fixed across when the regime switches between price flexibility and price coherence. Consequently, in their setting, when one performs the analogous exercises as we have thus far in this section, one finds that Condition 1 always holds with equality, but consumer surplus remains constant, as the price changes simply amount to a transfer between cash buyers and card buyers.

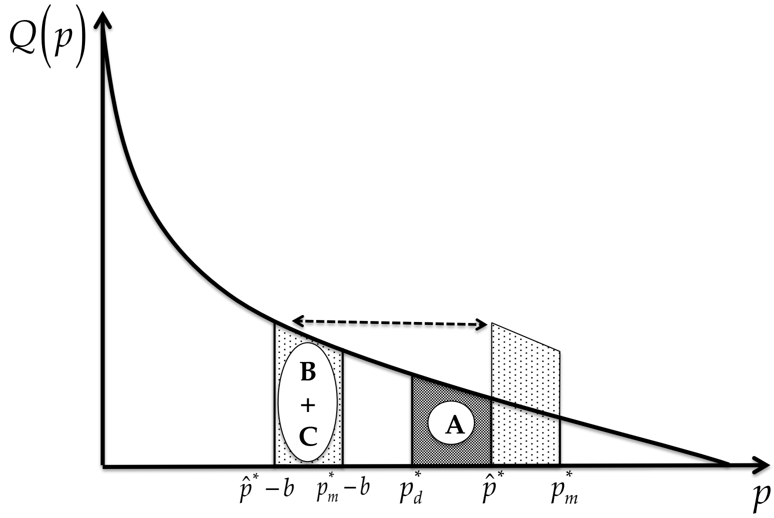


***Figure 2:*** *Change in consumer surplus following a shift in pricing regime, with a convex demand function. Here, unlike in Figure 1, Condition 1 fails to hold, yet area B + C may still exceed area A.*



Second, consider the relationship of our model to the environment in Section 3 of Chen and Schwartz (2015). There, a firm serves two markets with the same demand curve as one another, but it incurrs a low marginal cost in one of the markets and a high marginal cost in the other, and the authors compare consumer surplus under differential pricing and uniform pricing. As we mention in footnote 2, in their setting, Condition 1 is necessary *but not sufficient* for consumer surplus to be greater under uniform pricing. Indeed, in their setting, when demand is of the constant pass-through form defined in equation (6), Condition 1 always holds with equality (under any value of $\gamma$), but differential pricing still leads to higher consumer surplus. Our model's card usage benefit, $b$, absent in their setting, is the driving force behind this difference. Effectively, because of $b$, in our model, when the regime switches from price flexibility to price coherence, the valuation *for the good* of the marginal buyer in each group moves further apart, rather than closer together.

## 4 Equilibrium when transaction fees are endogenous

We now analyze the subgame-perfect equilibrium of the full game, including the first stage, in which the platform chooses both whether to impose price coherence or to allow the merchant price flexibility and the level of the transaction fee, $f$.

### 4.1 Linear demand

We focus, first, on the case of linear demand ($\gamma = 1$), which can be solved analytically. Proposition 3 highlights the important properties of equilibrium, which the proof fully characterizes. Here, we maintain the assumption (stated formally in the proof) that $b$ is small enough so that, under price coherence, the merchant does not exclude all cash buyers.

**Proposition 3.** *Under linear demand, in the unique subgame-perfect equilibrium of the game, the following statements hold.*

*(a) The platform chooses to impose price coherence.*

*(b) The transaction fee that the platform sets under price coherence is strictly lower than the transaction fee that it would set if it chose to allow price flexibility.*

*(c) Consumer surplus and total surplus are both greater under price coherence than under flexibility.*

The intuition behind Proposition 3 can most easily be appreciated by first recalling Lemmas 1 and 2. These state the maximum transaction fees that the merchant will tolerate before it refuses to accept the card, in the two respective regimes. They show that, under price flexibility, the merchant accepts a transaction fee as high as $b$, whereas, under price coherence, the maximum fee it agrees to is $\overline{f} < b$. Moreover, in the subgames corresponding two each of the two regimes, each of these constraints is binding for the platform; that is, $f^F = b$ and $f^C = \overline{f}$. Thus, it follows immediately from Proposition 2 that, with linear demand and the lower transaction fee under price coherence, Condition 1 strictly holds. Therefore, consumer surplus is greater under price coherence.

It is less straightforward to see why the platform should prefer price coherence to price flexibility. On the one hand, because $f^F < f^C$, the platform earns less per transaction under price coherence. On the other hand, however, the good is cheaper for cardholders under price coherence than it is under price flexibility: $\hat{p}^*\left(f^F\right) < p_m^*\left(f^C\right)$. Thus, price coherence leads to a larger volume of card transactions. Under linear demand, the latter effect dominates, leading



Preliminary Draft

the platform to prefer coherence.[3] Given this fact, it follows immediately that total surplus is greater under price coherence, because this regime favors both consumers and the platform, while the merchant is indifferent between the two regimes.

## 4.2 Constant pass-through demand

We next generalize to the case of constant pass-through demand, as defined in equation (6). This environment does not typically admit a closed-form solution for $\overline{f}$, the maximum transaction fee that the platform may charge under price coherence, so we solve the game numerically.

The following two features stand out. First, compared to the environment in which the transaction fee is exogenous, here, demand may be more convex and still lead consumers to be better off under price coherence. Second, as demand becomes too concave, even though consumers would be better off under price coherence, the platform chooses price flexibility at equilibrium. Example 2 illustrates these patterns by expanding on the analysis of Example 1.

**Example 2.** *Let $\lambda = 1/2$, $\overline{v} = 100$, $b = 5$ (as in Example 1), and let stage 1 unfold endogenously. The following table reports outcomes of interest for $\gamma \in \left\{\frac{1}{2}, 1, 2, 4\right\}$ and compares them to the case when $f$ is exogenous.*

| | Endogenous Stage 1 | | | Exogenous $f$ |
|---|---|---|---|---|
| | **Equilibrium Arrangement** | **Condition 1** | **Consumer Surplus** | |
| $\gamma = 1/2$ | Flexibility | ✓ | $S^C > S^F$ | $S^C > S^F$ |
| $\gamma = 1$ | Coherence | ✓ | $S^C > S^F$ | $S^C > S^F$ |
| $\gamma = 2$ | Coherence | ✗ | $S^C > S^F$ | $S^C > S^F$ |
| $\gamma = 4$ | Coherence | ✗ | $S^C > S^F$ | $S^C < S^F$ |

In order to interpret Example 2, note that, here, as under linear demand, the platform finds it optimal to set the transaction fee to the maximum level that the merchant is willing to accept: $\overline{f} = f^C < f^F = b$. This implies that the marginal cost to the merchant of selling to card buyers is lower under price coherence than it is under price flexibility. Consequently, some levels of convexity (e.g., $\gamma = 4$) that were high enough to yield greater consumer surplus under price flexibility when $f$ was exogenous now yield greater consumer surplus under price coherence.

[3] In Subsection 4.2, below, we explore this markup-versus-volume tradeoff for the platform in a more general setting.

The platform, however, continues to face the same tradeoff noted above in the case of linear demand. Price coherence leads to a larger volume of card transactions, but price flexibility allows it to set a higher transaction fee. When demand becomes too concave (e.g., $\gamma = 1/2$), the latter effect dominates, and the platform chooses to allow price flexibility.

The pattern that we illustrate here holds quite generally within the family of constant pass-through demand functions. In our numerical analysis, the main simplifying feature that we rely on is that, under both pricing arrangements, the platform's transaction fee constraints (discussed above) bind. A sufficient condition for this to hold is $\gamma \leq \overline{v}/b$. Thus, given our specific assumptions on $\overline{v}$ and $b$ in Example 2, these constraints bind for all $\gamma \leq 20$. Moreover, for all $\gamma$ below this threshold, when the transaction fee is set endogenously, then $S^C > S^F$.



# 5 Endogenous cardmembership

In this section, we endogenize buyers' decision of whether or not to become cardholders. Here, we find that, although price coherence always leads to higher joining costs for consumers, the potential for coherence to offer them higher surplus from transactions means that aggregate consumer surplus may be higher under either regime.

## 5.1 Setup

We assume that each member of the unit mass of buyers incurs a joining cost, $c$, if she chooses to become a cardholder, and that these costs are distributed according to distribution $H$. The new timing is as follows.

1. The platform sets the transaction fee, $f$, and it chooses which arrangement to use, price flexibility or price coherence.

2. Each buyer observes her value of $c$ and decides whether to join the platform. Simultaneously, the merchant chooses whether or not to accept the card. If it accepts, then, under price flexibility, it sets $p_d$ and $p_m$, whereas, under price coherence, it sets $\hat{p}$.

3. Each buyer learns her value of $v$ and chooses whether or not to purchase the good. In so doing, if the merchant has chosen to accept the card, cardholders can choose whether to pay by card or by cash.

The only difference here, compared to above, is the feature that, in stage 2, not only the merchant, but also buyers choose whether or not to join. Note that this timing matches that of Edelman and Wright (2015).

We solve the model numerically, adopting the following specifications. The distribution of joining costs, $H$, embeds our analysis above as a special case. Intuitively, this distribution can be described as follows. A share, $\xi \in [0,1)$, of consumers have joining cost of zero, while the complementary share, $1-\xi$, have (strictly) positive joining costs. Among those with positive joining costs, one share, $\phi \in [0,1]$, have costs that are uniformly distributed over the interval $(0,\bar{c})$, where $\bar{c}$ is large enough to be make joining prohibitively costly. The complementary share, $1-\phi$, among this group, all have prohibitively high joining cost, $\bar{c}$.[4]



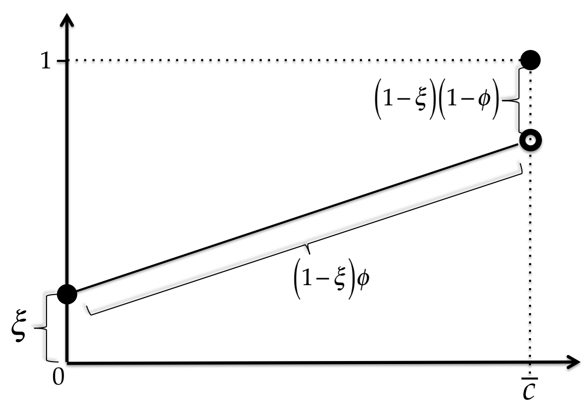


***Figure 3:*** *Cumulative distribution, $H$, on buyers' joining costs. It has has point mass $\xi$ at $c=0$, point mass $(1-\xi)\left(1-\phi\right)$ at $c=\bar{c}$, and the rest is uniformly distributed over $(0,\bar{c})$.*

Figure 3 plots this distribution. Note that, when $\phi=0$, this is equivalent to the setup studied above, in that $1-\xi=\lambda$ always choose to be non-cardholders, and $\xi=1-\lambda$ always choose to

[4]Formally, this distribution takes the form

$$H(c)=\begin{cases}\xi, & c=0,\\ \xi+(1-\xi)\,\phi\cdot c/\bar{c}, & 0<c<\bar{c},\\ 1, & c=\bar{c}.\end{cases}$$

join. Meanwhile, when $\phi = 1$ and $\xi = 0$, we have a simple uniform distribution over $(0, \overline{c})$.

Beyond merely embedding these two canonical cases, this specification has a natural interpretation. The share, $\xi$, of buyers with zero joining costs may obtain the card automatically under various possible circumstances, such as when signing up for a bank or brokerage account. They may also be approached by marketing staff at, for instance, entertainment events or airports, in a way that allows for essentially costless sign-up (or, indeed, in a way that makes it costlier to decline!) Meanwhile the share, $(1-\xi)(1-\phi)$, may include groups such as buyers with foreign residency/citizenship who are ineligible to sign up for the card.[5]

As in Subsection 4.1, we focus on the case of linear demand for the good; i,e., the distribution, $G$, of valuations for the good is uniform over the interval $[0, \overline{v}]$. We further assume that buyers joining costs and valuations for the good are independently distributed.

In the second stage, given transaction fee $f$ and the selected regime, each buyer joins the platform if and only if her joining cost, $c$, is below some threshold, $\tilde{c}$. Thus, under the respective regimes, the total mass of buyers that join is $1 - \lambda^F = H\left(\tilde{c}^F\right)$ and $1 - \lambda^C = H\left(\tilde{c}^C\right)$, and aggregate joining costs are given by $L^F \equiv \int_{x \leq \tilde{c}^F} x\, dH(x)$ and $L^C \equiv \int_{x \leq \tilde{c}^C} x\, dH(x)$. We maintain the definitions $S^F \equiv \lambda^F S(p_d) + \left(1 - \lambda^F\right) S(p_m - b)$ and $S^C \equiv \lambda^C S(\hat{p}) + \left(1 - \lambda^C\right) S(\hat{p} - b)$, which now denote consumer surplus derived from purchasing the good. Define *inclusive* consumer surplus (taking into account joining costs) as $W^F \equiv S^F - L^F$ and $W^C \equiv S^C - L^C$. Let $T^F$ and $T^C$ denote total surplus under the two regimes, defined as the sum of inclusive consumer surplus, merchant profits and platform profits.

### 5.2 Equilibrium

Example 3 characterizes the equilibrium of this model under specific numerical values for parameters $\overline{v}$, $b$, and $\overline{c}$. However, the outcomes we describe hold much more generally in the environment with uniform, independently distributed joining costs, $c$ and valuations, $v$. We have explored many other parameter values, beyond those reported here and have always found

[5]Note, however, that the size, $\xi$, of the former group with zero joining cost is more meaningful than the size, $(1-\xi)(1-\phi)$, of the latter group with joining cost $\overline{c}$, because increases in $\phi$ can be offset by increases in $\overline{c}$, so as to hold fixed the total mass of consumers with joining costs above any particular threshold.

qualitatively similar results, so long as (i) $b$ is not so large that the merchant is incentivized to fully exclude cash buyers in the price coherence regime, and (ii) $\overline{c} > b$, so that some buyers always choose not to join the platform.

**Example 3.** *Let $\overline{v} = 100$ and $b = 5$ (as in Examples 1 and 2), and let $\overline{c} = 10$. The following points hold.*

*(a) For all values of $\left(\xi, \phi\right) \in [0, 1]^2$,*

- *at equilibrium, the platform chooses to impose price coherence. It sets its transaction fee equal to the maximum level that the merchant is willing to accept, $\overline{f}$.*
- *more buyers join the platform under price coherence than under flexibility ($\tilde{c}^F < \tilde{c}^C$),*
- *consumer surplus derived from purchasing the good and total surplus are greater under price coherence than under flexibility ($S^C > S^F, T^C > T^S$).*

*(b) When the mass of buyers with zero joining costs is sufficiently large, relative to the mass buyers that are potentially marginal, then inclusive consumer surplus is greater under price coherence ($W^C > W^F$). Otherwise, the reverse is true.*

*Figures 4 and 5 further illustrate this example.*

These results raise the following three points. First, in this setting where consumers endogenously decide whether or not to join the platform, the main theme discussed in previous sections continues to hold. That is, the surplus that consumers derive from *transactions* (i.e., including their valuations for the good, $v$, and their benefit from using the card, $b$) can be greater under price coherence than under flexibility. Specifically, here, we study the case of linear demand for the good and find this to hold.

Second, *inclusive consumer surplus* may be greater under either regime. On the one hand, so long as there are *some* consumers "on the margin" between joining the card or not, aggregate joining costs are greater under price coherence. To see why, note that, under coherence, the equilibrium price of the good is lower than under flexibility, and, therefore, the joining cost of the marginal consumer must be higher. Consequently, the crucial factor determining whether price coherence helps consumers, overall, is whether increased surplus from transactions exceeds

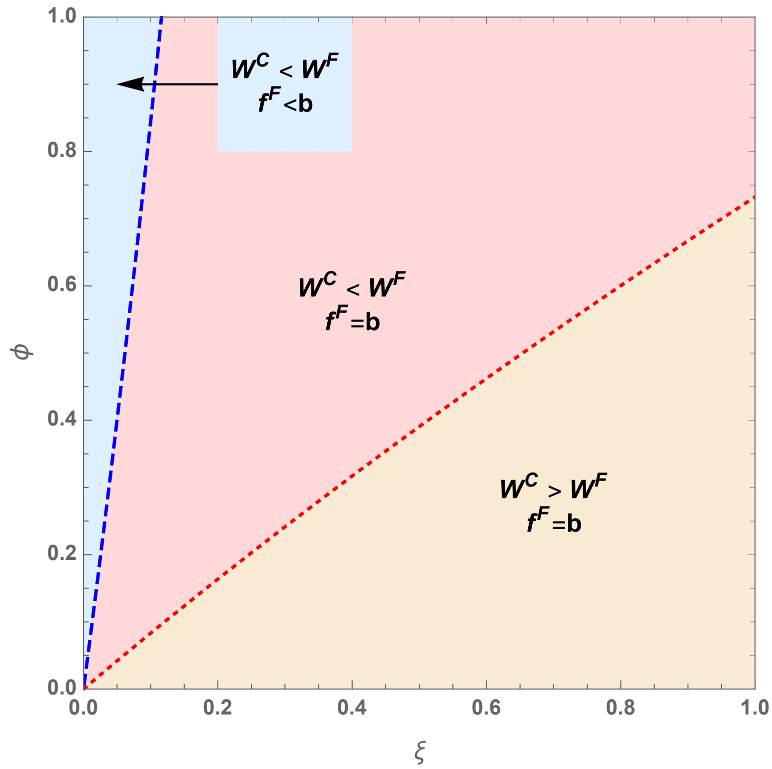


***Figure 4:*** *Comparisons, under the two regimes, of inclusive consumer surplus, W, and the platform's optimal transaction fee under price flexibility, $f^F$.*

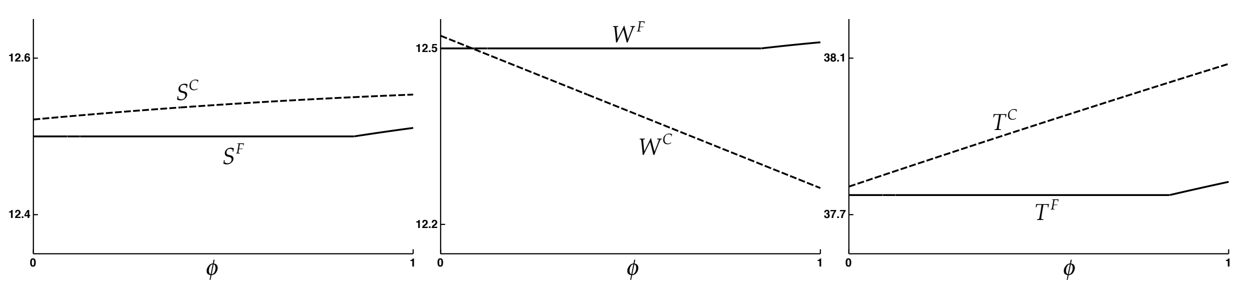


***Figure 5:*** *Here, $\xi$=0.1; these plot comparisons under the two regimes of, from left to right, consumer surplus derived from purchasing the good, S, inclusive consumer surplus, W, and total surplus, T.*

the higher joining costs. Figures 4 and 5 show circumstances in which each of these outcomes prevail. In particular, Figure 4 shows that when the relative mass of potentially undecided consumers is low, compared to those who could be swayed either way, then price coherence gives rise to higher inclusive consumer surplus.

On the other hand, when the set of potentially undecided consumers becomes significant enough, the opposite holds. Consumers are worse off at equilibrium, featuring price coherence, than they would be if the platform were required to allow price flexibility but could still set its merchant fee optimally. At one level, our results under such parameter values mirror the findings of Edelman and Wright (2015), who also find price flexibility to be better for consumers. However, the crucial difference is that, in their setting with Vickrey-Salop competition among merchants, consumer surplus from transactions is the same under the two regimes, and thus the sole driving force is the higher joining costs under price coherence. In our setting with a variable volume of total transactions, there is a potentially more interesting tradeoff.

For example, consider the following stylized description of a policy decision. Suppose that, at the *status quo* equilibrium, a competition authority observes a payment platform to be imposing price coherence. Moreover, assume for the sake of argument that the authority's objective is to maximize consumer surplus. Using the lens of our model, if the authority had the ability to forbid the platform from imposing price coherence, should it do so, provided that the parameters of the model give rise to greater inclusive consumer surplus ($W^F > W^C$) under flexibility? The answer does not seem obvious if the market is already relatively mature. If consumers have already joined the platform based on the anticipation of the pricing regime that arises under coherence, and their joining costs are sunk, a prohibition on price coherence has the potential to further harm them by leading to a less desirable pricing regime for the good. Finally, note that, when total surplus is the relevant measure, price coherence broadly performs better than price flexibility.

## 6 Conclusion

This paper examines a familiar yet unresolved and controversial question. Should a payment intermediary (e.g. American Express, PayPal, etc.) be allowed to require merchants to charge the same price for customers who pay via the platform as they charge other customers? We build a model that highlights certain aspects of this questions that we believe have been under-appreciated. Taking these into account, we find that other recent work on the topic may underestimate the possibility for such platform-imposed requirements to benefit consumers.

Broadly speaking, we find that when merchants are allowed to charge different prices to buyers who pay by card and those who pay by cash, this helps the merchant to extract from consumers paying by card the value of the additional convenience that the card offers. As a result, when the payment intermediary requires merchants to charge a single price to all consumers, it can make consumers better off. At the same time, however, when consumers endogenously decide whether to sign up with the platform, under such a single price rule, inefficiently many consumers to sign up. Whereas the latter effect has been well understood thanks to Edelman and Wright (2015), our results in this paper suggest that there may be a significant tradeoff between higher transaction surplus and excessive joining costs.

## References

Aguirre, Iñaki, Simon Cowan, and John Vickers (2010), "Monopoly Price Discrimination and Demand Curvature." *American Economic Review*, 100, 1601–1615.

Bourguignon, Helene, Renato Gomes, and Jean Tirole (2018), "Shrouded Transaction Costs: Must-Take Cards, Discounts and Surcharges." *Working Paper*. Available at https://sites.google.com/site/northwesternrenatogomes/Home/research.

Bulow, Jeremy I. and Paul Pfleiderer (1983), "A Note on the Effect of Cost Changes on Prices." *Journal of Political Economy*, 91, 182–85.

Chen, Yongmin and Marius Schwartz (2015), "Differential Pricing When Costs Differ: A Welfare Analysis." *RAND Journal of Economics*, 46, 442–460.

Cowan, Simon (2012), "Third-Degree Price Discrimination and Consumer Surplus." *Journal of Industrial Economics*, LX, 333–345.

Cowan, Simon (2016), "Welfare-increasing third-degree price discrimination." *RAND Journal of Economics*, 47, 326–340.

Edelman, Benjamin and Julian Wright (2015), "Price Coherence and Excessive Intermediation." *Quarterly Journal of Economics*, 130, 1283–1328.

Gomes, Renato and Jean Tirole (2018), "Missed Sales and the Pricing of Ancillary Goods." *Quarterly Journal of Economics*, 133, 2097–2169.

Schmalensee, Richard (1981), "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination." *American Economic Review*, 71, 242–247.

Schwartz, Marius and Daniel R. Vincent (2006), "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network." *Review of Network Economics*, 5, 72–102.

Weyl, E. Glen and Michal Fabinger (2013), "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition." *Journal of Political Economy*, 121, 528–583.

# Appendices

## A Proofs

***Proof of Lemma 1.*** Let $\tilde{p} \equiv p_m - b$ and rewrite equation (1) as

$$\lambda Q(p_d)\,p_d + (1-\lambda)\,Q(\tilde{p})\,(\tilde{p} - (f-b)). \tag{7}$$

If $f > b$, then the effective marginal cost of selling by card is positive. Therefore, it is more profitable for the merchant not to accept the card (or, equivalently, to accept the card and set $p_m$ to an arbitrarily high level that induces no card sales). If $f \le b$, even if the merchant were to set $\tilde{p} = p_d$, it would make (weakly) greater profits than if it didn't accept the card and charged that same price. When $f < b$, by separately setting prices optimally for each group, it makes strictly greater profits. □

***Proof of Lemma 2.*** We show that, if $f = 0$, accepting the card makes the merchant strictly better off, and, if $f = b$, doing so makes it strictly worse off. Note, first, that if the merchant refuses to accept the card, its maximal profits are $Q\left(p_d^*\right)p_d^*$. Now, suppose $f = 0$. If the merchant accepts the card and sets $\hat{p} = p_d^*$, it earns $\lambda Q\left(p_d^*\right)p_d^* + (1-\lambda)\,Q\left(p_d^* - b\right)p_d^* > Q\left(p_d^*\right)p_d^*$. Next, suppose $f = b$. Lemma 1 shows that, in this case, under price flexibility, the merchant is indifferent whether or not to accept the card. That is, $Q\left(p_d^*\right)p_d^* = \lambda Q\left(p_d^*\right)p_d^* + (1-\lambda)\,Q\left(p_m^* - b\right)\left(p_m^* - b\right)$, which implies that $p_d^* = p_m^* - b$. Since this profit level requires the merchant to charge the two groups different prices, it is not feasible under price coherence. □

***Proof of Lemma 3.*** First we show that $p_m^* - b < p_d^*$. Lemma 2 implies that the merchant accepts the card under price coherence only if $f \le \overline{f} < b$, so the term $f - b$ in equation (7) is strictly negative. Thus, the optimal prices for each group satisfy $p_d^* > \tilde{p}^* = p_m^* - b$.

Next we show that $p_d^* < p_m^*$. For any $f, b > 0$, denote by $p^*(f,b)$ the solution to

$$p = f + \frac{Q(p-b)}{-Q'(p-b)}.$$

We have $p_d^* = p^*(0,0)$ and $p_m^* = p^*(f,b)$. Implicit function theorem implies that

$$\frac{\partial p^*}{\partial f} = -\frac{-1}{1 + \frac{(Q')^2 - QQ''}{(Q')^2}} = \frac{1}{2 - \frac{Q''}{Q'}\Big/\frac{Q'}{Q}} > 0,$$

$$\frac{\partial p^*}{\partial b} = -\frac{-\frac{(Q')^2 - QQ''}{(Q')^2}}{1 + \frac{(Q')^2 - QQ''}{(Q')^2}} = \frac{1 - \frac{Q''}{Q'}\Big/\frac{Q'}{Q}}{2 - \frac{Q''}{Q'}\Big/\frac{Q'}{Q}} > 0,$$

since the log-concavity of $Q(\cdot)$ from Assumption 1 indicates that $\frac{Q''}{Q'} \Big/ \frac{Q'}{Q} < 1$ globally. Thus, $p_d^* = p^*(0,0) < p^*(f,b) = p_m^*$.

Finally, because the profit functions for each group are single-peaked, it follows that $p_d^* < \hat{p}^* < p_m^*$, with $Q(\hat{p}^*) + \hat{p}^* Q'(\hat{p}^*) < 0$ and $Q(\hat{p}^* - b) + (\hat{p}^* - f) Q'(\hat{p}^* - b) > 0$. This implies that $\hat{p}^* - b < p_m^* - b$. $\square$

***Proof of Proposition 1***. By assumption, $\hat{p} - b < p_m - b < p_d < \hat{p} < p_m$. Hence, Lagrange's mean value theorem implies that

$$\begin{aligned} S^C(\hat{p}) > S^F(p_d, p_m) &\Leftrightarrow \lambda S(\hat{p}) + (1-\lambda) S(\hat{p} - b) > \lambda S(p_d) + (1-\lambda) S(p_m - b) \\ &\Leftrightarrow (1-\lambda)(S(\hat{p} - b) - S(p_m - b)) > \lambda (S(p_d) - S(\hat{p})) \\ &\Leftrightarrow (1-\lambda) \cdot ((p_m - b) - (\hat{p} - b)) \cdot (-S'(\xi_1)) > \lambda \cdot (\hat{p} - p_d) \cdot (-S'(\xi_2)), \end{aligned}$$

where $\hat{p} - b < \xi_1 < p_m - b < p_d < \xi_2 < \hat{p}$. Since $S(\cdot)$ is convex, $-S'(\xi_1) > -S'(\xi_2)$. Thus, it suffices to have $(1-\lambda)(p_m - \hat{p}) \geq \lambda (\hat{p} - p_d)$, which is equivalent to Condition 1. $\square$

***Proof of Proposition 2***. From optimality condition (3), we have

$$p_d^* = \frac{\overline{v}}{1+\gamma}, \quad p_m^* = \frac{\overline{v} + b + \gamma f}{1+\gamma},$$

so

$$p_\lambda \equiv \lambda p_d^* + (1-\lambda) p_m^* = \frac{\overline{v} + (1-\lambda)(b + \gamma f)}{1+\gamma},$$

and Condition 1 is equivalent to $\hat{p}^* \leq p_\lambda$. Since $\hat{p}^*$ satisfies FOC (4)

$$\lambda (Q(\hat{p}^*) + \hat{p}^* Q'(\hat{p}^*)) + (1-\lambda)(Q(\hat{p}^* - b) + (\hat{p}^* - f) Q'(\hat{p}^* - b)) = 0,$$

$\hat{p}^* \leq p_\lambda$ if and only if the LHS of equation (4) is weakly negative when evaluated at $p_\lambda$, i.e.,

$$\lambda (Q(p_\lambda) + p_\lambda Q'(p_\lambda)) + (1-\lambda)(Q(p_\lambda - b) + (p_\lambda - f) Q'(p_\lambda - b)) \leq 0,$$

which is equivalent to

$$\lambda (1-\lambda)(b + \gamma f) \left( (\overline{v} + b - p_\lambda)^{\gamma - 1} - (\overline{v} - p_\lambda)^{\gamma - 1} \right) \leq 0 \quad \Leftrightarrow \quad \gamma \leq 1.$$

$\square$

***Proof of Proposition 3***. Under linear demand, $Q(p) = 1 - p/\overline{v}$. Thus,

$$S(p) = \int_p^{\overline{v}} Q(x)\, dx = \frac{(\overline{v} - p)^2}{2\overline{v}}.$$

Assume that $b < \overline{v}/2$. From optimality conditions (3) and (4), we obtain

$$p_m^*(f) = \frac{\overline{v} + b + f}{2}, \quad p_d^* = \frac{\overline{v}}{2}, \quad \hat{p}^*(f) = \frac{\overline{v} + (1-\lambda)(b+f)}{2}.$$

By definition, $\overline{f}$ solves

$$(1-\lambda)\, Q\left(\hat{p}^*\left(\overline{f}\right) - b\right) \cdot \left(\hat{p}^*\left(\overline{f}\right) - \overline{f}\right) + \lambda Q\left(\hat{p}^*\left(\overline{f}\right)\right) \hat{p}^*\left(\overline{f}\right) = Q\left(p_d^*\right) p_d^*,$$

which yields

$$\overline{f} = \frac{2\overline{v} + (1-\lambda)\, b}{\overline{v} + (1+\lambda)\, b + \sqrt{\overline{v}^2 + 4\lambda \overline{v} b + 4\lambda b^2}} b \in (0, b)\,.$$

Now consider the platform's optimal transaction fees under the two regimes, $f^F$ and $f^C$:

$$f^F = \operatorname*{argmax}_{f \leq b} \; Q\left(p_m^*(f) - b\right) f, \quad f^C = \operatorname*{argmax}_{f \leq \overline{f}} \; Q\left(\hat{p}^*(f) - b\right) f.$$

First consider price flexibility. FOD of the platform's profits evaluated at $f = b$ is

$$Q\left(p_m^*(b) - b\right) + b \cdot Q'\left(p_m^*(b) - b\right) \cdot \left.\frac{dp_m^*}{df}\right|_{f=b} = \frac{1}{2} + b \cdot (-1/\overline{v}) \cdot \frac{1}{2} = \frac{1 - b/\overline{v}}{2} > 0,$$

so it would profit from a marginal increase in $f$ when $f = b$ if it could ignore the merchant's acceptance constraint. Hence, the constrained maximization leads to corner solution, which implies that $f^F = b$.

Next consider price coherence. Similarly, FOD of the platform's profits evaluated at $f = b$ is

$$Q\left(\hat{p}^*(b) - b\right) + b \cdot Q'\left(\hat{p}^*(b) - b\right) \cdot \left.\frac{d\hat{p}^*}{df}\right|_{f=b} = \frac{1}{2} + \lambda\,(b/\overline{v}) + b \cdot (-1/\overline{v}) \cdot \frac{1-\lambda}{2} = \frac{1 - (1-3\lambda)\, b/\overline{v}}{2} > 0,$$

so, similarly, the constrained maximization leads to corner solution, which implies that $f^C = \overline{f} < b = f^F$, i.e., the transaction fee that the platform sets under price coherence is strictly lower than the transaction fee that it would set if it chose to allow price flexibility.

Note that

$$\begin{aligned} & Q\left(\hat{p}^*\left(f^C\right) - b\right) f^C - Q\left(p_m^*\left(f^F\right) - b\right) f^F \\ = & \frac{2\lambda(1-\lambda) b^3\, (\overline{v} + b)}{\overline{v}\left(\overline{v}^2 + (1+3\lambda)\, \overline{v} b + 4\lambda b^2 + (\overline{v} + (1+\lambda) b)\, \sqrt{\overline{v}^2 + 4\lambda \overline{v} b + 4\lambda b^2}\right)} > 0, \end{aligned}$$

so the platform chooses to impose price coherence.

It remains to compare surplus. Proposition 2 implies that if the transaction fee is set at an exogenously level $f = \overline{f}$, price coherence gives rise to greater consumer surplus than price flexibility. Now that price flexibility would lead to a strictly higher transaction fee, the consumers would be hurt even more. Therefore, consumer surplus is greater under price coherence than under flexibility. It follows immediately that total surplus is greater under price coherence, because this regime favors both consumers and the platform, while the merchant is indifferent between the two regimes (it would earn $Q\left(p_d^*\right)p_d^*$ under both). This completes the proof. $\square$